Notes on General Relativity

Jiashu Han

December 22, 2019

Contents

1 Manifolds and Tensor Fields

General relativity is a theory about curvature in spacetime - how things move inside a curved spacetime, and how these things change the curvature of spacetime. Before we delve into the physics of general relativity, we want a mathematical formalism that can be readily used to describe the concepts in GR. In this section we will briefly introduce some of the most important concepts, according to Robert M. Wald's *General Relativity* [1].

1.1 Manifolds

From our experience, spacetime is a "four-dimensional continuum" because it requires four numbers to specify an event. In special relativity we assume this is globally true and that all events in spacetime have one-to-one correspondence with the points in $\mathbb{R}^4$. However, in general relativity we are solving for the geometry of spacetime and we do not want to make assumptions about the global structure of spacetime in advance. For example, before Magellan set out to investigate the structure of the surface of the Earth, he might notice that he could characterize the positions in his vicinity by two numbers, but it would be wrong to assume this fact is globally true and all points on the Earth have a one-to-one correspondence with the points in $\mathbb{R}^2$.

In the case of the Earth, its surface lives in the higher dimensional Euclidean space $\mathbb{R}^3$ of all space points, so the study of two-dimensional surfaces embedded in $\mathbb{R}^3$ is adequate for analyzing the structure of the Earth's surface and we do not need to introduce the abstract concept of a manifold. However, in GR, as far as we know the spacetime does not live in a higher dimensional Euclidean space, and the concept of a manifold would be very useful.

First, consider an *open ball* in $\mathbb{R}^n$ of radius r centered around point $y = (y^1, ..., y^n)$, which consists of points x that satisfy $|x - y| < r$, where

$$|x - y| = \left[\sum_{\mu} (x^\mu - y^\mu)^2\right]^{1/2}. \tag{1.1.1}$$

An *open set* in $\mathbb{R}^n$ is any set that can be expressed as a union of open balls, and $\mathbb{R}^n$ is a topological space - a set of points along with a set of neighborhoods for each point that satisfies a set of axioms relating the points and neighborhoods. A **manifold** is a topological space that locally resembles a Euclidean space near each point p. It is a set made up of pieces that look like open subsets of

$\mathbb{R}^n$ such that they can be sewn together smoothly. Each point of an n-dimensional manifold has a neighborhood homeomorphic to an open subset of $\mathbb{R}^n$[1]. For example, one-dimensional manifolds include lines and circles, but not figure eights, which have crossing points that are not homeomorphic to $\mathbb{R}$.

A manifold locally resembles the Euclidean space, but this may not be true globally. For example, the surface of the sphere is not homeomorphic to the Euclidean plane for reasons such as compactness[2]. However, a region/open subset O_α of the manifold can be mapped into an open subset U_α of the Euclidean plane using a map projection (**chart**, a.k.a. **coordinate system**) ψ_α. When a region appears in two neighboring charts, the two charts are not exactly the same and we need a transformation (**transition map**) $\psi_\beta \circ \psi_\alpha^{-1}$ to connect them. Manifolds can include additional structure, such as a differentiable structure that allows calculus to be done on the manifolds. This special class of manifolds is called **differentiable manifolds**.

If the transition from one chart to another is differentiable, i.e. the charts are compatible, then computations done in one chart are valid in any other differentiable chart. A C^k-manifold is a topological manifold with a chart whose transition maps are all k-times continuously differentiable. A smooth manifold (C^∞-manifold) is a differentiable manifold for which all transition maps are smooth so it is a C^k-manifold for all k.

To summarize, an n-dimensional, C^∞, real manifold M satisfies the following properties:

(1) Each point $p \in M$ lies in at least one subset O_α, i.e. the collection $\{O_\alpha\}$ covers M.

(2) For each α, there is a bijective map $\psi_\alpha : O_\alpha \to U_\alpha$ where U_α is an open subset of $\mathbb{R}^n$.

(3) If any two sets O_α and O_β overlap, i.e. $O_\alpha \cap O_\beta \neq \varnothing$, then we have the map $\psi_\beta \circ \psi_\alpha^{-1}$ which takes points in $\psi_\alpha[O_\alpha \cap O_\beta] \subset U_\alpha \subset \mathbb{R}^n$ to points in $\psi_\beta[O_\alpha \cap O_\beta] \subset U_\beta \subset \mathbb{R}^n$.

We also require in the definition of M that all coordinate systems compatible with (2) and (3) are included so that one cannot define a new manifold just by adding or deleting in a coordinate system. Note that each chart ψ_α is also a homeomorphism.

$\mathbb{R}^n$ is a trivial example of a manifold and is covered by a single chart ($O = \mathbb{R}^n$, $\psi =$ identity map).We cannot map an entire 2-sphere into $\mathbb{R}^n$ in a continuous, one-to-one manner, but we can do so by defining six hemispherical open sets $O_i^\pm$, each of which can be mapped homeomorphically into an open disk D via the projection maps $f_i^\pm$ such that $f_1^+(x^1, x^2, x^3) = (x^2, x^3)$, etc. It can be checked that the transition maps/overlap functions $f_i^\pm \circ (f_j^\pm)^{-1}$ are C^∞.

For two manifolds M and M' of dimension n and n', we can make the product space $M \times M'$ consisting of all pairs (p, p') into an $(n+n')$-dimensional manifold by defining the chart $\psi_{\alpha\beta} : O_{\alpha\beta} \to U_{\alpha\beta} \subset \mathbb{R}^{n+n'}$ on $M \times M'$, where $O_{\alpha\beta} = O_\alpha \times O'_\beta$, $U_{\alpha\beta} = U_\alpha \times U'_\beta$, and $\psi_{\alpha\beta}(p, p') = [\psi_\alpha(p), \psi'_\beta(p')]$.

We can now define differentiability and smoothness of maps between manifolds. A map $f : M \to M'$ is C^∞ if for each α and β, the map $\psi'_\beta \circ f \circ \psi_\alpha^{-1}$ taking $U_\alpha \subset \mathbb{R}^n$ into $U'_\beta \subset \mathbb{R}^{n\prime}$ is C^∞. f is called a **diffeomorphism** if it is C^∞, bijective and its inverse is C^∞.

1.2 Vectors

In pre-GR physics, space has the natural structure of a three- or four-dimensional vector space with a point designated as origin and the vector space axioms apply. This structure is lost in curved

[1] A **homeomorphism** is the mapping that preserves all the topological properties of a given space.

[2] A generalization of the notion of closedness and boundedness in the Euclidean space.

geometries. For example, it is hard to define how to "add" two points on a sphere to obtain a third point. We can recover the vector space structure in the limit of infinitesimal displacements, or **tangent vectors**, about a point. For manifolds embedded in $\mathbb{R}^n$, a tangent vector at p can be thought of as a vector lying in the tangent plane of the surface at that point.

However, in cases where a manifold is not embedded in $\mathbb{R}^n$ we need to to define a tangent vector by referring only to the intrinsic structure of the manifold. We do so by treating the tangent vector as a **directional derivative**. There is a one-to-one correspondence between vectors $v = (v^1, ..., v^n)$ and directional derivatives $\sum_\mu v^\mu(\partial/\partial x^\mu)$ in $\mathbb{R}^n$. The latter are characterized by linearity and the Leibnitz rule when acting on functions, so we can define a vector v at point $p \in M$ to be a map $v : \mathcal{F} \to \mathbb{R}$, where $\mathcal{F}$ is the collection of C^∞ functions from M to $\mathbb{R}$, that satisfies the following properties:

(1) Linearity: $v(af + bg) = av(f) + bv(g)\ \forall f, g \in \mathcal{F}; a, b \in \mathbb{R}$

(2) Leibnitz rule: $v(fg) = v(f)g(p) + f(p)v(g)$

Note these rules imply that for a constant function $h(q) = c\ \forall q \in M$, we have $v(h) = 0$.

It is easy to see that the collection of tangent vectors V_p at $p \in O \subset M$ has the structure of a vector space under the laws of addition and scalar multiplication. Another property states that $\dim V_p = n$ for an n-dimensional manifold M. For a chart $\psi : O \to U \subset \mathbb{R}^n$, we can prove this by constructing a **coordinate basis** $\{X_\mu\}$ of the tangent space given by

$$X_\mu(f) = \frac{\partial}{\partial x^\mu}(f \circ \psi^{-1})\big|_{\psi(p)} \tag{1.2.1}$$

where x^μ are the Cartesian coordinates of $\mathbb{R}^n$. A detailed proof can be found on page 15 of Wald.

One can also denote X_μ as $\partial/\partial x^\mu$, and we would get a different coordinate basis $\{X'_\nu\}$ if we chose a different chart ψ'. We can relate the two bases using chain rule:

$$X_\mu = \sum_{\nu=1}^{n} \frac{\partial x'^\nu}{\partial x^\mu}\big|_{\psi(p)} X'_\nu \tag{1.2.2}$$

where x'^ν is the ν-th component of the map $\psi' \circ \psi^{-1}$. Therefore, the components v'^ν of a vector v in the new basis are related to the components v^μ in the old basis by the *vector transformation law*

$$v'^\nu = \sum_\mu v^\mu \frac{\partial x'^\nu}{\partial x^\mu} \tag{1.2.3}$$

1.2.1 Curves

A **smooth curve** C is a C^∞ map $C : \mathbb{R} \to M$, with a parameter t. We associate with C a tangent vector $T \in V_p$ at each point $p \in M$ by setting $T(f)$ equal to the derivative of the function $f \circ C : \mathbb{R} \to \mathbb{R}$ evaluated at p. Note that C on M will be mapped into a curve $x^\mu(t)$ on $\mathbb{R}^n$ with a choice of coordinate system ψ. For any $f \in \mathcal{F}$, we have

$$T(f) = \frac{d(f \circ C)}{dt}\Big|_p = \sum_\mu \frac{\partial}{\partial x^\mu}(f \circ \psi^{-1})\frac{dx^\mu}{dt} = \sum_\mu \frac{dx^\mu}{dt} X_\mu(f) \tag{1.2.4}$$

In any coordinate basis, the components of T are

$$T^\mu = \frac{dx^\mu}{dt} \tag{1.2.5}$$

We can also define a vector space V_q at a point $q \in M$, but there is no way of determining whether a tangent vector at q is the same as a tangent vector at p given only the structure of a manifold. Later, given a connection (derivative operator) on the manifold, we will introduce the notion of parallel transport along a curve joining p and q. The identification of the two vector spaces will depend on the choice of the curve.

A **tangent field** v on M is an assignment of a tangent vector $v\big|_p \in V_p$ at each point p. Although V_p and V_q are different vector spaces, we can still define what it means for v to vary smoothly from point to point. If a function f is smooth, then at each p, $v(f)$ is a function on M and $v\big|_p(f)$ is a number. The tangent field is smooth if $v(f)$ is smooth for each f. Since X_μ are smooth, we have that v is smooth if and only if its coordinate basis components v^μ are smooth.

1.2.2 Precise meaning of tangent vector

We can give precise meaning to the idea of tangent vectors as infinitesimal displacements. Consider a *one-parameter group of diffeomorphisms* ϕ_t, which is a C^∞ map from $\mathbb{R} \times M \to M$. For a fixed parameter $t \in \mathbb{R}$, $\phi_t : M \to M$ is a diffeomorphism and $\phi_t \circ \phi_s = \phi_{t+s}\ \forall t, s \in \mathbb{R}$. ϕ_0 is the identity map.

For a fixed $p \in M$, $\phi_t(p) : \mathbb{R} \to M$ takes in a parameter $t \in \mathbb{R}$ and maps it to a point on M, thus we can think of $\phi_t(p)$ as a curve, called the **orbit** of ϕ_t. $\phi_t(p)$ passes through p at $t = 0$. Define $v\big|_p$ as the tangent to this curve at $t = 0$, then the vector field v is the infinitesimal generator of these transformations ϕ_t. We thus associate the vector field v to ϕ_t.

We can also do the converse and find the integral curves of v given a smooth vector field v by solving the system

$$\frac{dx^\mu}{dt} = v^\mu(x^1, ..., x^n) \tag{1.2.6}$$

if we pick a coordinate system in the neighborhood of p. The point $\phi_t(p)$ is the point at parameter t on the integral curve of v starting at p.

1.2.3 Commutator

For two smooth vector fields v and w it is possible to define a new vector field called the **commutator**

$$[v, w](f) = v[w(f)] - w[v(f)] \tag{1.2.7}$$

The commutator of any two vector fields X_μ and X_ν occurring in a coordinate basis is zero. Conversely, given a collection $\{X_i\}$ of nonvanishing, linearly independent, commuting vector fields, one can always find a chart for which they are the coordinate basis vector fields.

1.3 Tensors

In the previous section we introduced displacement vectors, and there are many quantities that have linear or multilinear dependence on displacements. An example would be the measurement of the magnetic field - we do not need to measure the projection of the field in every possible

orientation of the probe. Only three linearly independent directions are needed because the field strength has a linear dependence on the probe direction.

This means the magnetic field can be treated as a vector, or more precisely as a dual vector. We can define a **dual vector** as a collection of three numbers associated with a basis of spatial displacement vectors that transform inn a certain way when the basis is changed. We can also define it as a linear map from spatial displacement vectors to numbers. We will show that we can associate any dual vector with an ordinary vector because space has a metric defined on it.

Similarly, if a quantity depends linearly on more than one spatial displacement vector, such as the force per unit area exerted on a material body in equilibrium, we obtain a tensor (in this case the *stress tensor*). A **tensor** is a multilinear (linear in each variable) map from vectors (or dual vectors) to numbers.

Let V be any finite-dimensional vector space over $\mathbb{R}$. Consider the collection V^* of linear maps $f : V \to \mathbb{R}$. One can get a natural vector space structure on V^* with definitions of addition and scalar multiplication on such linear maps. We call V^* the **dual vector space** to V and the elements of V^* are the dual vectors. If $v_1, ..., v_n$ is a basis of V, we can define elements $v^{1*}, ..., v^{n*} \in V^*$ by

$$v^{\mu *}(v_\nu) = \delta^\mu_{\ \nu} \tag{1.3.1}$$

where $\delta^\mu_{\ \nu}$ is the Kronecker delta. It follows directly that $\{v^{\mu *}\}$ is a basis of V^* called the **dual basis** to the basis $\{v_\mu\}$ of V, and we have $\dim V^* = \dim V$, i.e. V and V^* are isomorphic. This isomorphism depends on the choice of basis and there is no natural way of identifying V with V^*.

We can apply the procedure twice to obtain the **double dual** vector space V^{**}. A vector $v^{**} \in V^{**}$ is a linear map from V^* to $\mathbb{R}$. We can show that V^{**} is isomorphic to V because we can associate a $v^{**} \in V^{**}$ to each $v \in V$ such that $v^{**}(v) = w^*(v)$, where $w^* \in V^*$. This means we can naturally identify V with V^{**}.

Let us now formally define a tensor. Let V be a finite dimensional vector space and V^* be its dual space. A **tensor** T of type (k, l) over V is a multilinear map

$$T : \underbrace{V^* \times ... \times V^*}_{k} \times \underbrace{V \times ... \times V}_{l} \to \mathbb{R}. \tag{1.3.2}$$

This means T produces a number from k dual vectors and l ordinary vectors. If we fix all but one of the vectors or dual vectors, it is a linear map on that variable.

For example, a type $(0, 1)$ tensor is a dual vector, and a type $(1, 0)$ tensor is a double dual vector but since we identify V^{**} with V, it is also an ordinary vector. This idea allows us to view a type $(1, 1)$ tensor T as a linear map from V to V or a linear map from V^* to V^* because it is an element of V^{**} if we fix $v \in V$ and an element of V^* if we fix $v^* \in V^*$. Therefore, a type (k, l) tensor T can be viewed as either $T : V \to \mathcal{T}(k, l-1)$ or $T : V^* \to \mathcal{T}(k-1, l)$.

The collection $\mathcal{T}(k, l)$ of all tensors of type (k, l) has the structure of a vector space given rules of addition and scalar multiplication, and a tensor is unique specified by giving its values on vectors in a basis of V and dual vectors in the dual basis of V^*. The dimension of the vector space $\mathcal{T}(k, l)$ is n^{k+l} because there are n^{k+l} linearly independent ways of filling the slots of a type (k, l) tensor, where n is the dimension of V.

We introduce two important operations on tensors. A **contraction** with respect to the i-th (dual vector) and j-th (vector) slots is a map $C : \mathcal{T}(k, l) \to \mathcal{T}(k-1, l-1)$. If $T \in \mathcal{T}(k, l)$, then

$$CT = \sum_{\sigma=1}^{n} T(..., v^{\sigma *}, ...; ..., v_\sigma, ...) \tag{1.3.3}$$

where $\{v_\sigma\}$ is a basis of V and is inserted into the j-th slot; $\{v^{\sigma *}\}$ is its dual basis and is inserted into the i-th slot. The contraction operation is indepenent of the choise of basis. The contraction of a type $(1,1)$ tensor $T : V \to V$ is just the trace of the map.

Another operation is the outer product. The **outer product** of two tensors $T \in \mathcal{T}(k,l)$ and $T' \in \mathcal{T}'(k',l')$ is a type $(k+k', l+l')$ tensor denoted $T \otimes T'$. Given $(k+k')$ dual vectors $v^{1*}, ..., v^{k+k'*}$ and $(l+l')$ vectors $w_1, ..., w_{l+l'}$,

$$S = T \otimes T' = T(v^{1*}, ..., v^{k*}; w_1, ..., w_l) \cdot T'(v^{k+1*}, ..., v^{k+k'*}; w_{l+1}, ..., w_{l+l'}) \tag{1.3.4}$$

A tensor is called *simple* if it can be expresses as an outer product.

If $\{v_\mu\}$ is a basis of V and $\{v^{\nu *}\}$ is its dual basis, the n^{k+l} simple tensors $\{v_{\mu_1} \otimes ... \otimes v_{\mu_k} \otimes v^{\nu_1 *} \otimes ... \otimes v^{\nu_l *}\}$ form a basis of $\mathcal{T}(k,l)$, so we can express every $T \in \mathcal{T}(k,l)$ as a sum of simple tensors

$$T = \sum_{\mu_1,...,\nu_l=1}^{n} T^{\mu_1 ... \mu_k}{}_{\nu_1 ... \nu_l} v_{\mu_1} \otimes ... \otimes v^{\nu_l *}. \tag{1.3.5}$$

The basis expansion coefficients $T^{\mu_1 ... \mu_k}{}_{\nu_1 ... \nu_l}$ are called the **components** of T with respect to that basis. The standard convention is to use superscripts for labels associated with vectors $(V^* \to \mathbb{R})$ and subscripts for those associated with dual vectors $(V \to \mathbb{R})$.

In the component form, the contraction and outer product formulas can be written as

$$(CT)^{\mu_1 ... \mu_{k-1}}{}_{\nu_1 ... \nu_{l-1}} = \sum_{\sigma=1}^{n} T^{\mu_1 ... \sigma ... \mu_{k-1}}{}_{\nu_1 ... \sigma ... \nu_{l-1}} \tag{1.3.6}$$

$$S^{\mu_1 ... \mu_{k+k'}}{}_{\nu_1 ... \nu_{l+l'}} = T^{\mu_1 ... \mu_k}{}_{\nu_1 ... \nu_l} T'^{\mu_{k+1} ... \mu_{k+k'}}{}_{\nu_{l+1} ... \nu_{l+l'}}. \tag{1.3.7}$$

Let us now a consider the special case where V is the tangent space V_p at $p \in M$. V_p^* is the **cotangent space** at p and its vectors are the **cotangent vectors**[3]. In section 1.2 we defined the coordinate basis $\partial/\partial x^1, ..., \partial/\partial x^n$ of V_p. The associated dual basis is denoted as $dx^1, ..., dx^n$. We can think of the dual basis vectors dx^μ as the linear map defined by $dx^\mu(\partial/\partial x^\nu) = \delta^\mu_\nu$. From Eq. 1.2.3 and 1.3.1 we have the transformation law for a dual vector with components w_μ:

$$w'_{\mu'} = \sum_{\mu=1}^{n} w_\mu \frac{\partial x^\mu}{\partial x'^{\mu'}} \tag{1.3.8}$$

The general **tensor transformation law** (sometimes used as the definition of tensors) is

$$T'^{\mu'_1 ... \mu'_k}{}_{\nu'_1 ... \nu'_l} = \sum_{\mu_1,...,\nu_l=1}^{n} T^{\mu_1 ... \mu_k}{}_{\nu_1 ... \nu_l} \frac{\partial x'^{\mu'_1}}{\partial x^{\mu_1}} ... \frac{\partial x^{\nu_l}}{\partial x'^{\nu'_l}}. \tag{1.3.9}$$

An assignment of a tensor over V_p at each point $p \in M$ is called a **tensor field**. We define a covariant field w to be smooth (C^∞) if for each smooth vector field v the function $w(v)$ is smooth. A tensor $T \in \mathcal{T}(k,l)$ is smooth if $T(w^1, ..., w^k; v_1, ..., v_l)$ is a smooth function for all smooth covariant vector fields $w^1, ..., w^k$ and smooth contravariant vector fields $v_1, ..., v_l$.

[3]We also commonly refer to vectors in V_p as **contravariant vectors** and vectors in V_p^* as **covariant vectors**.

1.3.1 The metric tensor

A metric tells us the "infinitesimal squared distance" associated with an "infinitesimal displacement". In section 1.2 we explained that an "infinitesimal displacement" can be described by a tangent vector, so a **metric** g must be "quadratic" in the tangent vector. g is therefore a type $(0,2)$ tensor field and a linear map $V_p \times V_p \to \mathbb{R}$. A metric must be **symmetric**, meaning $g(v_1, v_2) = g(v_2, v_1)\ \forall v_1, v_2 \in V_p$, and **nondegenerate**, meaning $g(v, v_1)\ \forall v \in V_p$ only if $v_1 = 0$. A metric is an inner product at each point in the tangent space. We sometimes denote g as ds^2, and the component expansion of g is

$$ds^2 \equiv g = \sum_{\mu,\nu} g_{\mu\nu} dx^\mu \otimes dx^\nu = \sum_{\mu,\nu} g_{\mu\nu} dx^\mu dx^\nu \tag{1.3.10}$$

Given a metric g we can always find an **orthonormal basis** v_i of the tangent space at each poin p, such that $g(v_\mu, v_\nu) = 0$ if $\mu \neq \nu$ and $g(v_\mu, v_\mu) = \pm 1$. There exists other orthonormal bases at p but the number of basis vectors with $g(v_\mu, v_\mu) = +1$ and the number with $g(v_\mu, v_\mu) = -1$ are always the same. The number of occurences of $+$ and $-$ is called the **signature** of the metric. In ordinary differential geometry the metric is usually positive definite ($+ + + + ...$, a.k.a. *Riemannian*). The metric of spacetime has signature $- + ++$ and is called *Lorentzian*. We can also view g as a linear map $V_p \to V_p^*$ via $v \to g(\cdot, v)$, thus establishing connection between vectors and dual vectors.

1.4 The Abstract Index Notation

Manipulations of high type tensors are very cumbersome, and we have seen that a tensor can be viewed in many equivalent ways, so it is important to have a simple, unambiguous notational scheme for tensor operations. The component notation introduced in section 1.3 solves these problems, but it does not make a distinction between equations that hold between tensors and equations for their components that hold in a specific basis.

We therefore use the **abstract index notation**, which is just a slight modification of the component notation. We simply replace the Greek indices with Latin indices, which serve not as basis components but as reminders of the number and type of variables the tensor acts on. For example, $T^{abc}{}_{de}$ denotes a type $(3,2)$ tensor. Note that the same letter must be used to represent the same slot on both sides of an equation.

Under this notation, we denote the contraction of a tensor by repeating the index on the contracted slots and omitting the summation. For example, $T^{abc}{}_{be}$ denotes the contraction of $T^{abc}{}_{de}$ with respect to the second contravariant and first covaraint slots. The outer product of two tensors $T^{abc}{}_{de}$ and $S^a{}_b$ is denoted by $T^{abc}{}_{de} S^f{}_g$.

Additional rules apply to the metric tensor, denoted by g_{ab}. For a vector v^a we denote the dual vector $g_{ab}v^b$ as simply v_a. The inverse metric is a type $(2,0)$ tensor $(g^{-1})^{ab}$ but we denote it as simply g^{ab}. By definition, we have $g^{ab}g_{bc} = \delta^a{}_c$ which is the identity map from V_p to V_p. In general, raising and lowering indices correspond to applying the inverse metric or metric to that slot.

We can also express symmetry properties of tensors using this notation. If a tensor T_{ab} takes a pair of vectors (v^a, w^a) into a number $T_{ab}v^a w^b$, then we can denote the tensor obtained by interchanging the order in which T_{ab} acts on the pair as T_{ba}. A **symmetric tensor** follows $T_{ab} = T_{ba}$. We also introduce a notation for the totally symmetric and totally antisymmetric parts of

tensors:

$$T_{(ab)} = \frac{1}{2}(T_{ab} + T_{ba}) \tag{1.4.1}$$

$$T_{[ab]} = \frac{1}{2}(T_{ab} - T_{ba}) \tag{1.4.2}$$

In general

$$T_{(a_1 \ldots a_l)} = \frac{1}{l!} \sum_{\pi} T_{a_{\pi(1)} \ldots a_{\pi(l)}} \tag{1.4.3}$$

$$T_{[a_1 \ldots a_l]} = \frac{1}{l!} \sum_{\pi} \delta_\pi T_{a_{\pi(1)} \ldots a_{\pi(1)}} \tag{1.4.4}$$

where the sum is taken over all permutations π and δ_π is $+1$ for even permutations and -1 for odd permutations. A totally antisymmetric type $(0, l)$ tensor field $T_{a_1 \ldots a_l} = T_{[a_1 \ldots a_l]}$ is called a **differential l-form**[4].

2 Curvature

Our intuition of curvature comes from two-dimensional surfaces embedded in $\mathbb{R}^3$ and is based on how surfaces bend in $\mathbb{R}^3$. As far as we know the spacetime manifold M is not embedded in any higher dimensional space, so we need to develop an *intrinsic* notion of curvature. This can be defined in terms of *parallel transport* (keeping a vector "pointing in the same direction" in the tangent space). On a plane, if one parallel transports a vector around any closed the path, the final vector always coincides with its initial values. This is not true on a sphere. Here we can see the idea of parallel transport can be used to characterize the curvature of any manifold.

Another way to characterize curve is through geodesics. A *geodesic* is a curve whose tangent is parallel transported along itself, i.e. it is the straightest possible curve. A space is curved if and only if some initially parallel geodesics fail to remain parallel.

There is no natural notion of parallel transport given only the manifold structure of space because the tangent spaces of two distinct vectors are different. We therefore need to relate parallel transport along a curve to taking derivatives of vector fields in the direction of the curve. We can define a vector to be parallel transported if its derivative along the curve is zero. The failure of a vector to return to its original value after parallel transported around an infinitesimal closed curve translates to the lack of commutativity of derivatives, and we can define curvature in terms of the failure of successive differentiations on tensor field to commute.

2.1 Derivative Operators and Parallel Transport

A **derivative operator** ∇, a.k.a. **covariant derivative**, on a manifold M is a map that takes each differentiable type (k, l) tensor field to a differentiable type $(k, l+1)$ tensor field. For $T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l} \in \mathcal{T}(k, l)$, we denote the tensor field resulting from acting ∇ on T by $\nabla_c T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l}$. Note that ∇_c is not a dual vector. ∇ satisfies the following properties:

(1) Linearity: $\forall A, B \in \mathcal{T}(k, l)$ and $\alpha, \beta \in \mathbb{R}$,
$\nabla_c(\alpha A^{a_1 \ldots a_k}{}_{b_1 \ldots b_l} + \beta B^{a_1 \ldots a_k}{}_{b_1 \ldots b_l}) = \alpha \nabla_c A^{a_1 \ldots a_k}{}_{b_1 \ldots b_l} + \beta \nabla_c B^{a_1 \ldots a_k}{}_{b_1 \ldots b_l}$

[4]Sometimes we denote an l-form $T_{a_1 \ldots a_l}$ as simply T when dealing strictly with differential forms.

(2) Leibnitz rule: $\forall A \in \mathcal{T}(k,l), B \in \mathcal{T}(k',l')$,
$\nabla_e[A^{a_1\ldots a_k}{}_{b_1\ldots b_l} B^{c_1\ldots c_{k'}}{}_{d_1\ldots d_{l'}}] = [\nabla_e A^{a_1\ldots a_k}{}_{b_1\ldots b_l}]B^{c_1\ldots c_{k'}}{}_{d_1\ldots d_{l'}} + A^{a_1\ldots a_k}{}_{b_1\ldots b_l}[\nabla_e B^{c_1\ldots c_{k'}}{}_{d_1\ldots d_{l'}}]$

(3) Commutativity with contraction: $\forall A \in \mathcal{T}(k,l)$,
$\nabla_d(A^{a_1\ldots c\ldots a_k}{}_{b_1\ldots c\ldots b_l}) = \nabla_d A^{a_1\ldots c\ldots a_k}{}_{b_1\ldots c\ldots b_l}$

(4) Consistency with the notion of tangent vectors as directional derivatives on scalar fields: $\forall f \in \mathcal{F}$ and $t^a \in V_p$, $t(f) = t^a \nabla_a f$

(5) Torsion free[5]: $\forall f \in \mathcal{F}$, $\nabla_a \nabla_b f = \nabla_b \nabla_a f$

The fifth condition is assumed satisfied by ∇ in GR but it is not always imposed in other theories of gravitation. The last three conditions allow us to derive an expression for the commutator of two vector fields v^a, w^b in terms of any ∇_a:

$$\begin{aligned}
[v,w](f) &= v\{w(f)\} - w\{v(f)\} \\
&= v^a\nabla_a(w^b\nabla_b f) - w^a\nabla_a(v^b\nabla_b f) \\
&= \{v^a\nabla_a w^b - w^a\nabla_a v^b\}\nabla_b f \qquad (2.1.1) \\
[v,w]^b &= v^a\nabla_a w^b - w^a\nabla_a v^b. \qquad (2.1.2)
\end{aligned}$$

To show that derivative operators exist, let ψ be a coordinate system and $\{\partial/\partial x^\mu\}$ and dx^μ be the associated coordinate bases. We may define an **ordinary derivative** ∂_a in the region covered by these coordinates. For any smooth tensor field $T^{a_1\ldots a_k}{}_{b_1\ldots b_l}$ with components $T^{\mu_1\ldots\mu_k}{}_{\nu_1\ldots\nu_l}$ in this basis, $\partial_c T^{a_1\ldots a_k}{}_{b_1\ldots b_l}$ is the tensor whose components in this basis are the partial derivatives $\partial(T^{\mu_1\ldots\mu_k}{}_{\nu_1\ldots\nu_l})/\partial x^\sigma$. All five properties follow from the properties of partial derivatives, so we can construct an associated derivative operator ∂_a for any ψ. This operator is coordinate dependent and is not naturally associated with the structure of the manifold.

Condition (4) states that any two derivative operators ∇_a and $\tilde{\nabla}_a$ must produce the same result when acting on scalar fields, but it is possible that they disagree in their action on higher rank tensors such as dual vector fields. Let w_b be a dual vector field on V_p^*, we can use conditions (2) and (4) to show that for an arbitrary scalar field f,

$$\tilde{\nabla}_a(fw_b) - \nabla_a(fw_b) = f(\tilde{\nabla}_a w_b - \nabla_a w_b) \qquad (2.1.3)$$

At point p, both $\tilde{\nabla}_a w_b$ and $\nabla_a w_b$ should depend on how w_b changes as we move away from p by the definition of a derivative, but their difference does not. Suppose w'_b is a different dual vector field that equals w_b at p, we can find smooth functions $f_{(\alpha)}$ that vanish at p and smooth dual vector fields $\mu^{(\alpha)}$ such that (similar to a component expansion): $w'_b - w_b = \sum_{\alpha=1}^n f_{(\alpha)}\mu^{(\alpha)}$. Plug this into Eq. 2.1.3 we have

$$\begin{aligned}
\tilde{\nabla}_a(w'_b - w_b) - \nabla_a(w'_b - w_b)|_p &= \sum_\alpha \{\tilde{\nabla}_a(f_{(\alpha)}\mu_b^{(\alpha)}) - \nabla_a(f_{(\alpha)}\mu_b^{(\alpha)})\} \qquad (2.1.4) \\
&= \sum_\alpha f_{(\alpha)}\{\tilde{\nabla}_a\mu_b^{(\alpha)} - \nabla_a\mu_b^{(\alpha)}\} = 0 \qquad (2.1.5)
\end{aligned}$$

[5]If this is not imposed, it can be shown that there exists a tensor $T^c{}_{ab}$ antisymmetric in a and b such that $\nabla_a\nabla_b f - \nabla_b\nabla_a f = -T^c{}_{ab}\nabla_c f$. $T^c{}_{ab}$ is called the **torsion tensor**.

since each $f_{(\alpha)} = 0$ at p. This shows $(\tilde{\nabla}_a - \nabla_a)w'_b = (\tilde{\nabla}_a - \nabla_a)w_b$ at p, which means the difference in the actions of the two operators on dual vector fields depends only on the value of w_b at p, because we were able make this difference the same for the two dual vector fields at a point by making the two fields equal at that point. Therefore, $\tilde{\nabla}_a - \nabla_a$ defines a linear map of dual vectors at p (as opposed to dual vector fields defined in the neighborhood of p) to type $(0,2)$ tensors at p. This map itself is a type $(1,2)$ tensor at p and we denote it as ${C^c}_{ab}$. We have shown that for any two derivative operators $\tilde{\nabla}_a$ and ∇_a there exists a ${C^c}_{ab}$ such that

$$\nabla_a w_b = \tilde{\nabla}_a w_b - {C^c}_{ab} w_c \tag{2.1.6}$$

This describes the possible disagreements of the actions of the two operators on dual vector fields.

Using the torsion free condition, we can show a symmetry property of ${C^c}_{ab}$. Let $w_b = \nabla_b f = \tilde{\nabla}_b f$, we have $\nabla_a \nabla_b f = \tilde{\nabla}_a \tilde{\nabla}_b f - {C^c}_{ab} \nabla_c f$. Since both $\nabla_a \nabla_b f$ and $\tilde{\nabla}_a \tilde{\nabla}_b$ are symmetric in a and b, ${C^c}_{ab}$ must also be symmetric in a and b, i.e. ${C^c}_{ab} = {C^c}_{ba}$. This is not necessarily true if we do not impose the torsion free condition.

The difference in the action of $\tilde{\nabla}_a$ and ∇_a on vector fields and all higher rank tensor fields is determined by conditions (2) and (4), and Eq. 2.1.6. For every vector field t^a and one form field w_a:

$$\begin{aligned} (\tilde{\nabla}_a - \nabla_a)(w_b t^b) &= ({C^c}_{ab} w_c) t^b + w_b (\tilde{\nabla}_a - \nabla_a) t^b = 0 \\ w_b [(\tilde{\nabla}_a - \nabla_a) t^b + {C^b}_{ac} t^c] &= 0 \\ \nabla_a t^b &= \tilde{\nabla}_a t^b + {C^b}_{ac} t^c \end{aligned} \tag{2.1.7}$$

A general formula for $T \in \mathcal{T}(k,l)$ can be derived in a similar manner:

$$\nabla_a {T^{b_1 \ldots b_k}}_{c_1 \ldots c_l} = \tilde{\nabla}_a {T^{b_1 \ldots b_k}}_{c_1 \ldots c_l} + \sum_i {C^{b_i}}_{ad} {T^{b_1 \ldots d \ldots b_k}}_{c_1 \ldots c_l} - \sum_j {C^d}_{ac_j} {T^{b_1 \ldots b_k}}_{c_1 \ldots d \ldots c_l}. \tag{2.1.8}$$

The difference between the two derivative operators is thus completely characterized by the tensor field ${C^c}_{ab}$, or **affine connection**. Since on an n-dimensional manifold ${C^c}_{ab}$ has $n^2(n+1)/2$ independent components at each point, there is a lot of freedom in the choice of a derivative operator.

In the case where $\tilde{\nabla}_a$ is an ordinary derivative ∂_a, ${C^c}_{ab}$ is called a **Christoffel symbol** and denoted ${\Gamma^c}_{ab}$:

$$\nabla_a t^b = \partial_a t^b + \Gamma^b_{ac} t^c. \tag{2.1.9}$$

If we know ${\Gamma^c}_{ab}$, we can compute ∇_a since we know how to compute an ordinary derivative. Note that ${\Gamma^c}_{ab}$ is also coordinate dependent because it is associated with ∇_a and the coordinate system we used to define ∂_a.

With the derivative operator now defined, we can say a vector v^a given at each point on a curve C with tangent vector t^a is **parallel transported** as one moves along C if

$$t^a \nabla_a v^b = 0 \tag{2.1.10}$$

is satisfied along C. This can be generalized to a tensor of arbitrary rank simply by replacing v^b with the tensor. In terms of components in a coordinate basis and the parameter t along the curve, this equation becomes

$$\frac{dv^\nu}{dt} + \sum_{\mu,\lambda} t^\mu {\Gamma^\nu}_{\mu\lambda} v^\lambda = 0. \tag{2.1.11}$$

This shows the parallel transport of v^a depends only on its value along the curve. Also, it follows from properties of ODEs that Eq. 2.1.11 always has a unique solution for any given initial value of v^a, so a vector at point p uniquely defines a parallel transported vector everywhere on the curve. Thus, given a derivative operator and a curve connecting p and q, we can map into each other the vector spaces associated with the two points. This mathematical structure is called a **connection**.

Although there are many possible derivative operators on the manifold, there is a unique definition of parallel transport that preserves inner products of all pairs of vectors given a metric of any signature. Let us require that for two vectors v^a and w^a, their inner product remain unchanged when they are parallel transported along any curve, i.e. $t^a\nabla_a(g_{bc}v^bw^c) = 0$ for v^b and w^c satisfying Eq. 2.1.10, then we obtain $t^av^bw^c\nabla_a g_{bc} = 0$. Since we require this to hold for any curve and vectors, we get

$$\nabla_a g_{bc} = 0 \tag{2.1.12}$$

This is the **metric compatibility** condition, and it uniquely determines ∇_a. A proof is given on page 35 in Wald. In terms of ordinary derivative, this unique choice is given by

$$\Gamma^c{}_{ab} = \frac{1}{2}g^{cd}\{\partial_a g_{bd} + \partial_b g_{ad} - \partial_d g_{ab}\} \tag{2.1.13}$$

In component form, this is

$$\Gamma^\rho{}_{\mu\nu} = \frac{1}{2}\sum_\sigma g^{\rho\sigma}\left\{\frac{\partial g_{\nu\sigma}}{\partial x^\nu} + \frac{\partial g_{\mu\sigma}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma}\right\} \tag{2.1.14}$$

2.2 Curvature

As mentioned before, we want to use the path dependence of parallel transport to define the intrinsic notion of curvature. Let ∇_a be a derivative operator, w_a be a dual vector field, and f be smooth function. Again using Leibnitz rule, we obtain

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)(fw_c) = f(\nabla_a\nabla_b - \nabla_b\nabla_a)w_c. \tag{2.2.1}$$

By the same reasoning following Eq. 2.1.3 in the previous section, we have that the tensor $(\nabla_a\nabla_b - \nabla_b\nabla_a)w_c$ at p depends only on the value of w_c at p. Thus $(\nabla_a\nabla_b - \nabla_b\nabla_a)$ defines a type $(1,3)$ tensor, i.e. a linear map from dual vectors at p to type $(0,3)$ tensors at p, and there exists a tensor field $R_{abc}{}^d$ such that for all w_c,

$$\nabla_a\nabla_b w_c - \nabla_b\nabla_a w_c = R_{abc}{}^d w_d. \tag{2.2.2}$$

$R_{abc}{}^d$ is called the **Riemann curvature tensor**.

$R_{abc}{}^d$ is related to the failure of a vector to return to its original value when parallel transported around a closed curve. Let us consider a surface S with coordinates t and s that contains the point p (chosen to be $(0,0)$). We construct a loop by moving Δt along $s = 0$, followed by moving Δs along $t = \Delta t$ and then moving back by Δt and Δs. Let v^a be a vector at p that is not necessarily tangent to S and parallel transport v^a around this loop. To compute the change in v^a when it returns to p, we can choose an arbitrary dual vector field w_a and find the change in the scalar v^aw_a

around the loop. For small Δt, the change in the scalar in the first leg of the curve is

$$\begin{aligned}\delta_1 &= \Delta t \frac{\partial}{\partial t}(v^a w_a)\big|_{(\Delta t/2,0)} \\ &= \Delta t T^b \nabla_b (v^a w_a)\big|_{(\Delta t/2,0)} \\ &= \Delta t v^a T^b \nabla_b w_a\big|_{(\Delta t/2,0)} \end{aligned} \tag{2.2.3}$$

where T^b is the tangent to the curves of constant s and $T^b \nabla_b v^a = 0$ by Eq. 2.1.10. This expression is similar for other parts of the curve, and the parts corresponding to the Δt variations are

$$\delta_1 + \delta_3 = \Delta t \{ v^a T^b \nabla_b w_a\big|_{(\Delta t/2,0)} - v^a T^b \nabla_b w_a\big|_{(\Delta t/2,\Delta s)} \} \tag{2.2.4}$$

As $\Delta s \to 0$, the term in brackets goes to zero, so the total change in $v^a w_a$, as well as the total change in v^a, vanishes. This means parallel transport is independent of path to *first order* in Δt and Δs. This means v^a at $(\Delta t/2, \Delta s)$ equals the parallel transport of v^a at $(\Delta t/2, 0)$ along the curve $t = \Delta t/2$ to first order in Δs. However, to first order the term $T^b \nabla_b w_a$ at $(\Delta t/2, \Delta s)$ will differ from its parallel transport from $(\Delta t/2, 0)$ by $\Delta s S^c \nabla_c (T^b \nabla_b w_a)$, where S^c is the tangent to the curves of constant t. Therefore, the total change to second order in Δt and Δs is

$$\delta_1 + \delta_3 = -\Delta t \Delta s \; v^a S^c \nabla_c (T^b \nabla_b w_a). \tag{2.2.5}$$

We evaluate all tensors at p and include all parts of the curve. Using commutativity of the coordinate vector fields T^a and S^b, we find the total changes in $v^a w_a$ and v^a are

$$\begin{aligned}\delta(v^a w_a) &= \Delta t \Delta s v^a \{ T^c \nabla_c (S^b \nabla_b w_a) - S^c \nabla_c (T^b \nabla_b w_a) \} \\ &= \Delta t \Delta s \; v^a T^c S^b (\nabla_c \nabla_b - \nabla_b \nabla_c) w_a \\ &= \Delta t \Delta s \; v^a T^c S^b R_{cba}{}^d w_d \qquad (2.2.6) \\ \delta v^a &= \Delta t \Delta s \; v^d T^c S^b R_{cbd}{}^a. \qquad (2.2.7)\end{aligned}$$

Analogous to our derivation of Eq. 2.1.7, we can show that for a vector field t^a and

$$(\nabla_a \nabla_b - \nabla_b \nabla_a) t^c = -R_{abd}{}^c t^d, \tag{2.2.8}$$

and for an arbitrary tensor field $T^{c_1 \ldots c_k}{}_{d_1 \ldots d_l}$

$$(\nabla_a \nabla_b - \nabla_b \nabla_a) T^{c_1 \ldots c_k}{}_{d_1 \ldots d_l} = -\sum_{i=1}^{k} R_{abe}{}^{c_i} T^{c_1 \ldots e \ldots c_k}{}_{d_1 \ldots d_l} + \sum_{j=1}^{l} R_{abd_j}{}^e T^{c_1 \ldots c_k}{}_{d_1 \ldots e \ldots d_l}. \tag{2.2.9}$$

Four key properties of the Riemann tensor are

(1) $R_{abc}{}^d = -R_{bac}{}^d$.

(2) $R_{[abc]}{}^d = 0$.

(3) For the metric compatible derivative ∇_a, i.e. $\nabla_a g_{bc} = 0$, $R_{abcd} = -R_{abdc}$.

(4) The Bianchi identity: $\nabla_{[a} R_{bc]d}{}^e = 0$.

Properties (1)-(3) also lead to a symmetry property of $R_{abc}{}^{d}$: $R_{abcd} = R_{cdab}$. Full proofs of these properties can be found on page 39 of Wald. To summarize: (1) follows from the definition of $R_{abc}{}^{d}$; (2) comes from $R_{[abc]}{}^{d} = 2\nabla_{[a}\nabla_{b}w_{c]} = 0$, which can be proven from Eq. 2.1.8 in the case $\tilde{\nabla}_a = \partial_a$; (3) follows from Eq. 2.2.9 applied to g_{ab}; (4) can be derived from Eq. 2.2.9 and the commutator of derivative operators.

The Riemann tensor can be decomposed into a "trace part" and a "trace free part." The trace of the Riemann tensor over its first two or last two indices is zero by properties (1) and (3). We define the **Ricci tensor** R_{ac} as the trace of $R_{abc}{}^{d}$ over its second and fourth indices (or equivalently the first and the third)

$$R_{ac} = R_{abc}{}^{b} \tag{2.2.10}$$

R_{ac} is symmetric by the symmetric property of the Riemann tensor. The trace of the Ricci tensor is the **scalar curvature** $R = R_{a}{}^{a}$. The trace free part is called the **Weyl tensor**[6] C_{abcd}, defined for manifolds of dimension $n \geq 3$ by

$$R_{abcd} = C_{abcd} + \frac{2}{n-2}(g_{a[c}R_{d]b} - g_{b[c}R_{d]a}) - \frac{2}{(n-1)(n-2)}Rg_{a[c}g_{d]b}. \tag{2.2.11}$$

C_{abcd} satisfies the properties (1)-(3) and is trace free on all of its indices.

By contracting over the indices a and e, the Bianchi identity becomes

$$\nabla_a R_{bcd}{}^{a} + \nabla_b R_{cd} - \nabla_c R_{bd} = 0. \tag{2.2.12}$$

Raising d and contracting over b and d, we get

$$\nabla_a R_{c}{}^{a} + \nabla_b R_{c}{}^{b} - \nabla_c R = 0 \tag{2.2.13}$$

or $\nabla^a G_{ab} = 0$, where

$$G_{ab} = R_{ab} - \frac{1}{2}Rg_{ab} \tag{2.2.14}$$

is the **Einstein tensor**.

2.3 Geodesics

Geodesics are the straightest possible lines we can draw on a curved geometry, and we define a **geodesic** as a curve whose tangent vector T^a is parallel transported along itself, i.e.

$$T^a\nabla_a T^b = 0. \tag{2.3.1}$$

This is actually a stronger condition than what we need for the curve to satisfy the requirement of being the straightest possible line because it requires the tangent vector to both point in the same direction and maintain the same length when parallel transported. We can drop the second requirement to get a weaker condition

$$T^a\nabla_a T^b = \alpha T^b \tag{2.3.2}$$

It can be shown that we can always reparameterize this curve so that it satisfies the stronger condition, and this is called **affine parameterization**. We require a geodesic to be affinely parameterized.

[6] It is sometimes called the **conformal tensor.**

If we map the geodesic into a curve $x^\mu(t)$ in $\mathbb{R}^n$ using a coordinate system ψ, the component form of the **geodesic equation** is

$$\frac{dT^\mu}{dt} + \sum_{\sigma,\nu} \Gamma^\mu{}_{\sigma\nu} T^\sigma T^\nu = \frac{d^2x^\mu}{dt^2} + \sum_{\sigma,\nu} \Gamma^\mu{}_{\sigma\nu} \frac{dx^\sigma}{dt}\frac{dx^\nu}{dt} = 0 \tag{2.3.3}$$

This includes n coupled second order ODEs, and a unique solution exists for any initial x^μ and dx^μ/dt. Thus given $p \in M$ and $T^a \in V_p$, there always exists a unique geodesic through p with T^a. This allows us to construct some convenient coordinate systems.

For $p \in M$, we define the **exponential map** $V_p \to M$ that maps $T^a \in V_p$ to a point in M lying at unit affine parameter from p along the geodesic through p with tangent T^a. There always exists a sufficiently small neighborhood of the origin of V_p on which the exponential map is one-to-one. The dimension of V_p is n so we can identify it with $\mathbb{R}^n$, and use the exponential map to give us a coordinate system called **Riemannian normal coordinates** at p. In these coordinates all geodesics through p are mapped into straight lines through the origin of $\mathbb{R}^n$. Eq. 2.3.3 tells us the components of $\Gamma^\mu{}_{\sigma\nu}$ are zero at p, and this makes the Riemannian normal coordinates useful for calculations at a given point.

Let S be a hypersurface, i.e. an $(n-1)$ dimensional submanifold embedded in n dimensional M. At each point $p \in S$, we can view the tangent space $\tilde{V}_p$ of S as an $(n-1)$ dimensional subspace of V_p of M. A vector $n^a \in V_p$ will be orthogonal to all vectors in $\tilde{V}_p$ with respect to the metric g_{ab} and we say it is normal to S. If S is not a null hypersurface[7], we normalize n^a with $g_{ab}n^a n^b = \pm 1$. In this case we can use the **Gaussian normal coordinates**, or **synchronous coordinates**, given a a metric-compatible ∇_a. For each $p \in S$ we construct a unique geodesic through p with tangent n^a. We choose arbitrary coordinates $(x^1, ..., x^{n-1})$ in a small portion of S around p, and label points in the neighborhood by these coordinates and the parameter t of the geodesic on which the point lies. This defines a chart in a sufficiently small neighborhood of p.

A property of the Gaussian normal coordinates is the geodesics remain orthogonal to all hypersurfaces S_t defined by a constant t. We can prove this by showing the tangent field n^a of the geodesic remains orthogonal to all coordinate basis fields X_i^a that generate the tangent space to S_t. Note that n^a and X^b commute because they are elements of a coordinate basis on M:

$$\begin{aligned} n^b\nabla_b(n_a X^a) &= n_a n^b \nabla_b X^a + \cancel{X^a n^b \nabla_b n_a} \\ &= n_a X^b \nabla_b n^a = \frac{1}{2} X^b \nabla_b (n^a n_a) = 0. \end{aligned} \tag{2.3.4}$$

This shows that $n_a X^a$ remains zero.

Geodesics of a metric-compatible derivative operator the length of curves connecting two points as measured by the metric. For a differentiable curve C with tangent T^a on a manifold M with Riemannian metric g_{ab}, we define the length of C as

$$l = \int (g_{ab}T^a T^b)^{1/2} dt. \tag{2.3.5}$$

For a Lorentzian metric with signature $-+ + ... +$, a curve is said to be: *timelike* if $g_{ab}T^aT^b < 0$; *null* if $g_{ab}T^aT^b = 0$; and *spacelike* if $g_{ab}T^aT^b > 0$. The length of spacelike curves the length is

[7] In the case of a Riemannian metric, n^a cannot lie in $\tilde{V}_p$; in the case of a metric of indefinite signature, n^a could be a null vector ($g_{ab}n^a n^b = 0$) and lie in $\tilde{V}_p$, and we say S is a **null hypersurface** at p.

defined the same way as in Eq. 2.3.5. For null curves the length is zero. For timelike curves, we define the **proper time**:

$$\tau = \int (-g_{ab}T^aT^b)^{1/2}dt. \tag{2.3.6}$$

Note that the length or proper time does not depend on parameterization, and the length of curves which change from timelike to spacelike is not defined. Since the tangent of a geodesic is parallel transported with constant norm, the geodesic cannot change from timelike to spacelike or null in a Lorentz manifold.

We can derive the condition for a curve C that extremizes the length between its endpoints $p = C(a)$ and $q = C(b)$, meaning that the length of the curve does not change to first order under arbitrary smooth deformation that keeps the endpoints fixed. We will consider a spacelike curve and work in $\mathbb{R}^n$ with a chart. In a coordinate basis, Eq. 2.3.5 becomes

$$l = \int_a^b \left[\sum_{\mu,\nu} g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}\right]^{1/2} dt. \tag{2.3.7}$$

We will use variations to extremize l in the same way we vary the action in Lagrangian mechanics. Assuming the curve is parameterized so that $g_{ab}T^aT^b = 1$, the extremization condition is

$$0 = \int_a^b \sum_{\alpha,\beta}\left\{-\frac{d}{dt}\left(g_{\alpha\beta}\frac{dx^\alpha}{dt}\right) + \frac{1}{2}\sum_\lambda \frac{\partial g_{\alpha\lambda}}{\partial x^\beta}\frac{dx^\alpha}{dt}\frac{dx^\lambda}{dt}\right\}\delta x^\beta dt \tag{2.3.8}$$

Note that δx^β vanishes at the endpoints. It can be shown this equation holds for arbitrary δx^β if an only if the geodesic equation (Eq. 2.3.3) holds. Thus, a curve extremizes the length between its endpoints if and only if it is a geodesic. This derivation can be applied to the proper time of a timelike curve as well, and it also shows Eq. 2.3.3 can be obtained from the variations of the Lagrangian

$$L = \sum_{\mu,\nu} g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}. \tag{2.3.9}$$

In many cases, the most efficient way to compute the Christoffel symbol to start with the Lagrangian, write down the Euler-Lagrange equations, and read off $Gamma^\mu_{\sigma\nu}$ by comparing with Eq. 2.3.3.

There are always curves of arbitrary lengths (with a lower bound) connecting two points on a Riemannian manifold, and the shortest path between two points is always a "straightest possible path." However, a given geodesic between two points is not necessarily the shortest path: in a Lorentzian manifold, we can always find timelike curves of arbitrarily small proper time for two points that can be connected by a timelike curve. If a curve of greatest proper time exists, it must be a timelike geodesic.

We can now study the relation between the curvature of the manifold and the tendency for geodesics to accelerate toward or away from each other. Let $\gamma_s(t)$ be a smooth one-parameter family of geodesics. For each $s \in \mathbb{R}$, γ_s is a geodesic parameterized by affine parameter t; and the map $(t, s) \to \gamma_s(t)$ is smooth, one-to-one, and has smooth inverse. Let Σ be the two-dimensional submanifold spanned by the curves $\gamma_s(t)$. The tangent vector field $T^a = (\partial/\partial t)^a$ satisfies Eq. 2.3.1. The vector field $X^a = (\partial/\partial s)^a$ is the **deviation vector** and represents the displacement to an infinitesimally close geodesic. X^a has a "gauge freedom" in the sense that it changes by adding

a multiple of T^a under a reparameterization $t \to t' = b(s)t + c(s)$, and we can set $X^a T_a = 0$ by choosing the appropriate parameterization. Since T^a and X^a are coordinate vector fields (we chose t and s to be the coordinates), they commute:

$$T^b \nabla_b X^a = X^b \nabla_b T^a; \tag{2.3.10}$$

so $X^a T_a$ is constant along each geodesic by the same argument as in Eq. 2.3.4.

The relative velocity between geodesics $v^a = T^b \nabla_b X^a$ is the rate of change of X^a along a geodesic. Similarly, the relative acceleration is

$$\begin{aligned} a^a &= T^c \nabla_c v^a = T^c \nabla_c (T^b \nabla_b X^a) && (2.3.11) \\ &= T^c \nabla_c v^a = T^c \nabla_c (X^b \nabla_b T^a) \\ &= (T^c \nabla_c X^b)(\nabla_b T^a) + X^b T^c \nabla_c \nabla_b T^a \\ &= (X^c \nabla_c T^b)(\nabla_b T^a) + X^b T^c \nabla_c \nabla_b T^a - R_{cbd}{}^a X^b T^c T^d \\ &= X^c \nabla_c (T^b \nabla_b T^a) - R_{cbd}{}^a X^b T^c T^d \\ &= -R_{cbd}{}^a X^b T^c T^d && (2.3.12) \end{aligned}$$

where we used the "anti-Leibnitz" rule in the second to last line. Eq. 2.3.11 is the **geodesic deviation equation**. It states that the geodesics will deviate from each other if and only if $R_{abc}{}^d \neq 0$.

2.4 Methods for Computing Curvature

Previously we defined the Riemann tensor in section 2.2 simply by pointing out that there must exist such a tensor that describes the difference in action of the operators $\nabla_a \nabla_b$ and $\nabla_b \nabla_a$ on dual vector fields. However, we do not know how to calculate $R_{abc}{}^d$, and we will discuss methods for calculating $R_{abc}{}^d$ in this section.

2.4.1 Coordinate component method

We start by choosing a coordinate system and expressing ∇_a in terms of ∂_a and the Christoffel symbol: $\nabla_b w_c = \partial_b w_c - \Gamma^d{}_{bc} w_d$. Plugging this into Eq. 2.2.2, we get

$$R_{abc}{}^d w_d = [-2\partial_{[a}\Gamma^d{}_{b]c} + 2\Gamma^e{}_{c[a}\Gamma^d{}_{b]e}] w_d. \tag{2.4.1}$$

This holds for all w_d so we can drop w_d to get an expression for $R_{abc}{}^d$. In component form, this is

$$R^\sigma_{\mu\nu\rho} = \frac{\partial}{\partial x^\nu}\Gamma^\sigma{}_{\mu\rho} - \frac{\partial}{\partial x^\mu}\Gamma^\sigma{}_{\nu\rho} + \sum_\alpha (\Gamma^\alpha{}_{\mu\rho}\Gamma^\sigma{}_{\alpha\nu} - \Gamma^\alpha{}_{\nu\rho}\Gamma^\sigma{}_{\alpha\mu}). \tag{2.4.2}$$

We can then plug in values of $\Gamma^\sigma{}_{\mu\nu}$ calculated through any methods we introduced earlier.

We will discuss some useful facts about calculations in coordinate bases. We can write the components of the metric $g_{\mu\nu}$ as a matrix $(g_{\mu\nu})$. We define g to be the determinant of $(g_{\mu\nu})$

$$g = \det(g_{\mu\nu}) \tag{2.4.3}$$

so the natural volume element on the manifold induced by g_{ab} is $\sqrt{|g|}dx^1...dx^n$.

We can use Eq. 2.1.14 to derive the contracted Christoffel symbol $\Gamma^a{}_{ab}$

$$\Gamma^a{}_{a\mu} = \sum_\nu \Gamma^\nu{}_{\nu\mu} = \frac{1}{2}\sum_{\nu,\alpha} g^{\nu\alpha}\frac{\partial g_{\nu\alpha}}{\partial x^\mu} = \frac{1}{2g}\frac{\partial g}{\partial x^\mu} = \frac{\partial}{\partial x^\mu}\ln\sqrt{|g|}. \tag{2.4.4}$$

This appears in the component form of the Ricci tensor, as well as the divergence of any vector field T^a:

$$\nabla_a T^a = \partial_a T^a + \Gamma^a{}_{ab}T^b = \sum_\mu \frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x^\mu}(\sqrt{|g|}T^\mu). \tag{2.4.5}$$

2.4.2 Orthonormal basis (tetrad) methods

A coordinate basis $\{\partial/\partial x^\mu\}$ is not orthonormal except in the case of flat spacetime in Cartesian coordinates, so we introduce a "nonholonomic" (noncoordinate) orthonormal basis of smooth vector fields $(e_\mu)^a$ satisfying

$$(e_\mu)^a(e_\nu)_a = \eta_{\mu\nu} = \mathrm{diag}(-1,...,-1,1,...,1). \tag{2.4.6}$$

In four dimensions, $\{(e_\mu)^a\}$ is called a **tetrad**. Eq. 2.4.6 implies

$$\sum_{\mu,\nu}\eta^{\mu\nu}(e_\mu)^a(e_\nu)_b = \delta^a{}_b \tag{2.4.7}$$

where $\eta^{\mu\nu} = (\eta_{\mu\nu})^{-1} = \eta_{\mu\nu}$.

There are three key requirements for the tetrad methods: (1) the derivative operator ∇_a is metric compatible; (2) ∇_a is torsion free; and (3) the Riemann tensor is related to ∇_a by Eq. 2.2.2. In the coordinate basis methods, (2) is expresses by the symmetric property of $C^c{}_{ab}$, (1) is given by Eq. 2.1.13, and (3) is expressed by Eq. 2.4.2.

We begin by defining the **connection 1-forms**, $w_{a\mu\nu}$:

$$w_{a\mu\nu} = (e_\mu)^b\nabla_a(e_\nu)_b. \tag{2.4.8}$$

The components of $w_{a\mu\nu}$ are called the **Ricci rotation coefficients**:

$$w_{\lambda\mu\nu} = (e_\lambda)^a(e_\mu)^b\nabla_a(e_\nu)_b. \tag{2.4.9}$$

Using the metric compatibility condition, the orthonormality of $\{(e_\mu)^a\}$ implies:

$$w_{a\mu\nu} = (e_\mu)^b\nabla_a(e_\nu)_b = -(e_\nu)^b\nabla_a(e_\mu)_b = -w_{a\nu\mu}. \tag{2.4.10}$$

This is the expression for (1) in the orthonormal basis approach. Note that the antisymmetry of the Ricci rotation coefficients means it has $n^2(n-1)/2$ ($= 24$ when $n = 4$) independent components, while the Christoffel symbol has $n^2(n+1)/2$ ($= 40$ when $n = 4$) components.

The components of R_{abcd} in the orthonormal basis are

$$R_{\rho\sigma\mu\nu} = R_{abcd}(e_\rho)^a(e_\sigma)^b(e_\mu)^c(e_\nu)^d = (e_\rho)^a(e_\sigma)^b(e_\mu)^c(\nabla_a\nabla_b - \nabla_b\nabla_a)(e_\nu)_c.' \tag{2.4.11}$$

However, we have the following

$$\begin{aligned}(e_\mu)^c\nabla_a\nabla_b(e_\nu)^c &= \nabla_a\{(e_\mu)^c\nabla_b(e_\nu)_c\} - [\nabla_a(e_\mu)^c][\nabla_b(e_\nu)_c] \\ &= \nabla_a\{(e_\mu)^c\nabla_b(e_\nu)_c\} - [\nabla_a(e_\mu)^f]\delta^c{}_f[\nabla_b(e_\nu)_c] \\ &= \nabla_a\{(e_\mu)^c\nabla_b(e_\nu)_c\} - \sum_{\alpha,\beta}\eta^{\alpha\beta}[\nabla_a(e_\mu)^f](e_\alpha)^c(e_\beta)_f[\nabla_b(e_\nu)_c]. \end{aligned} \tag{2.4.12}$$

From the definition of connection 1-forms, we obtain

$$R_{\rho\sigma\mu\nu} = (e_\rho)^a(e_\sigma)^b\{\nabla_a w_{b\mu\nu} - \nabla_b w_{a\mu\nu} - \sum_{\alpha,\beta}\eta^{\alpha\beta}[w_{a\beta\mu}w_{b\alpha\nu} - w_{b\beta\mu}w_{a\alpha\nu}]\} \quad (2.4.13)$$

$$= (e_\rho)^a\nabla_a w_{\sigma\mu\nu} - (e_\sigma)^a\nabla_a w_{\rho\mu\nu} - \sum_{\alpha,\beta}\eta^{\alpha\beta}[w_{\rho\beta\mu}w_{\sigma\alpha\nu} - w_{\sigma\beta\mu}w_{\rho\alpha\nu} + w_{\rho\beta\sigma}w_{\alpha\mu\nu} - w_{\sigma\beta\rho}w_{\alpha\mu\nu}], \quad (2.4.14)$$

where the last two terms compensate for taking the components of $w_{a\mu\nu}$ inside the derivative in the first two terms, and we can replace the derivatives ∇_a by ∂_a because they act on scalars. This expresses requirement (3). The Ricci tensor is $R_{\rho\nu} = \eta^{\sigma\nu}R_{\rho\sigma\mu\nu}$.

To express requirement (2), we can use two approaches. The first one involves noting that Eq. 2.1.1 holds for all vector fields in a basis if and only ∇_a is torsion free. We can then express (2) by the **commutation relations** of the basis vector fields

$$(e_\sigma)_a[e_\mu, e_\nu]^a = (e_\sigma)_a\{(e_\mu)^b\nabla_b(e_\nu)^a - (e_\nu)^b\nabla_b(e_\mu)^a\} = w_{\mu\sigma\nu} - w_{\nu\sigma\mu}. \quad (2.4.15)$$

This gives the $n^2(n-1)/2$ equations needed to solve for $w_{\sigma\mu\nu}$.

The second approach is to realize that from the definition of connection 1-forms we have

$$\nabla_{[a}(e_\sigma)_{b]} = \sum_{\mu,\nu}(e_\mu)_{[a}w_{b]\sigma\nu}. \quad (2.4.16)$$

We can replace ∇ with ∂ here because the torsion fee condition implies that the antisymmetrized derivative of a 1-form is independent of derivative operator. The converse is also true.

Eq. 2.4.13 and 2.4.16 can be expressed in the notations of differential forms by dropping the dual index a and use boldface letters to designate forms:

$$\mathbf{w}_\nu{}^\mu = \sum_\sigma \eta^{\mu\sigma}\mathbf{w}_{\nu\sigma}. \quad (2.4.17)$$

Then Eq. 2.4.13 and 2.4.16 become

$$\mathbf{R}_\mu{}^\nu = d\mathbf{w}_\mu{}^\nu + \sum_\alpha \mathbf{w}_\mu{}^\alpha \wedge \mathbf{w}_\alpha{}^\nu \quad (2.4.18)$$

$$d\mathbf{e}_\sigma = \sum_\mu e_\mu \wedge \mathbf{w}_\sigma{}^\mu \quad (2.4.19)$$

These are sometimes referred to as the **equations of structure**. In addition to the methods above, there is a third "null tetrad" method by Newman and Penrose for calculating the curvature, and we will not introduce it here.

3 Einstein's Equation

3.1 The Geometry of Space in Prerelativity Physics; General and Special Covariance

In prerelativity physics space is assumed to have the manifold structure of $\mathbb{R}^3$ with **Cartesian coordinates**. Many such rigid grid systems are possible and can be put into one-to-one correspondence with elements of the six-parameter group of rotations and translations of $\mathbb{R}^3$. Thus the

Cartesian coordinates (x^1, x^2, x^3) of a point in space do not have any intrinsic meaning. However, the distance between two points x and $\bar{x}$, defined in terms of Cartesian coordiantes by $D^2 = (x^1 - \bar{x}^1)^2 + (x^2 - \bar{x}^2)^2 + (x^3 - \bar{x}^3)^2$, is independent of the choice of Cartesian coordinate system and describes an intrinsic property of space. This definition of distance gives rise to a metric:

$$ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2, \tag{3.1.1}$$

or in the index notation (in the Cartesian coordinate basis):

$$h_{ab} = \sum_{\mu,\nu} h_{\mu\nu}(dx^\mu)_a(dx^\nu)_b, \tag{3.1.2}$$

with $h_{\mu\nu} = \text{diag}(1, 1, 1)$. This definition is independent of the choice of Cartesian coordinate system. Since the components of h_{ab} are constants, the ordinary derivative ∂_a satisfies $\partial_a h_{bc} = 0$, and it is metric compatible. The Christoffel symbols therefore vanish for this coordinate system. Also note that ordinary derivatives commute on all tensors, the curvature (Eq. 2.2.2) vanishes, i.e. h_{ab} is flat. Thus the prerelativity assumptions have led to the conclusion that *space is the manifold* $\mathbb{R}^3$ *with a flat Riemannian metric defined on it* (the converse is also true).

All physics experiments measure numbers, so all physical quantities must be reducible to numbers, thus tensor fields (maps from vector/dual vectors to numbers) encompass a wide range of quantities. An important principle applies to the form of laws of physics in any descriptions of space and is the motivation for SR and GR. the principle of **general covariance** states that the metric of space is the only quantity *pertaining to space* that can appear in any laws of physics. The italicized part, however, is not precisely defined.

To give an example of how this principle is violated. Consider a preferred vector field v^a, by which we mean it is possible to choose a coordinate system such that $v^a = (\partial/\partial x^1)^a$. This vector therefore is a quantity *pertaining to space*. If we write out an equation of physics *without explicitly incorporating* v^a *into the equation* but rather substitute it with the components $v^\mu = (1, 0, 0, ..., 0)$, the form of the equation would not be preserved when we make a coordinate transformation which would invalidate $v^a = (\partial/\partial x^1)^a$. We can also conclude that these equations are not tensor equations because the components fail to transform according to the tensor transformation law (Eq. 1.3.9). An implication of the principle of general covariance is that the Christoffel symbol $\Gamma^c{}_{ab}$ cannot appear in any laws of physics, because it is equivalent to specifying an ordinary derivative ∂_a, which is an additional geometric quantity pertaining to space not derivable from the metric (unless ∂_a coincides with the metric compatible derivative and $\Gamma^c{}_{ab}$ vanishes). In the other viewpoint, $\Gamma^c{}_{ab}$ cannot appear because it does not transform according to Eq. 1.3.9 under general coordinate transformations.

The metric of space has a nontrivial number of isometries, including the six parameter group of translations and rotations of $\mathbb{R}^3$ and the discrete parity symmetry. Consider a family of observers O and another family of observers O' obtained by acting on O with an isometry (distance preserving transformations between metric spaces). The principle of **special covariance** states that any physically possible set of measurements obtained by O is also a physically possible set for O'. This principle implies the existence of an action of the isometry group on the state of the physical fields being measured. It is closely related to the principle of general covariance. Suppose a physical object is described by a tensor field $T^{a\ldots}{}_{b\ldots}$. General covariance requires that the equations governing $T^{a\ldots}{}_{b\ldots}$ involve only $T^{a\ldots}{}_{b\ldots}$, the metric h_{ab}, and quantities determined by the metric such as the derivative operator. All quantities measurable by O can be expressed as scalars resulting

from contracting $T^{a\cdots}{}_{b\cdots}$ and its derivatives with O's basis vector fields $(e_\alpha)^a$. These assumptions will lead to special covariance. Therefore we can view general covariance as a formulation of the idea of special covariance in the absence of isometries.

Special covariance of the laws of physics for tensor fields implies that given a coordinate system, if we write out the coordinate component equations *without explicitly incorporating the metric* but rather with its components $h_{\mu\nu}$, then the equations will be preserved under a special group of coordinate transformations corresponding to the isometries.

3.2 Special Relativity

In special relativity, spacetime is assumed to have the manifold structure of $\mathbb{R}^4$ and there exist preferred families of motion in spacetime ("inertial" or "nonaccelerating" motions). It is also assumed that each event can be specified by the coordinates (t, x^1, x^2, x^3), called the **global inertial coordinate system**. Many such systems exist[8], so the labels do not have intrinsic meaning, as discussed at the beginning of the previous section. Similar to the distance in $\mathbb{R}^3$, the spacetime interval $I = -(x^0 - \bar{x}^0)^2 + (x^1 - \bar{x}^1)^2 + (x^2 - \bar{x}^2)^2 + (x^3 - \bar{x}^3)^2$ is the same for all global inertial coordinate systems and can be viewed as an intrinsic property of spacetime, and we define the **metric of spacetime** by

$$\eta_{ab} = \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu}(dx^\mu)_a(dx^\nu)_b \tag{3.2.1}$$

with $\eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$. Again the ordinary derivative ∂_a is metric compatible and is the derivative operator associated with η_{ab}. The commutativity of ∂_a implies zero curvature for η_{ab}. Thus, SR asserts that *spacetime is the manifold* $\mathbb{R}^4$ *with as flat metric of Lorentz signature* (and the converse). The principle of general covariance applies to SR mostly in the same way but with one modification. Two further aspects, the space orientation and time orientation of spacetime, can appear in physical laws[9].

SR asserts that the paths in spacetime of material particles are always timelike curves, i.e. nothing travels faster than light. Timelike curves can be parameterized by the proper time τ as defined in Eq. 2.3.6 with g_{ab} replaced by η_{ab}. According to SR, τ is the time that would elapse on a clock carried along the given curve. The maximum elapsed time between two events is given by the geodesic (i.e. inertial) motion.

The tangent vector u^a to a timelike curve parameterized by τ is called the **4-velocity** of the curve and it has unit length, i.e. $u^a u_a = -1$. With no external forces, a particle will travel on a geodesic, i.e. $u^a \partial_a u^b = 0$. The **4-momentum** of a particle with rest mass m is $p^a = mu^a$, and the **energy** (recognized as the time component of p^a) as measured by an observer with 4-velocity v^a at the site of the particle is $E = -p_a v^a$. We may define the energy as measured by a distance observer to be the energy measured by an observer at the site with 4-velocity parallel to that of the distant observer because parallel transport is path independent in a flat spacetime.

Continuous matter distributions are described by the symmetric **stress-energy-momentum tensor** T_{ab}. For an observer with 4-velocity v^a, the component $T_{ab}v^a v^b$ is interpreted as energy density and is nonnegative for normal matter. If x^a is orthogonal to v^a, we interpret the component $-T_{ab}v^a x^b$ as the momentum density of the matter in the x^a direction. If y^a is also orthogonal to v^a, then $T_{ab}x^a y^b$ is the x^a-y^a component of the stress tensor as defined in section 1.3.

[8] They can be put into one-to-one correspondence with elements of the 10-parameter Poincaré group.

[9] For details of these aspects, see page 60 of Wald.

A **perfect fluid** is defined to be a continuous distribution of matter with stress-energy tensor:

$$T_{ab} = \rho u_a u_b + P(\eta_{ab} + u_a u_b), \tag{3.2.2}$$

where u^a represents the 4-velocity of the fluid. The functions ρ and P are the mass-energy density and pressure respectively, as measured in the rest frame. This fluid is perfect in the sense that there is no terms describing heat conduction or viscosity. The equation of motion of a perfect fluid subject to no external forces is

$$\partial^a T_{ab} = 0. \tag{3.2.3}$$

In terms of ρ, P and u^a, and projecting the equation parallel and perpendicular to u^b, we get

$$u^a \partial_a \rho + (\rho + P)\partial^a u_a = 0 \tag{3.2.4}$$

$$(P + \rho)u^a \partial_a u_b + (\eta_{ab} + u_a u_b)\partial^a P = 0 \tag{3.2.5}$$

In the nonrelativistic limit, $P \ll \rho$, $u^\mu = (1, \vec{v})$, and $v\frac{dP}{dt} \ll |\vec{\nabla} P|$; and these equations become

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{v}) = 0, \tag{3.2.6}$$

$$\rho \left\{ \frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \vec{\nabla})\vec{v} \right\} = -\vec{\nabla} P, \tag{3.2.7}$$

which are the continuity (conservation of mass) equation and Euler's equation, respectively.

Consider a family of inertial observers with parallel 4-velocities v^a so that $\partial_b v^a = 0$. The above interpretation of T_{ab} gives us the mass-energy 4-current density of the fluid as measured by these observers:

$$J_a = -T_{ab} v^b, \tag{3.2.8}$$

and Eq. 3.2.3 implies $\partial^a J_a = 0$. Using Gauss's law, this implies that over the three-dimensional boundary S of any four-dimensional spacetime volume V, we have

$$\int_S J_a n^a dS = 0, \tag{3.2.9}$$

where n^a is the unit normal vector. This equation implies conservation of energy, and Eq. 3.2.3 holds for all continuous matter distributions.

3.2.1 Examples: scalar field and electromagnetic field

No classical scalar field exists in nature, but we may still consider a scalar field ϕ satisfying the Klein-Gordon equation

$$\partial^a \partial_a \phi - m^2 \phi = 0. \tag{3.2.10}$$

The stress-energy tensor of this scalar field is

$$T_{ab} = \partial_a \phi \partial_b \phi - \frac{1}{2}\eta_{ab}(\partial^c \phi \partial_c \phi + m^2 \phi^2). \tag{3.2.11}$$

In prerelativity physics, the electric field $\vec{E}$ and the magnetic field $\vec{B}$ are separate spatial vectors. In SR, they are combined into a single, antisymmetric tensor field F_{ab}, which has six independent

components. For an observer with 4-velocity v^a, we interpret the quantity $E_a = F_{ab}v^b$ as the electric field measured by the observer, and $B_a = -\frac{1}{2}\epsilon_{ab}{}^{cd}F_{cd}v^b$ as the measured magnetic field. Here ϵ_{abcd} is the **Levi-Civita symbol**, a totally antisymmetric tensor of positive orientation with norm $\epsilon_{abcd}\epsilon^{abcd} = -24$. In a right-handed orthonormal basis we have ϵ_{0123}. Thus, Maxwell's equations can be written as

$$\partial^a F_{ab} = -4\pi j_b, \tag{3.2.12}$$

$$\partial_{[a}F_{bc]} = 0, \tag{3.2.13}$$

where j^a is the 4-current density of electric charge. The antisymmetry of F_{ab} implies that $\partial^b\partial^a F_{ab} = -4\pi\partial^b j_b = 0$, i.e. electric charge is conserved. The Lorentz force law is the equation of motion for a particle of charge q

$$u^a\partial_a u^b = \frac{q}{m}F^b{}_c u^c. \tag{3.2.14}$$

The stress-energy tensor of the electromagnetic field is

$$T_{ab} = \frac{1}{4\pi}\left\{F_{ac}F_b{}^c - \frac{1}{4}\eta_{ab}F_{de}F^{de}\right\}. \tag{3.2.15}$$

Both T_{ab}'s above satisfiy the energy condition and Eq. 3.2.3. We may write F_{ab} in terms of the **vector potential** A^a:

$$F_{ab} = \partial_a A_b - \partial_b A_a. \tag{3.2.16}$$

A^a has a gauge freedom and the gauge transformation $A^a \to A^a + \partial_a\chi$ keeps F_{ab} unchanged. If impose the Lorenz gauge condition

$$\partial^a A_a = 0, \tag{3.2.17}$$

Maxwell's equation becomes

$$\partial^a\partial_a A_b = -4\pi j_b, \tag{3.2.18}$$

which may be solved with oscillating wave solutions of constant amplitude $A_a = C_a\exp(iS)$, where S is the phase of the wave. According to the above equations, for $j^a = 0$, S must satisfy the following

$$\partial^a\partial_a S = 0, \tag{3.2.19}$$

$$\partial_a S\partial^a S = 0, \tag{3.2.20}$$

$$C_a\partial^a S = 0. \tag{3.2.21}$$

Note for any function f on any manifold with a metric, $\nabla^a f$ is normal to the surfaces of constant f. Eq. 3.2.20 states that the normal $k^a = \partial^a S$ to surfaces of constant S is a null vector. Such a surface is called a **null hypersurface**. Null hypersurfaces have their normal vector tangent to the hypersurface. We can differentiate Eq. Eq. 3.2.20 to show that the integral curves of k^a are null geodesics. The **frequency** of the wave as measured by an observer with 4-velocity v^a is $\omega = -v^a\partial_a S = -v^a k_a$. For plane waves,

$$S = \sum_{\mu=0}^{3} k_\mu x^\mu, \tag{3.2.22}$$

where $\{x^\nu\}$ are global inertial coordinates and k_μ are constants. All well behaved solutions of Maxwell's equations which vanish at large spacial distances sufficiently rapidly can be expressed as superpositions of plane waves. The above analysis suggests that light signals propagate on null geodesics and gives rise to the idea of a light cone.

3.3 General Relativity

Einstein did not develop a theory that generalizes Newton's theory and make it compatible with special relativity like what Maxwell did with electromagnetism. The primary motivation for his development of a new theory include the equivalence principle (all bodies fall the same way in a gravitational field) and Mach's principle (all matter in the universe should contribute to the local definition of "inertial motion").

To see how these principles tie into the theory of gravitation, let us consider how we measure the electromagnetic field in special relativity. We have inertial observers who are not subject to EM forces or any other forces. We then release a charged test particle, whose world line should satisfy the Lorentz force law. This allows us to determine F_{ab} by observing the deviation from inertial motion for the particle.

The problem of this method with gravity is that we cannot insulate the observer from the gravitational force, as the observer will move exactly the same way as the test body. This means there is no simple way of measuring the gravitational force field. Therefore, the theory of general relativity makes the following hypothesis: *the spacetime metric is not flat as was assumed in special relativity. The world lines of freely falling bodies in a gravitational field are the geodesics of the curved spacetime metric.* In this way, the "background observers" (geodesics of the spacetime metric) coincide with what was previously viewed as motion in a "gravitational force field", so gravity is viewed instead as an aspect of spacetime structure. There is no meaning to an absolute gravitational force, but relative gravitational force (tidal force) still has meaning and can be measured.

In the Newtonian viewpoint, the gravitational force on an object on Earth's surface is balanced by the force exerted by the surface. In GR, only the surface exerts a force, which makes the object deviate from geodesic motion at a rate 9.8 $\mathrm{m/s^2}$. The object remains in a stationary state because in the curved spacetime geometry near the Earth, the orbits of time translation symmetry are different from the geodesics of the metric. This time translation symmetry allows us to define a preferred set of background observers, and the Earth's gravitational force field can be defined as the negative acceleration an object need in order to remain stationary. Without time translation symmetry, we cannot have a well defined gravity force.

General relativity allows a Lorentz metric g_{ab} to be curved and places no a priori restriction on the spacetime manifold. It asserts that spacetime is curved in all situations where a gravitational field is present, and Einstein's equation relates the spacetime geometry to matter distribution. The statement about the spacetime structure now becomes: *Spacetime is a manifold M with a Lorentz metric g_{ab}.*

The laws of physics are governed by the principle of general covariance and the requirement that equations can be reduced to the equations satisfied in special relativity when the metric is flat. Since the only modification from SR is that the spacetime manifold can be different from $\mathbb{R}^4$, we can continue to represent physical quantities by the same type of tensor fields. We modify the equations satisfied in SR by replacing η_{ab} with g_{ab} and the operator ∂_a with ∇_a ("minimal substitution"). Thus, a free particle satisfies the geodesic equation $a^b = u^a \nabla_a u^b = 0$, where a^b is the (absolute) acceleration of the particle. Unless specified, all equations in the previous section will change accordingly.

One importance difference is that parallel transport is now path dependent due to spacetime being curved. This means we cannot define the energy of a distant particle for a given observer. There is an altered interpretation for the modified Eq. 3.2.3:

$$\nabla^a T_{ab} = 0 \tag{3.3.1}$$

A family of observers is represented by a unit timelike vector field v^a. If we can find a covariantly constant v^a (i.e. $\nabla_a v_b = 0$), or $\nabla_{(a}v_{b)} = 0$ (Killing's equation[10]), then $\nabla^a(T_{ab}v^b) = 0$. The curved spacetime version of Gauss's law again leads to conservation of energy. However, in curved spacetime in general one cannot find a v^a satisfying $v^a v_a = -1$ and $\nabla_{(a}v_{b)} = 0$, and the conclusion of energy conservation from Eq. 3.3.1 is only approximately true in a spacetime region of dimension small compared with the radius of curvature, because tidal forces can do work on the fluid and change its locally measured energy. Thus, Eq. 3.3.1 may be interpreted as a local conservation of material energy over small regions of spacetime, and this holds for all matter and fields.

It is worth pointing out that in addition to the natural generalization of Eq. 3.2.10 using the minimal substitution rule, other generalizations exist, such as

$$\nabla^a\nabla_a\phi - m^2\phi - \alpha R\phi = 0. \tag{3.3.2}$$

The generalized Maxwell's equations allow us to introduce a vector potential A_a locally. However, due to the commutation of derivatives, Eq. 3.2.18 has an explicit curvature term:

$$\nabla^a\nabla_a A_b - R^d{}_b A_d = -4\pi j_b. \tag{3.3.3}$$

Without the curvature term, this equation would not satisfy current conservation.

In situations where the spacetime scale of variation of the EM field is much smaller than that of the curvature, we would again expect an oscillating wave solution for Maxwell's equations, but with *nearly* constant amplitude, i.e. derivatives of C_a are small. Substituting this solution into Eq. 3.3.3 and neglecting the small terms ($\nabla_b\nabla^b C_a$ and the Ricci term), we get $\nabla_a S\nabla^a S = 0$. Under this approximation (known as the **geometrical optics approximation**), the result suggests that light travels on null geodesics.

We have described how the laws of physics and the motion of particles would change in a curved spacetime. Mach's principle would lead us to the question: how is the spacetime geometry influenced by the matter distribution of the universe? To find the equation that describes the relation between spacetime geometry and the matter distribution, we can compare the description of tidal force in Newtonian gravity and GR. In the Newtonian theory, the gravitational field may be represented by a potential ϕ, and the tidal acceleration of two nearby particles is given by $-(\vec{x}\cdot\vec{\nabla})\vec{\nabla}\phi$, where $\vec{x}$ is the separation vector. In GR, the tidal acceleration is given by $-R_{cbd}{}^a v^c x^b v^d$ according to Eq. 2.3.11, where v^a is the 4-velocity of the particles and x^a is the deviation vector. This implies the correspondence

$$R_{cbd}{}^a v^c v^d \leftrightarrow \partial_b\partial^a\phi. \tag{3.3.4}$$

Poisson's equation[11] tells us that:

$$\nabla^2\phi = 4\pi\rho, \tag{3.3.5}$$

where ρ is the mass/energy density of matter. As described before, in SR and GR the energy properties of matter are described by T_{ab}, so we have the correspondence

$$T_{ab}v^a v^b \leftrightarrow \rho, \tag{3.3.6}$$

[10]The solutions v^a are called the **Killing vector field**, often denoted ξ^a, which describes the direction of time translation invariance.

[11]Note that we have $G_N = c = 1$ throughout the text.

where v^a is the 4-velocity of the observer. Combining the three equations above, we would get $R_{cad}{}^a v^c v^d = 4\pi T_{cd} v^c v^d$, which suggests the field equation $R_{cd} = 4\pi T_{cd}$. This is the equation originally postulated by Einstein. However, it has some problems. As discussed before, the stress tensor satisfies $\nabla^c T_{cd} = 0$. On the other hand, Eq. 2.2.13 tells us that $\nabla^c(R_{cd} - \frac{1}{2}g_{cd}R) = 0$. An equality of R_{cd} and $4\pi T_{cd}$ would imply $\nabla_d R = 0$, which means R, and hence $T = T^a_{\ a}$, is constant throughout the universe. This is highly unphysical. This problem is resolved by **Einstein's equation**:

$$G_{ab} \equiv R_{ab} - \frac{1}{2}Rg_{ab} = 8\pi T_{ab}, \tag{3.3.7}$$

under which the Bianchi identity implies loval energy conservation. The correspondences that motivated the previous equation are not destroyed. Taking the trace of Eq. 3.3.7, we find $R = -8\pi T$ and thus

$$R_{ab} = 8\pi(T_{ab} - \frac{1}{2}g_{ab}T). \tag{3.3.8}$$

In the Newtonian limit, the energy of matter as measured by an observer roughly at rest with respect to the masses will be much greater than the material stresses, so we have $T \approx -\rho = -T_{ab}v^a v^b$. This leads to $R_{ab}v^a v^b \approx 4\pi T_{ab} v^a v^b$.

The entire content of general relativity may be summarized as follows: *spacetime is a manifold M with a Lorentz metric g_{ab}, whose curvature is related to the matter distribution in spacetime by Einstein's equation.* Before we move on to the solutions of Einstein's equation, here are a few remarks: (1) Based on results in section 2.4.1, Einstein's equation is equivalent to a coupled system of nonlinear second order PDEs for the metric components $g_{\mu\nu}$; (2) Although Einstein's equation is analogous to Maxwell's equation in some sense, we must solve simultaneously for g_{ab} and T_{ab} because T_{ab} depends on the metric, while in the latter's case one may find A_a simply by specifying j_a; (3) Einstein's equation implies the equations of motion $\nabla^a T_{ab} = 0$, as well as the **geodesic hypothesis** that the world lines of test bodies are geodesics of the spacetime metric.

3.3.1 Einstein-Hilbert action

Alternatively, we can obtain Einstein's equation through the principle of least action, and the action in this case is given by

$$S = \frac{1}{16\pi}\int_M d^4x\sqrt{-g}R + (S_{\text{matter}}) \tag{3.3.9}$$

This is the **Einstein-Hilbert action**. The variation $g^{ab} \to g^{ab} + \delta g^{ab}$ produces Einstein's equation, in which T_{ab} would be represented by

$$T_{ab} = -\frac{2}{\sqrt{-g}}\frac{\delta S_{\text{matter}}}{\delta g^{ab}}. \tag{3.3.10}$$

The simplest extension to the Einstein-Hilbert action, however, is the cosmological constant Λ, which corresponds to an ideal fluid with $P = -\rho$:

$$S = \frac{1}{16\pi}\int_M d^4x\sqrt{-g}(R - 2\Lambda), \tag{3.3.11}$$

where the factor of 2 is by convention. This yields

$$R_{ab} - \frac{1}{2}g_{ab}R + \Lambda g_{ab} = 8\pi T_{ab} \tag{3.3.12}$$

An example of the extension S_{matter} is the action of the scalar field and the EM field

$$S_{\text{matter}} = \frac{1}{2}\int d^4x\sqrt{-g}(-\nabla^a\phi\nabla_a\phi - m^2\phi^2) - \frac{1}{4e^2}\int d^4x\sqrt{-g}(g^{ac}g^{bd}F_{ab}F_{cd}). \tag{3.3.13}$$

While varying the E-H action, we would come across a surface term, which needs to be canceled by adding an additional term in the action

$$\frac{1}{8\pi}\int_{\partial M} d^3xK, \tag{3.3.14}$$

where $K = h^{ab}K_{ab}$ and $K_{ab} = \nabla_a n_b$ is the change of the unit normal of the manifold M. ∂M is the surface of M, and

$$h_{ab} = g_{ab} \pm n_a n_b \tag{3.3.15}$$

is the induced metric on ∂M, with $+/-$ corresponding to a spacelike/timelike ∂M. Note that $n^a h_{ab} = n_b + (n^a n_a)n_b = 0$. This K describes the **extrinsic curvature** on the boundary of the region.

3.4 Linearized Gravity

In this section we consider the case in which gravity is weak. In the context of GR this means the spacetime metric is nearly flat, and we can treat this problem in a perturbative way, i.e. we write the metric and inverse metric as

$$g_{ab} = \eta_{ab} + \gamma_{ab}, \tag{3.4.1}$$
$$g^{ab} = \eta^{ab} - \gamma^{ab}, \tag{3.4.2}$$

with the perturbation γ_{ab} being small. In this context, small means the components $\gamma_{\mu\nu}$ in some global inertial coordinate system is much smaller than 1. "linearized gravity" is the approximation to GR obtained by substituting Eq. 3.4.1 for g_{ab} in Einstein's equation and retaining terms linear in γ_{ab}. We will raise and lower tensor indices with η_{ab} and η^{ab} instead of g_{ab} and g^{ab}, in order not to have γ_{ab} hidden in a raised or lowered index.

In a global inertial coordinate system, to linear order in γ_{ab} the Christoffel symbol is

$$\Gamma^c{}_{ab} = \frac{1}{2}\eta^{cd}(\partial_a\gamma_{bd} + \partial_b\gamma_{ad} - \partial_d\gamma_{ab}) \tag{3.4.3}$$

and the Ricci tensor is

$$R^{(1)}_{ab} = \partial_c\Gamma^c{}_{ab} - \partial_a\Gamma^c{}_{cb}. \tag{3.4.4}$$

Defining $\bar{\gamma}_{ab} \equiv \gamma_{ab} - \frac{1}{2}\eta_{ab}\gamma$, where $\gamma = \gamma^c{}_c$, the linearized Einstein equation is

$$G^{(1)}_{ab} = R^{(1)}_{ab} - \frac{1}{2}\eta_{ab}R^{(1)} = -\frac{1}{2}\partial^c\partial_c\bar{\gamma}_{ab} + \partial^c\partial_{(b}\bar{\gamma}_{a)c} - \frac{1}{2}\eta_{ab}\partial^c\partial^d\bar{\gamma}_{cd} = 8\pi T_{ab}. \tag{3.4.5}$$

There is a gauge freedom in GR corresponding to the group of diffeomorphisms. If $\phi : M \to M$ is a diffeomorphism of spacetime, then the metrics g_{ab} and $\phi^* g_{ab}$ represent the same spacetime geometry, where ϕ^* is the map on tensor fields induced by ϕ:

$$\begin{aligned}&(\phi^*T)^{a_1,\ldots,a_k}{}_{b_1,\ldots,b_l}(\mu_1)_{b_1}\ldots(\mu_k)_{b_k}(t_1)^{a_1}\ldots(t_l)^{a_l}\\ &= T^{a_1,\ldots,a_k}{}_{b_1,\ldots,b_l}(\phi^*\mu_1)_{b_1}\ldots(\phi_*\mu_k)_{b_k}([\phi^{-1}]^*t_1)^{a_1}\ldots([\phi^{-1}]^*t_l)^{a_l}.\end{aligned} \tag{3.4.6}$$

In section 1.2, we mentioned that an infinitesimal diffeomorphism is generated by a vector field ξ^a, and the change in a tensor field induced by this diffeomorphism defines a Lie derivative $\mathcal{L}_\xi$. In the linear approximation, this gauge freedom implies that two perturbations γ_{ab} and $\gamma_{ab} + \mathcal{L}_\xi \eta_{ab}$ represent the same physical perturbation since they differ by the action of an infinitesimal diffeomorphism on η_{ab}. Note that $\mathcal{L}_\xi \eta_{ab} = \partial_a \xi_b + \partial_b \xi_a$, so the gauge freedom can be expressed as

$$\gamma_{ab} \to \gamma_{ab} + \partial_a \xi_b + \partial_b \xi_a. \tag{3.4.7}$$

This is analogous to the gauge freedom of the electromagnetic vector potential A_a. This gauge freedom allows us to simplify the linearized Einstein equation. By solving $\partial^b \partial_b \xi_a = -\partial^b \bar{\gamma}_{ab}$ for ξ^a, we can make a gauge transformation to obtain a condition analogous to the Lorenz gauge:

$$\partial_b \bar{\gamma}_{ab} = 0. \tag{3.4.8}$$

This condition gives the simplified linearized Einstein equation:

$$\partial^c \partial_c \bar{\gamma}_{ab} = -16\pi T_{ab}. \tag{3.4.9}$$

In vacuum ($T_{ab} = 0$), Eq. 3.4.8 and 3.4.9 describe a massless spin-2 field propagating in flat spacetime. Thus we may view GR as a theory of a massless spin-2 field that undergoes nonlinear self-interaction.

3.4.1 The Newtonian limit

In order to test the validity of the theory of GR, we can check the Newtonian limit, in which we assume the following: (1) gravity is weak; (2) relative motion of the sources is much smaller than c (i.e. we may neglect the "time-space" components of T_{ab}); and (3) the material stress is much smaller than the mass-energy density (i.e. we may neglect the "space-space" components of T_{ab}).

Given (1), the linear approximations to GR should be valid, and we can incorporate the other assumptions into the following statement: there exists a global inertial coordinate system of η_{ab} such that $T_{ab} \approx \rho t_a t_b$, where $t^a = (\partial/\partial x^0)^a$ is the time direction of the system. Since the sources are slowly varying, we also assume the time derivatives of $\bar{\gamma}_{ab}$ are negligible, and we arrive at the following equation:

$$\nabla^2 \bar{\gamma}_{\mu\nu} = -16\pi\rho \; 2\delta_{\mu 0}\delta_{\nu 0}. \tag{3.4.10}$$

The only solution for $\gamma_{\mu\nu}$ ($\mu \neq \nu$) that is well behaved at infinity is $\bar{\gamma}_{\mu\nu} = 0$, so the solution for the perturbed metric γ_{ab} is

$$\gamma_{ab} = \bar{\gamma}_{ab} - \frac{1}{2}\eta_{ab}\bar{\gamma} = -(4t_a t_b + 2\eta_{ab})\phi \tag{3.4.11}$$

where $\phi \equiv -\frac{1}{4}\bar{\gamma}_{00}$ satisfies Poisson's equation $\nabla^2 \phi = 4\pi\rho$. The motion of test bodies in this curved spacetime geometry is governed by the geodesic equation (Eq. 2.3.3 with τ). For $v \ll c$, we may approximate $dx^\alpha/d\tau$ as $(1, 0, 0, 0)$ and the proper time τ is roughly the coordinate t, so we have

$$\frac{d^2 x^\mu}{dt^2} = -\Gamma^\mu{}_{00} = \frac{1}{2}\frac{\partial \gamma_{00}}{\partial x^\mu} = -\frac{\partial \phi}{\partial x^\mu}. \tag{3.4.12}$$

This is the familiar equation

$$\vec{a} = -\vec{\nabla}\phi. \tag{3.4.13}$$

If we take into account the lowest order effects of the motion of the sources, we can approximate the stress energy tensor as

$$T_{ab} = 2t_{(a}J_{b)} - \rho t_a t_b, \tag{3.4.14}$$

where $J_b = -T_{ab}t^a$ is the mass-energy 4-current density. The space-space components of $\bar{\gamma}_{ab}$ again satisfy the source free wave equation, but the other components now satisfy

$$\partial^a \partial_a \bar{\gamma}_{0\mu} = 16\pi J_\mu. \tag{3.4.15}$$

The quantity $-\frac{1}{4}\bar{\gamma}_{ab}t^b$ satisfies Maxwell's equations in the Lorenz gauge and we can interpret it as a vector potential A_a. With the same assumptions about $\bar{\gamma}_{ab}$ as before, we can get

$$\vec{a} = -\vec{E} - 4\vec{v} \times \vec{B}, \tag{3.4.16}$$

where $\vec{E}$ and $\vec{B}$ are defined the same way in terms of A_a as in electromagnetism. This is similar to the Lorentz force law except for an overall minus sign and an extra factor of 4.

3.4.2 Gravitational radiation

GR implies the existence of gravitational radiation, similar to how Maxwell's equations led to electromagnetic radiation. The propagation of gravitational radiation can be described by the source-free, linearized Einstein equation with the Lorenz gauge choice:

$$\partial^a \bar{\gamma}_{ab} = 0, \tag{3.4.17}$$

$$\partial^c \partial_c \bar{\gamma}_{ab} = 0. \tag{3.4.18}$$

Here we will choose the Coulomb gauge (a.k.a. radiation gauage), $\partial^b \partial_b \xi^a = 0$, which leaves Eq. 3.4.17 unchanged. We use this condition to achieve $\gamma = 0, \gamma_{0\mu} = 0$ for $\mu = 1, 2, 3$ in a source free region. If no sources are present throughout spacetime, we also get γ_{00}. The radiation gauge can be achieve by solving for the initial values and first derivatives of the four components of ξ_a in the following equations on the initial surface $t = t_0$:

$$2\left(-\frac{\partial \xi_0}{\partial t} + \vec{\nabla} \cdot \vec{\xi}\right) = -\gamma, \tag{3.4.19}$$

$$2\left[-\nabla^2 \xi_0 + \vec{\nabla} \cdot \left(\frac{\partial \vec{\xi}}{\partial t}\right)\right] = -\frac{\partial \gamma}{\partial t}, \tag{3.4.20}$$

$$\frac{\partial \xi_\mu}{\partial t} + \frac{\partial \xi_0}{\partial x^\mu} = -\gamma_{0\mu} \ (\mu = 1, 2, 3), \tag{3.4.21}$$

$$\nabla^2 \xi_\mu + \frac{\partial}{\partial x^\mu}\left(\frac{\partial \xi_0}{\partial t}\right) = -\frac{\partial \gamma_{0\mu}}{\partial t} \ (\mu = 1, 2, 3). \tag{3.4.22}$$

Under the radiation gauge, $\gamma = 0$, $\gamma_{00} = 0$, and $\gamma_{0\mu} = 0$ for $\mu = 1, 2, 3$. Eq. 3.4.17 then gives us

$$\frac{\partial \gamma_{00}}{\partial t} = 0, \tag{3.4.23}$$

and the linearized Einstein equation becomes:

$$\nabla^2 \gamma_{00} = -16\pi T_{00}. \tag{3.4.24}$$

This equation yields a constant solution for γ_{00} that can be transformed to 0 if $T_{00} = 0$ throughout the spacetime.

Using this radiation gauge, solutions to the source free linearized Einstein equation (Eq. 3.4.18) are plane waves:

$$\gamma_{ab} = H_{ab} \exp\left(i \sum_{\mu=0}^{3} k_\mu x^\mu \right), \tag{3.4.25}$$

where H_{ab} is a constant tensor field, if and only if

$$\sum_{\mu} k_\mu x^\mu = \sum_{\mu,\nu} \eta^{\mu\nu} k_\mu x_\nu = 0. \tag{3.4.26}$$

The radiation gauge conditions require (for $\nu = 0, 1, 2, 3$)

$$\sum_{\mu=0}^{3} k^\mu H_{\mu\nu} = 0, \tag{3.4.27}$$

$$H_{0\nu} = 0, \tag{3.4.28}$$

$$\sum_{\mu=0}^{3} H^\mu{}_\mu = 0. \tag{3.4.29}$$

The last two equations both imply $\sum_\nu H_{0\nu} k^\nu = 0$, so only eight of these nine equations are independent. Since there are 10 independent components $H_{\mu\nu}$, we have two linearly independent solutions for H_{ab}, which describe the two polarization states of gravitational waves.

The most straightforward way to detect gravitational waves is to measure the gravitational tidal force (relative acceleration) of two masses. This is described by the geodesic deviation equation (Eq. 2.3.11) for two nearby free falling bodies. Note that in the Newtonian limit, $\tau \approx t$:

$$\frac{d^2 X^\mu}{dt^2} \approx \sum_{\nu} R_{\nu 00}{}^{\mu} X^\nu, \tag{3.4.30}$$

where x^a is the deviation vector. In the radiation gauge, we can use Eq. 2.4.2 to get an expression for the relevant components of the linearized Riemann tensor:

$$R_{\nu 00 \mu} = \frac{1}{2} \frac{\partial^2 \gamma_{\mu\nu}}{\partial t^2}. \tag{3.4.31}$$

Strong gravitational waves are produced by sources such as collapse phenomena where gravity is not weak and the linear approximation cannot be used. For the purpose of illustrating the generation of radiation, we will stick with the linear approximation and solve Eq. 3.4.9, and the solutions are given in terms of the sources by the retarded Green's function as used in electromagnetism:

$$\bar{\gamma}_{\mu\nu}(x) = 4 \int_\Lambda \frac{T_{\mu\nu}(x')}{|\vec{x} - \vec{x}'|} dS(x'), \tag{3.4.32}$$

where Λ is the past lightcone of the point x, and the volume element on the lightcone is $dS = r^2 dr d\Omega$.

In the slow motion limit, where the typical source velocities are much smaller than c (or the spatial extent of the source is much smaller than the wavelength of the emitted radiation), we can

make the **dipole approximation**. We first Fourier transform all quantities in the time variable as follows:

$$\hat{\bar{\gamma}}_{\mu\nu}(\omega,\vec{x}) = \frac{1}{\sqrt{2}}\int_{-\infty}^{\infty} \bar{\gamma}_{\mu\nu}(t,\vec{x})e^{i\omega t}dt. \tag{3.4.33}$$

From Eq. 3.4.32 it follows that

$$\hat{\bar{\gamma}}_{\mu\nu}(\omega,\vec{x}) = 4\int \frac{\hat{T}_{\mu\nu}(\omega,\vec{x}')}{|\vec{x}-\vec{x}'|}\exp\left(i\omega|\vec{x}-\vec{x}'|\right)d^3x'. \tag{3.4.34}$$

We only need to solve for the space-space components of $\hat{\bar{\gamma}}_{\mu\nu}$ since we can use Eq. 3.4.8 to find $\hat{\bar{\gamma}}_{0\mu}$:

$$-i\omega\hat{\bar{\gamma}}_{0\mu} = \sum_{\nu=1}^{3}\frac{\partial\hat{\bar{\gamma}}_{\nu\mu}}{\partial x^{\nu}}. \tag{3.4.35}$$

For simplicity, let us consider the far zone radiation, i.e. $R \gg 1/\omega$, wher R is the distance from the source. In this case, the factor $\exp\left(i\omega|\vec{x}-\vec{x}'|\right)$ is roughly constant over the source, so we can replace $\frac{\exp(i\omega|\vec{x}-\vec{x}'|)}{|\vec{x}-\vec{x}'|}$ with $\frac{\exp(i\omega R)}{R}$ and pull it out of the integral. The remaining integral is

$$\int \hat{T}^{\mu\nu}d^3x = ... = -\frac{\omega^2}{2}\int \hat{T}^{00}x^{\mu}x^{\nu}d^3x \equiv -\frac{\omega^2}{6}\hat{q}_{\mu\nu}, \tag{3.4.36}$$

where $\hat{q}_{\mu\nu}$ is the Fourier transform of the quadrupole moment tensor, and I omitted several steps involving integration by parts and applications of Gauss's law and conservation of T_{ab}. The far zone solution is therefore

$$\hat{\bar{\gamma}}_{\mu\nu}(\omega,\vec{x}) = -\frac{2\omega^2}{3}\frac{e^{i\omega R}}{R}\hat{q}_{\mu\nu}(\omega) \quad (\mu,\nu=1,2,3), \tag{3.4.37}$$

and the inverse Fourier transform yields

$$\bar{\gamma}_{\mu\nu}(t,\vec{x}) = \frac{2}{3R}\frac{d^2q_{\mu\nu}}{dt^2}\Big|_{\text{ret}} \quad (\mu,\nu=1,2,3), \tag{3.4.38}$$

where the derivative is evaluated at $t' = t - R$. The absence of dipole radiation in this case can be attributed to the conservation of momentum, and this implies gravitational radiation is smaller than the electromagnetic radiation in comparable situations.

The notion of local energy density is hard to define in GR, because the metric g_{ab} describes both the background spacetime structure and the dynamical aspects of the gravitational field. However, we may define a *total energy* for an isolated system observed at large distances. For small deviations from flat spacetime, we can expect that the total energy and energy flux of the gravitational field are quadratic in the field γ_{ab}, in analogy to electromagnetism. We start by considering the linearized vaccum Einstein equation

$$G^{(1)}_{ab}[\gamma_{cd}] = 0, \tag{3.4.39}$$

which states that the Einstein tensor for the metric $\eta_{ab}+\gamma_{ab}$ vanishes to first order in γ_{ab}. However, to second order in γ_{ab}, this equation will not be satisifed in general, and the Ricci tensor quadratic in γ_{ab} is

$$\begin{aligned} R^{(2)}_{ab} = {} & \frac{1}{2}\gamma^{cd}\partial_a\partial_b\gamma_{cd} - \gamma^{cd}\partial_c\partial_{(a}\gamma_{b)d} + \frac{1}{4}(\partial_a\gamma_{cd})\partial_b\gamma^{cd} + (\partial^d\gamma^c{}_b)\partial_{[d}\gamma_{c]a} \\ & + \frac{1}{2}\partial_d(\gamma^{dc}\partial_c\gamma_{ab}) - \frac{1}{4}(\partial^c\gamma)\partial_c\gamma_{ab} - (\partial_d\gamma^{cd} - \frac{1}{2}\partial^c\gamma)\partial_{(a}\gamma_{b)c}. \end{aligned} \tag{3.4.40}$$

Thus, in order to maintain a solution for the vacuum Einstein equation, we must correct γ_{ab} by adding to it a term $\gamma^{(2)}_{ab}$, which satisfies

$$G^{(1)}_{ab}[\gamma^{(2)}_{cd}] + G^{(2)}_{ab}[\gamma_{cd}] = G^{(1)}_{ab}[\gamma^{(2)}_{cd}] - 8\pi t_{ab} = 0, \tag{3.4.41}$$

where we have $G^{(2)}_{ab} = R^{(2)}_{ab} - \frac{1}{2}\eta_{ab}R^{(2)}$ in the case $R^{(1)}_{ab} = 0$, and $t_{ab} = -\frac{1}{8\pi}G^{(2)}_{ab}[\gamma_{cd}]$. Thus, we may view t_{ab} as the effective stress-energy tensor valid to second order in deviation from flatness, but we cannot take this interpretation too literally because t_{ab} is not gauge invariant. However, the total energy associated with γ_{ab},

$$E = \int_\Sigma t_{00} d^3x \tag{3.4.42}$$

is gauge invariant because the inertial components of γ_{ab} and its derivatives go to zero as $r \to \infty$: $\gamma_{\mu\nu} = \mathcal{O}(1/r)$, $\partial_\rho \gamma_{\mu\nu} = \mathcal{O}(1/r^2)$, and $\partial_\gamma \partial_\rho \gamma_{\mu\nu} = \mathcal{O}(1/r^3)$, and the perturbed spacetime metric is asymptotically flat (note these conditions are not appropriate in the time dependent regime because we expect $\partial_\rho \gamma_{\mu\nu} = \mathcal{O}(1/r)$ from Eq. 3.4.32). Similarly, although the local energy flux $-t^a{}_0$ is not gauge invariant, the total radiated energy

$$\Delta E = -\int_S t_{a0} dS^a \tag{3.4.43}$$

is gauge independent if the spacetime is initially time independent and ends up being time independent. We can calculate from Eq. 3.4.41 and 3.4.43 the enrgy carried away by gravitational radiation produced by a slowly varying source:

$$\Delta E = \int P \, dt = \frac{1}{45} \sum_{\mu,\nu=1}^{3} \int \left(\frac{d^3 Q_{\mu\nu}}{dt^3}\Big|_{\text{ret}} \right)^2 dt, \tag{3.4.44}$$

where $Q_{\mu\nu}$ is the trace free quadrupole moment tensor

$$Q_{\mu\nu} = q_{\mu\nu} - \frac{1}{3}\delta_{\mu\nu} q. \tag{3.4.45}$$

As an example, the gravitational energy flux from a rod of mass M and length L that spins about its center at frequency Ω (T_{ab} oscillates at 2Ω) is

$$P_{\text{rod}} = \frac{2G}{45c^5} M^2 L^4 \Omega^6. \tag{3.4.46}$$

4 Homogeneous, Isotropic Cosmology

4.1 Homogeneity and Isotropy

In cosmology, we must realize that the observational data only covers a small region in the universe, and we must rely on certain philosophical prejudices in the development of cosmological models. Two commonly accepted assumptions are that we do not occupy a privileged position in our universe and the basic characteristics of our surroundings would be the same if we were at a different region in the universe (homogeneity), and that there are no preferred directions in space (isotropy). These assumptions have been confirmed on the largest scales through observations.

We will give mathematical definitions to these assumptions. A spacetime is said to be **homogeneous** if there exists a one-parameter family of spacelike hypersurfaces Σ_t such that for each t and for any points $p, q \in \Sigma_t$ there exists an isometry of the spacetime metric g_{ab} that takes p into q. As for isotropy, we must point out that at each point, at most one observer can see the universe at isotropic, and any observer with a relative motion to the matter distribution will see an anisotropic velocity distribution of matter. A spacetime is spatially **isotropic** at each point if there exists a congruence of timelike curves (observers) with tangents u^a filling the spacetime and satisfying the following property: given any point p and two unit spatial tangent vectors $s_1^a, s_2^a \in V_p$ (i.e. vectors at p orthogonal to u^a), there exists an isometry of g_{ab} which leaves p and u^a fixed but rotates s_1^a into s_2^a. It is therefore impossible to construct a geometrically preferred tangent vector orthogonal to u^a in an isotropic universe.

With the above definitions, it is easy to see that the surfaces Σ_t of homogeneity are orthogonal to the tangents u^a to the world lines of isotropic observers. The spacetime metric g_{ab} induces a Riemannian metric $h_{ab}(t)$ on each surface by restricting the action of g_{ab} at each p to vectors tangent to Σ_t. The assumptions of homogeneity and isotropy place the following requirements on the induced geometry of Σ_t: (1) there must be isometries of h_{ab} that carry $p \in \Sigma_t$ into any $q \in \Sigma_t$; (2) it is impossible to construct any geometrically preferred vectors on Σ_t.

Consider the Riemann tensor ${}^{(3)}R_{abc}{}^{d}$ constructed from h_{ab} on Σ_t. If we raise the third index, we may view ${}^{(3)}R_{ab}{}^{cd}$ at point p as a linear map $L : W \to W$, where W is the vector space of two forms, i.e. antisymmetric rank $(0,2)$ tensors, at p. By the symmetric property of the Riemann tensor, L is symmetric and is a self-adjoint map. Therefore, W has an orthonormal basis of eigenvectors of L. If the eigenvalues are distinct, the we would be able to pick out a preferred two-form at p and therefore a preferred vector at p. Thus, to not violate isotropy, all eigenvalues of L must be equal, meaning that L is a multiple of the identity operator:

$$L = K\mathbb{1}. \tag{4.1.1}$$

Thus, we have

$${}^{(3)}R_{ab}{}^{cd} = K\delta^{c}_{\ [a}\delta^{d}_{\ b]} \text{ and } {}^{(3)}R_{abcd} = Kh_{c[a}h_{b]d}. \tag{4.1.2}$$

Both homogeneity and isotropy require that K must be a constant and it cannot vary from point to point on Σ_t. We can show this by plugging Eq. 4.1.2 into the Bianchi identity to obtain

$$0 = D_{[e}{}^{(3)}R_{ab]cd} = (D_{[e}K)h_{|c|a}h_{b]d}, \tag{4.1.3}$$

where D_a denotes the derivative operator associated with h_{ab} on Σ_t. On a manifold of dimension three or higher, the right hand side of this equation will vanish if and only if $D_eK = 0$, i.e. K is constant.

A space where Eq. 4.1.2 is satisfied is called a **space of constant curvature**. It can be shown that any two such spaces of the same dimension and metric signature with equal values of K must be locally isometric. Thus, we may enumerate spaces of constant curvature for all values of K to determine the possible spatial geometries of Σ_t. All positive values of K are attained by 3-spheres, defined as surfaces in $\mathbb{R}^4$ whose Cartesian coordiantes satisfy

$$x^2 + y^2 + z^2 + w^2 = R^2. \tag{4.1.4}$$

In spherical coordinates, the metric of the unit 3-sphere is

$$ds^2 = d\psi^2 + \sin^2\psi(d\theta^2 + \sin^2\theta d\phi^2). \tag{4.1.5}$$

$K = 0$ is attained by the ordinary three-dimensional flat space with metric $ds^2 = dx^2 + dy^2 + dz^2$. All negative values of K are attained by the three-dimensional hyperboloids, defined as surfaces in a four dimensional flat Lorentz signature space (Minkowski spacetime) whose global inertial coordinates satisfy

$$t^2 - x^2 - y^2 - z^2 = R^2. \tag{4.1.6}$$

In hyperbolic coordinates, the metric of the unit hyperboloid is

$$d\psi^2 + \sinh^2\psi(d\theta^2 + \sin^2\theta d\phi^2) \tag{4.1.7}$$

In prerelativity physics as well as special relativity, it is assumed that $K = 0$. It is worth pointing out that the 3-sphere spatial geometry of the positive K case gives us a compact manifold that describes a finite universe with no boundary. We refer to this as a "closed" universe, while universes described by the other two possibilities are called "open."

Since the isotropic observers are orthogonal to the homogeneous surfaces, we can write g_{ab} as

$$g_{ab} = -u_a u_b + h_{ab}(t), \tag{4.1.8}$$

where for each t, $h_{ab}(t)$ is the metric of either a sphere, a flat Euclidean space, or a hyperboloid on Σ_t. We choose the appropriate coordiantes on the hypersurfaces, and carry these coordinates to each of the other homogeneous hypersurfaces by assigning a fixed spatial coordinate label to each observer. We label each hypersurfaces by the proper time τ of any of the isotropic observers. Thus, we can label each event in the universe with τ and the spatial coordinates, and the spacetime metric is given by

$$ds^2 = -d\tau^2 + a^2(\tau)\begin{cases} d\psi^2 + \sin^2\psi(d\theta^2 + \sin^2\theta d\phi^2) \\ d\psi^2 + \psi^2(d\theta^2 + \sin^2\theta d\phi^2) \quad (= dx^2 + dy^2 + dz^2) \\ d\psi^2 + \sinh^2\psi(d\theta^2 + \sin^2\theta d\phi^2) \end{cases} \tag{4.1.9}$$

where $a(\tau)$ is some arbitrary positive function called the **scale factor**. Eq. 4.1.9 is called the Friedmann-Robertson-Walker metric. We will use Einnstein's equation to determine the spatial geometry and $a(\tau)$.

4.2 Dynamics of a Homogeneous, Isotropic Universe

We can substitute Eq. 4.1.9 into Einstein's equation to obtain predictions for the dynnamical evolution of the universe. To do so, we need to first describe the matter content in terms of the stress-energy T_{ab}. On the cosmic scales, we can treat the galaxies as "grains of dust." The random velocities of the galaxies are small, so we may neglect the "pressure" of these galaxy dusts. Thus, we may approximate the stress-energy tensor of matter in the present universe as

$$T_{ab} = \rho u_a u_b, \tag{4.2.1}$$

where ρ is the average mass density of matter. However, there exists other forms of mass-energy in the universe. The cosmic microwave background, for example, may be described by a perfect fluid stress-energy tensor with nonzero pressure (for massless radiation, $P = \frac{1}{3}\rho$). The contribution of this radiation to the stress-energy of the present universe is negligible, but it is important in the early universe. Thus, we will take T_{ab} to have the general perfect fluid form

$$T_{ab} = \rho u_a u_b + P(g_{ab} + u_a u_b). \tag{4.2.2}$$

By plugging Eq. 4.1.9 and 4.2.2 into Einstein's equation, we will get 10 equations corresponding to the 10 independent components of a symmetric two-index tensor. However, spacetime symmetries reduce these down to two independent equations. Namely, $G^{ab}u_b$ cannot have a spatial component, which would violate isotropy, so the "time-space" components of Einstein's equation are identically zero. If we project both indices of G_{ab} onto the homogeneous hypersurface and raise an index with h_{ab}, then by a similar argument we have that the resulting tensor is a multiple of the identity. Thus, the off-diagonal "space-space" components of Einstein's equation must vanish. The diagonal "space-space" components yield the same equations, so we have

$$G_{\tau\tau} = 8\pi T_{\tau\tau} = 8\pi\rho, \tag{4.2.3}$$

$$G_{**} = 8\pi T_{**} = 8\pi P, \tag{4.2.4}$$

where $G_{\tau\tau} = G_{ab}u^a u^b$ and $G_{**} = G_{ab}s^a s^b$. s^a is any unit vector tangent to the homogeneous hypersurfaces.

We will compute $G_{\tau\tau}$ and G_{**} in terms of $a(\tau)$ for the case of flat spatial geometry using the coordinate basis method. By Eq. 2.1.14, the nonvanishing components of the Christoffel symbol are

$$\Gamma^{\tau}{}_{xx} = \Gamma^{\tau}{}_{yy} = \Gamma^{\tau}{}_{zz} = a\dot{a}, \tag{4.2.5}$$

$$\Gamma^{x}{}_{x\tau} = \Gamma^{x}{}_{\tau x} = \Gamma^{y}{}_{y\tau} = \Gamma^{y}{}_{\tau y} = \Gamma^{z}{}_{z\tau} = \Gamma^{z}{}_{\tau z} = \dot{a}/a, \tag{4.2.6}$$

where $\dot{a} = da/d\tau$. The independent components of the Ricci tensor are calculated to be

$$R_{\tau\tau} = -3\frac{\ddot{a}}{a}, \tag{4.2.7}$$

$$R_{**} = a^{-2}R_{xx} = \frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a}, \tag{4.2.8}$$

$$R = -R_{\tau\tau} + 3R_{**} = 6\left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2}\right), \tag{4.2.9}$$

Thus we have

$$G_{\tau\tau} = R_{\tau\tau} + \frac{1}{2}R = 3\frac{\dot{a}^2}{a^2} = 8\pi\rho, \tag{4.2.10}$$

$$G_{**} = R_{**} - \frac{1}{2}R = -2\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} = 8\pi P. \tag{4.2.11}$$

Repeating the calculation for the cases of spherical and hyperbolic geometries, and rewriting the second equation, we get the general evolution equations for homogeneous, isotropic cosmology:

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi\rho}{3} - \frac{k}{a^2}, \tag{4.2.12}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}(\rho + 3P), \tag{4.2.13}$$

where $k = +1$ for the 3-sphere, $k = 0$ for flat space, and $k = -1$ for the hyperboloid. Eq. 4.2.12 is often known as the **Friedmann equation**. Eq. 4.2.13 is sometimes called the **acceleration equation**. We may solve these equations exactly for some special cases, such as a matter/dust

dominated universe ($P = 0$) and a radiation dominated universe ($P = \rho/3$), the results are listed in Table 5.1 of Wald.

Eq. 4.2.12 and 4.2.13 imply that the universe cannot be static if $\rho > 0$ and $P \geq 0$. The universe must always be either contracting or expanding. Note that the distance scale between all isotropic observers changes with time, but there is no preferred center of expansion or contraction. If the distance between two isotropic observers at time τ is R, the rate of change of R is

$$v \equiv \frac{dR}{d\tau} = \frac{R}{a}\frac{da}{d\tau} = HR, \tag{4.2.14}$$

where $H(\tau) = \dot{a}/a$ is the **Hubble parameter**. Eq. 4.2.14 is known as **Hubble's law**. Note that v can be greater than the speed of light for large enough R, but this does not contradict the theory of relativity because the postulate refers to the locally measured relative velocity between two objects at the same spacetime event, not a globally defined velocity between two distance objects.

Although various observations have confirmed the prediction of GR, Einstein was not happy with the prediction of a dynamic universe, and he proposed a modification of his equation by adding a new term:

$$G_{ab} + \Lambda g_{ab} = 8\pi T_{ab}, \tag{4.2.15}$$

where Λ is the **cosmological constant**. It can be shown that Eq. 4.2.15 gives the most general modification which does not grossly alter the basic properties of Einstein's equation. With this additional one-parameter degree of freedom, a static universe is possible. The original motivation for the introduction of Λ was lost after Hubble demonstrated the expansion of the universe. However, Λ has since been reintroduced on many occasions to account for discrepancies between theory and observations.

Given that the universe is expanding ($\dot{a} > 0$), we know from Eq. 4.2.13 that $\ddot{a} < 0$ (this is not true at the present time with a nonzero Λ). These observations would suggest that at some point in the distant past, we have $a = 0$, and this is referred to as the **big bang**. Since the spacetime structure itself is singular at the big bang, it does not make sense to ask about the state of the universe before the big bang, and there is no natural way to extend the spacetime manifold and metric beyond the singularity. It is important to point out, however, at the extreme conditions very near the big bang singularity, one expects quantum effects to become important and the predictions of GR should breakdown.

Before we discuss the qualitative predictions of GR for the future evolution of the universe, it is useful to obtain an equation for the evolution of the mass density. By manipulating Eq. 4.2.12 and 4.2.13, we have a statement about energy-momentum conservation:

$$\dot{\rho} + 3(\rho + P)\frac{\dot{a}}{a} = 0. \tag{4.2.16}$$

In order to obtain solutions of Eq. 4.2.12 and 4.2.13, it is often convenient to characterize the different components of the universe with an **equation of state**, given by

$$P = w\rho, \tag{4.2.17}$$

where the parameter w depends on the property of the component. For dust/non-relativistic matter, $w = 0$; for radiation, $w = \frac{1}{3}$; for the cosmological constant, $w = -1$. We usually consider the condition $-1 \leq w \leq 1$, where the lower bound is given by the null energy condition (for all null

vectors k^μ, $T_{\mu\nu}k^\mu k^\nu \geq 0$) and the upper bound is given by the causal energy condition. Solving Eq. 4.2.16 with our new parameterization, we get

$$\rho \propto a^{-3(1+w)}, \tag{4.2.18}$$

For dust ($w = 0$), we find $\rho a^3 =$ constant, which expresses conservation of rest mass; for radiation ($w = 1/3$), we find $\rho a^4 =$ constant. The energy density decreases more rapidly for radiation as a increases. We can think in terms of photons: the photon number density decreases as a^{-3}, but each photon loses energy as a^{-1} due to redshift. This result also confirms that the contribution of radiation to the total energy density dominates over that of ordinary matter in the early universe.

Plugging Eq. 4.2.18 into Eq. 4.2.12 to eliminate ρ, we can see that for a flat, single component universe,

$$a(\tau) \propto \tau^{\frac{2}{3(1+w)}}. \tag{4.2.19}$$

We see that in a matter dominated universe, $a(\tau) \propto \tau^{2/3}$; in a radiation dominated universe, $a(\tau) \propto \tau^{1/2}$; and in the case of $w = -1$, $a(\tau) \propto e^{C\tau}$, where C is a constant. The solution for the dust-filled universe with 3-sphere geometry is called the **Friedmann cosmology**.

Assuming the universe has three major components (matter, radiation, and the cosmological constant), it is often convenient to rearrange Eq. 4.2.12 and divide both sides by H^2 to obtain

$$1 = \frac{\rho_{\mathrm{m}}}{\rho_c} + \frac{\rho_{\mathrm{rad}}}{\rho_c} + \frac{\rho_\Lambda}{\rho_c} - \frac{k}{a^2H^2} \equiv \Omega_{\mathrm{m}} + \Omega_{\mathrm{rad}} + \Omega_\Lambda + \Omega_k, \tag{4.2.20}$$

where $\rho_c \equiv \frac{3H^2}{8\pi G_N}$. Ω_i is called the density parameter. Equation 4.2.20 gives a simplified description of how the various components contributes to the expansion and curvature of the universe. Qualitatively, Eq. 4.2.12 shows that if $k = 0$ or -1, $\dot{a}$ can never be zero, so if the universe is currently expanding, it will continue to expand forever. However, it also shows that for $k = +1$, the universe cannot expand forever and there exists a critical value a_c such that $a \leq a_c$. Experiments have shown that k is very close to 0, so it is reasonable for us to work with the flat geometry when solving Eq. 4.2.12.

4.3 The Cosmological Redshift; Horizons

4.3.1 Redshift

Suppose that at event p_1 at time τ_1 an isotropic observer emits a photon of frequency ω_1, and this photon is observed by another isotropic observer at event p_2 at time τ_2. We wish to find ω_2, the frequency measured by the second observer.

The solution of all redshift problems in relativity is governed by two facts: (1) in the geometric optics approximation, light travels on null geodesics; (2) the frequency of a light signal of wave vector k^a measured by an observer with 4-velocity u^a is $\omega = -k_a u^a$. Thus, we can always calculate the null geodesic determined by the initial values of k^a at the emission point and then the frequency at the observation point.

However, with symmetries, we may use a shortcut for the calculation of the observed frequency. Let ξ^a be a Killing vector field, i.e. a vector field that generates a one-parameter group of isometries. Let t^a be the tangent to the geodesic curve. Then $t^a\xi_a$ is constant along the geodesic. We may notice that for all three choices of spatial geometry, we can find a spacetime ξ^a that points in the direction of the projection of k^a onto Σ_1 at p_1 *and* points in the direction of the projection of k^a

onto Σ_2 at p_2. For example, in the case of flat spatial geometry, without loss of generality, we may assume that the projection of k^a onto Σ_1 at p_1 is in the $(\partial/\partial x)^a$ direction. We initially have $k^a(\partial/\partial y)_a = k^a(\partial/\partial z)_a = 0$, and since $(\partial/\partial y)^a$ and $(\partial/\partial z)^a$ are Killing vector fields, these inner products must also vanish at p_2. Thus, teh projection of k^a onto Σ_2 at p_2 also points in the $(\partial/\partial x)^a$ direction. In all cases, the length of ξ^a at p_2 changes from its length at p_1 in proportion to the change in the length scale factor a of the universe in going from Σ_1 to Σ_2:

$$\frac{\sqrt{\xi^a\xi_a}|_{p_1}}{\sqrt{\xi^a\xi_a}|_{p_2}} = \frac{a(\tau_1)}{a(\tau_2)}. \tag{4.3.1}$$

We note that since k^a is null, at any point its projection onto u^a must have the same magnitude as its projection onto Σ, so at p_1

$$k_a u_1^a = -k_a \left[\frac{\xi^a}{\sqrt{\xi^b\xi_b}}\right]\Big|_{p_1}. \tag{4.3.2}$$

Thus we have

$$\omega_1 = \left[\frac{k_a\xi^a}{\sqrt{\xi^b\xi_b}}\right]\Big|_{p_1}, \text{ and } \omega_2 = \left[\frac{k_a\xi^a}{\sqrt{\xi^b\xi_b}}\right]\Big|_{p_2}. \tag{4.3.3}$$

We have already shown that the inner product $k_a\xi^a$ is invariant, so we have

$$\frac{\omega_2}{\omega_1} = \frac{\sqrt{\xi^a\xi_a}|_{p_1}}{\sqrt{\xi^a\xi_a}|_{p_2}} = \frac{a(\tau_1)}{a(\tau_2)}. \tag{4.3.4}$$

This implies that the wavelength of each photon increases in proportion to the amount of expansion of the universe. Note that this notion of redshift is different from the Doppler shift. It describes the stretch of the wavelength between two observers at rest in the isotropic frame. The redshift factor is given by

$$z \equiv \frac{\lambda_2 - \lambda_1}{\lambda_1} = \frac{\omega_1}{\omega_2} - 1 = \frac{a(\tau_2)}{a(\tau_1)} - 1. \tag{4.3.5}$$

For light from nearby galaxies we have $\tau_2 - \tau_1 \approx R$ where R is the present proper distance to the galaxy. Using the approximation $a(\tau_2) \approx a(\tau_1) + (\tau_2 - \tau_1)\dot{a}$ we can get the linear redshift-distance relationship discovered by Hubble: $z \approx H_0 R$.

4.3.2 Particle horizons

We may ask the question: how much of our universe can be observed at a given event p? Or more preciesly, which isotropic observers could have sent a signal that reaches a given isotropic observer at or before p? The boundary between the world lines that can be seen at p and those that cannot is called the **particle horizon** at p. We may expect that all isotropic observers can communicate with each other near the big bang singularity, but this not the case for Robertson-Walker models which expand sufficiently fast from the initial big bang singularity. We will demonstrate this in the flat spatial geometry with the metric specified in Eq. 4.1.9.

We can make a coordinate transformation $\tau \to \eta$ defined by

$$\eta = \int \frac{d\tau}{a(\tau)} \tag{4.3.6}$$

so that we can express the metric as

$$ds^2 = a^2(\eta)(-d\eta^2 + dx^2 + dy^2 + dz^2). \tag{4.3.7}$$

This metric is just a multiple of the flat Minkowski metric and is called **conformally flat**. The coordinate η is called the **conformal time**. A vector will be timelike, null, or spacelike in Eq. 4.3.7 if and only if it has the same property with respect to the flat Minkowski metric. It is then not difficult to see that an observer at an event p will be able to receive a signal from all other isotropic observers if and only if the integral defining η in Eq. 4.3.6 diverges as $\tau \to 0$ (approaching the big bang singularity), which will be the case if $a(\tau) \leq \alpha\tau$ for some constant α, and there will be no particle horizon. However, if the integral converges, the Robertson-Walker model will be conformally related to only a portion of the Minkowski spacetime above a $\eta =$ constant surface, and there will be particle horizons. In fact, the integral converges for all spatially flat R-W solutions of Einstein's equation.

For the hyperboloid and spherical geometries, the behavior of $a(\tau)$ approaches that of the flat case since the term involving k becomes negligible. In the case of spherical geometry, the spatial extent of the universe is finite and depending on the nature of its components, the particle horizon may cease to exist at some point.

The cosmic microwave background provides strong evidence for the homogeneity and isotropy of the universe. In ordinary systems such as gas in a box, we can explain its homogeneity and isotropy by stating that the particles have time to self-interact and thermalize. However, this reasoning does not apply to a universe with particle horizons. To explain this, we may either postulate that the universe either began in an extremely homogeneous, isotropic state, or had its inhomogeneity and anisotropy damped out at some later time. Both theories have encountered difficulties, and a more widely accepted explanation involves an inflation phase of the universe, which drastically expanded the particle horizon to allow interactions between particles.

4.4 The Evolution of Our Universe

We will outline the history of the universe from the big bang to the present, assuming it is well-described by a Robertson-Walker solution throughout its history. For more detailed descriptions about the theory and observational evidence, see Peebles (1971) and Weinberg (1972).

As we have discussed earlier, the energy densities of radiation and matter scale differently with the scale factor a, and we would expect the energy density contribution of radiation to dominate in the early universe, when the scale factor was much smaller. In fact, at the matter-radiation equality, a was more than 3000 times smaller (i.e. $z \sim 3000$) than its present value. Thus, we would expect a radiation filled model to be a good approximation for the dynamics of the universe before this stage, and a matter/dust filled model to be a good approximation afterwards.

If the early universe was radiation dominated, we would expect that for all spatial geometries, as $a \to 0$, the dependence of a and ρ on τ goes over to the flat solution:

$$a(\tau) = (4C')^{1/4}\tau^{1/2}, \tag{4.4.1}$$

$$\rho = \frac{3}{32\pi G\tau^2}, \tag{4.4.2}$$

where we have restored the constants c and G. If the radiation is thermally distributed, we can

derive the mass density from statistical mechanics:

$$\rho = \sum_{i=1}^{n} \alpha_i g_i \frac{\pi^2}{30\hbar^3 c^5} (kT)^4, \tag{4.4.3}$$

where n is the number of species of radiation, g_i is the spin degeneracy, and α_i is 1 for bosons and 7/8 for fermions. Note that particles with mass much less than kT act like massless particles and are considered a "species of radiation."

Next, we want to get an idea of the time scale at which the matter-radiation interactions occur. This would help us determine whether thermalization occurs locally or not. The expansion time scale t_E of the universe is simply

$$t_E \sim a/\dot{a} = 2\tau. \tag{4.4.4}$$

On the other hand, the interaction time scale is

$$t_I \sim \frac{1}{n\sigma c} \propto \frac{a^3}{\sigma} \propto \tau^{3/2}/\sigma(T), \tag{4.4.5}$$

where the number of interacting particles is assumed to be conserved so that $n \propto a^{-3}$. The above equations show that unless σ falls off rapidly at high energies, we will have $t_I \ll t_E$ at very early times of the universe, and thermalization can be achieved. Eventually, as the temperature drops, we will have $t_I > t_E$, and the matter distribution will drop out of thermal equilibrium with radiation.

During the very beginning of the evolutionary history of the universe predicted by classical general relativity, the spactime curvature was greater than the Planck length, so we should expect quantum effects to play an important role. All statements about the behavior of matter during this epoch are speculative.

It is worth pointing out two important effects that may have occurred shortly after the Planck time. The first involves a phase transition of the thermal equilibrium state of the quantum field of a unified theory of strong and electro-weak interactions. This may have caused the universe to go through a phase in which the dynamics of the universe was dominated by a large, positive cosmological constant, leading to an inflationary phase. The second effect concerns the production of baryons. In the present universe there is a matter-antimatter asymmetry. It may be the case that the universe was simply born with with an excess amount of baryons over antibaryons. However, it is also possible that more baryons were produced in the very early universe. For the second statement to be true, the high energy particle interactions mus satisfy the following properties: (1) they do not conserve baryon number; (2) they do not preserve charge conjugation C and the composition of C with parity, CP; (3) they must result in departures from thermal equilibrium.

At $\tau = 1$ second, the density of the universe was $\rho \approx 5 \times 10^5$ g/cm^3 and the temperature was $T \approx 10^{10}$ K. These are low enough for us to make solid predictions. The matter of the universe consisted almost entirely of photons, neutrinos, electrons, positrons, neutrons, and protons in thermal equilibrium. At this stage, the interactions of neutrinos become weak and they decouple from the rest of the matter. We therefore expect to see a cosmic neutrino background at temperature $T \approx 2$ K.

As the universe continues to cool, the rates of reactions between neutrons and protons drop quickly to below the expansion rate, leading to a "**freeze-out**" of the neutron-proton ratio at 1/6 at $\tau \sim 1.5$ seconds. Then, at $\tau = 4$ seconds, the temperature dropped to approximately the mass of electrons and positrons (0.5 MeV). At this stage the production rate drops below the annihilation

rate, and all positrons are annihilated, heating the photons to a temperature about 1.4 times that of the neutrinos.

Nucleosythesis producing ^{4}He began when the temperature drops to about 10^9 K at $\tau \approx 3$ minutes. Very little nucleosynthesis occurbefore this time before because deuterium, an important ingredient, was not abundant until this time. The large Coulomb barrier and lack of stable isotopes limit the production of heavier nuclei. The percentage of ^{4}He is not very sensitive to the baryon density, and depends mainly on the neutron-proton ratio at "freeze-out." The abundances of other elements are more sensitive to baryon density. This process (**big bang nucleosythesis**) accounts for most of the 25% of helium in the universe.

The next major event, called **recombination**, occurred at $\tau \sim 4 \times 10^5$ years, when the temperature was about 4000 K. At this point the free electrons and protons combined to form neutral hydrogen. The interaction between matter and radiation drops as the scattering cross-section for neutral atoms is much smaller than that of charged particles, and the photons decouples from the matter. This is the origin of the **cosmic microwave background**, a blackbody radiation at temperature $T \approx 2.7$ K filling the universe. Its isotropy at the large scales is strong evidence for the homogeneity and isotropy of the universe at the time of recombination.

As the matter and radiation decouple, gravitational perturbations, no longer inhibited by radiation pressure, started to grow and led to the formation of galaxies. The density fluctuations in the matter distribution that acted as seeds of galaxies are being actively studied. Around the same time ($10^3 \lesssim \tau \lesssim 10^7$ years), matter became the dominant form of energy in the universe. More recently (in the cosmological sense), another form of energy with an equation of state parameter $w \approx -1$ (i.e. the cosmological constant) starts to dominate, and is responsible for the accelerating expansion of the universe.

If one accepts the picture of the universe predicted by general relativity, we will arrive at some constraints on (1) the masses of stable, weakly interacting, elementary particles, and (2) the number of species of massless particles that are in thermal equilibrium in the early universe. As for the first constraint, suppose some massive stable particle (electron, neutrino, etc.) has mass m. The behavior of this particle in the early universe would be the same as that of a massless neutrino. However, in the present universe, it contributes energy m per particle instead of $\sim 10^{-4}$ eV per particle (or $T \sim 2$ K). If the particles are very massive, their population would have been set at a low value when they dropped out of thermal equilibrium and they would not contribute much to the energy density of the present universe. They would be the dominant contributors to the energy density, however, if their mass lies within 100 eV $\lesssim m \lesssim$ 100 GeV.

Because the energy of "massless" particles would get redshifted away as the universe expands, the existence of other species of such particles in thermal equilibrium in the early universe would not contribute much to the present energy density of the universe. However, according to Eq. 4.4.3, they would affect the relation between ρ and T in the early universe and make T smaller for a given ρ. This means a given temperature will occur at an earlier time when the expansion rate is higher. As a result, "freeze-out" occurs at a higher temperature and a higher percentage of neutrons is produced, leading to more helium from the nucleosynthesis. Assuming the baryon density corresponds to the observed value today, it is found that much more helium would be produced if there are more than four species of neutrinos.

The future of the universe depends on its components as well as its geometry. There are many ways to determine whether the universe is open or closed, including the redshift-apparent magnitude relation for distant objects, the present mass density of the universe, the age of the universe, the

cosmic abundance of deuterium, etc. The current observation results indicate that the universe is close to being flat, with a tight upper bound placed on the value of Ω_k.

5 The Schwarzschild Solution

In addition to cosmological observations, we can also test general relativity by measuring the gravitational fields around the sun. Thus, we wish to determine a solution of Einstein's equation corresponding to the exterior of a static, spherically symmetric body, which is a good approximation of the sun and many other bodies. This problem was solved by Karl Schwarzschild in 1916, a few months after the publication of Einstein's vacuum field equations.

As we have mentioned in section 3.4.1, predictions of general relativity reduce to those of Newtonian theory in the slow motion, weak field limit. The Schwarzschild solution, however, accurately predicts the deviations of planetary motion from the Newtonian theory, and as well as gravitational lensing, gravitational redshift, and time delay effects. In addition, the vacuum Schwarzschild solution can describe the entire spacetime geometry around the end product of gravitational collapse, and it contains a spacetime singularity hidden inside a black hole.

5.1 Derivation of the Schwarzschild Solution

We wish to find all four-dimensional Lorentz signature metrics whose Ricci tensor vanishes (because it is associated with a flow of the metric with respect to time)[12] and which are static and possess spherical symmetry. The first task is to give precise definitions to the terms "static" and "spherically symmetric, and to choose a convenient coordinate system.

A spacetime is said to be **stationary** is there exists a one-parameter group of isometries ϕ_t whose orbits are timelike curves. This group expresses time translation symmetry of the spacetime. Equivalently, a stationary spacetime possesses a timelike Killing vector field ξ^a. A spacetime is said to be **static** if it is stationary *and* if there exists a spacelike hypersurface Σ which is orthogonal to the orbits of the isometry. By Frobenius's theorem this is equivalent to

$$\xi_{[a}\nabla_b\xi_{c]} = 0. \tag{5.1.1}$$

We can interpret this extra condition by introducing a convenient coordinates for static spacetimes as follows. If $\xi^a \neq 0$ everywhere on Σ, then in a neighborhood of Σ, every point will lie on a unique orbit of ξ^a which passes through Σ. Assuming $\xi^a \neq 0$, we choose arbitrary coordinates $\{x^\mu\}$ on Σ and label each point in this neighborhood by the parameter t of the orbit which starts from Σ and ends at p, and the coordinates x^1, x^2, x^3 of the orbit at Σ. Since this coordinate system employs the Killing parameter t as one of the coordinates, the metric components in this coordinate basis will be independent of t. Also since the surface Σ_t (the set of points with "time coordinate" t) is the image of Σ under teh isometry ϕ_t, it follows that each Σ_t is also orthogonal to ξ^a. In these coordinates, the metric components take the form

$$ds^2 = -V^2(x^1, x^2, x^3)dt^2 + \sum_{\mu,\nu=1}^{3} h_{\mu\nu}(x^1, x^2, x^3)dx^\mu dx^\nu, \tag{5.1.2}$$

[12] We can also contract the vacuum Einstein equation $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 0$ to get $R - \frac{1}{2}\delta^\mu{}_\mu R = 0$. In four dimensions, $\delta^\mu{}_\mu = 4$, so we obtain $-2R = 0$ and hence $R_{\mu\nu} = 0$.

where $V^2 = -\xi_a \xi^a$, and the absence of $dtdx^\mu$ cross terms expresses the orthogonality of ξ^a with Σ. These cross terms are present for a stationary but nonstatic metric.

From Eq. 5.1.2, we can see the diffeomorphism defined by $t \to -t$ which takes a point on each Σ_t to a point with the same spatial coordinate on Σ_{-t} is an isometry. Thus static spacetimes possess a "time reflection" symmetry in addition to the "time translation" symmetry possessed by stationary spacetimes. Field that are time translation invariant can fail to be time reflection invariant when rotational motion, whose direction would be changed by time reflection, is involved. The failure of Eq. 5.1.1 to hold implies that the neighboring orbits of ξ^a "twist" around each other.

A spacetime is said to be **spherically symmetric** if its isometry group contains a subgroup isomorphic to SO(3), and the orbits of this subgroup are two-dimensional spheres. We may interpret the SO(3) isometries as rotations, and a spherically symmetric spacetime has its metric invariant under rotations. The spacetime metric induces a metric on each orbit 2-sphere which must be a multiple of the metric of a unit 2-sphere due to rotational symmetry. Thus, this induced metric can be completely characterized by the total area A of the 2-sphere, and we introduce a function r defined by $r = (A/4\pi)^{1/2}$. Therefore, in spherical coordinates, the metric on each orbit 2-sphere is the familiar

$$ds^2 = r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{5.1.3}$$

We interpret r as the radius of the sphere in flat, three-dimensional Euclidean space.

If a spacetime is both static ans spherically symmetric, and if the static Killing field ξ^a is unique, then ξ^a must be orthogonal to the orbit 2-spheres, because its rotational invariance requires its projection onto any orbit sphere to be zero. Thus, the orbit spheres lie wholly within the hypersurfaces Σ_t. We choose the coordinates as follows. We select a sphere on $\Sigma = \Sigma_0$ and choose spherical coordinates (θ, ϕ) on it, and "carry" these coordinates to other spheres of Σ by means of geodesics orthogonal to the 2-sphere. Provided that $\nabla_a r \neq 0$, we choose (r, θ, ϕ) as coordinates in Σ_t and (t, r, θ, ϕ) as coordinates for the spacetime according to Eq. 5.1.2. The most general form of such as metric has the form

$$ds^2 = -f(r)dt^2 + h(r)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{5.1.4}$$

It is worth pointing out that, in addition to the breakdown of the spherical coordinates at the north and south poles, this coordinate system breaks down at points where $\xi^a = 0$ or $\nabla_a r = 0$. This occurs in the strong field region of the Schwarzschild solution.

With the form of the metric determined, we have reduced the problem from solving for 10 functions corresponding to the 10 metric components $g_{\mu\nu}$ of 4 variables down to solving for two functions of one variable. We will compute the Ricci tensor of the the metric in Eq. 5.1.4 and solve $R_{ab} = 0$ for f and h. This can be done with the tetrad method introduced in section 2.4.2. A convenient basis is

$$(e_0)_a = f^{1/2}(dt)_a, \tag{5.1.5}$$
$$(e_1)_a = h^{1/2}(dr)_a, \tag{5.1.6}$$
$$(e_2)_a = r(d\theta)_a, \tag{5.1.7}$$
$$(e_3)_a = r\sin\theta(d\phi)_a. \tag{5.1.8}$$

Detailed calculations of the Riemann tensor using the tetrad method can be found on pages 121-123 of Wald. By setting the Ricci tensor to zero, we will arrive at the vacuum Einstein equation for

static, spherically symmetric spacetime:

$$0 = R_{00} = \frac{1}{2}(fh)^{-1/2}\frac{d}{dr}[(fh)^{-1/2}f'] + (rfh)^{-1}f', \tag{5.1.9}$$

$$0 = R_{11} = -\frac{1}{2}(fh)^{-1/2}\frac{d}{dr}[(fh)^{-1/2}f'] + (rh^2)^{-1}h', \tag{5.1.10}$$

$$0 = R_{22} = R_{33} = -\frac{1}{2}(rfh)^{-1}f' + \frac{1}{2}(rh^2)^{-1}h' + r^{-2}(1-h^{-1}), \tag{5.1.11}$$

where $R_{\mu\nu} \equiv R_{ab}(e_\mu)^a(e_\nu)^b$. The off-diagonal components of $R_{\mu\nu}$ vanish due to symmetry. Adding Eq. 5.1.9 and 5.1.10, we obtain

$$\frac{f'}{f} + \frac{h'}{h} = 0, \tag{5.1.12}$$

which implies $f = K/h$, where K is a constant. We may set K to 1 by rescaling the time coordinate $t \to K^{1/2}t$, and Eq. 5.1.11 yields

$$-f' + \frac{1-f}{r} = 0, \text{ i.e., } \frac{d}{dr}(rf) = 1. \tag{5.1.13}$$

This implies $f = 1 + C/r$, where C is a constant. Combining these results we obtain the Schwarzschild solution

$$ds^2 = -\left(1 + \frac{C}{r}\right)dt^2 + \left(1 + \frac{C}{r}\right)^{-1}dr^2 + r^2 d\Omega^2, \tag{5.1.14}$$

where $d\Omega^2 \equiv d\theta^2 + \sin^2\theta d\phi^2$. We should notice that this solution is asymptotically flat, i.e. its metric components approach those of Minkowski spacetime as $r \to \infty$. This means we can interpret the Schwarzschild metric as the exterior gravitational field of an isolated body. By comparing the behavior of a test body in the weak field regime ($r \to \infty$) with the parameter C with the behavior of a test body of M in Newtonian theory, we can show that $M = -C/2$. Thus, we can interpret $-C/2$ as the total mass of the Schwarzschild field, and we can rewrite the **Schwarzschild metric** as

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1}dr^2 + r^2 d\Omega^2. \tag{5.1.15}$$

We should note that the metric components of the Schwarzschild solution become singular in the strong field regime at $R = 0$ and $R = 2M$. This behavior can be attributed to either (i) a breakdown of the coordinates used to obtain the general form of the metric (Eq. 5.1.4) because $\xi^a = 0$ or $\nabla_a r = 0$, or (ii) a true singularity of the spacetime structure. We will see later that the singularity at $r = 2M$ is due to (i), while $r = 0$ is a true, physical singularity. Note that the "singularity" at $r = 2M$ occurs at a numerical value given by

$$r_S = \frac{2GM}{c^2} \approx 3\left(\frac{M}{M_\odot}\right)\text{km}. \tag{5.1.16}$$

Thus, the Schwarzschild radius r_S for an ordinary body such as the sun is well inside the radius of the body, where the vacuum solution is no longer valid. The two singularities are relevant only for bodies that have undergone complete gravitational collapse.

The vacuum Einstein equation can also be solved for a general spherically symmetric spacetime, but it has been shown by Birkhoff (1923) that the Schwarzschild solution is the only solution to this more general system of equations. This result is analogous to the fact that the Coulomb solution is the only spherically symmetric solution of Maxwell's equations in vacuum, which we can interpret as there exists no monopole radiation.

5.2 Interior Solutions

We now look for a static, spherically symmetric solution of Einstein's equation with a perfect fluid stress-energy tensor $T_{ab} = \rho u_a u_b + P(g_{ab} + u_a u_b)$. For the static symmetry of the spacetime to hold, the 4-velocity of the fluid must point in the same direction as the static Killing vector field, i.e.

$$u^a = -(e_0)^a = -f^{1/2}(dt)^a, \tag{5.2.1}$$

where f is the function appearing in Eq. 5.1.4. We want to find solutions that describe the possible interior fluid sources of the exterior Schwarzschild metric, so we will be looking for equations of structure for static, fluid objects such as stars.

We can simply add the stress-energy terms to equations 5.1.9 - 5.1.11 to obtain Einstein's equation with a fluid:

$$\begin{aligned} 8\pi T_{00} = 8\pi\rho = G_{00} &= R_{00} + \frac{1}{2}(R_0{}^0 + R_1{}^1 + R_2{}^2 + R_3{}^3) \\ &= (rh^2)^{-1}h' + r^{-2}(1 - h^{-1}), &(5.2.2) \\ 8\pi T_{11} = 8\pi P = G_{11} &= R_{11} - \frac{1}{2}(R_0{}^0 + R_1{}^1 + R_2{}^2 + R_3{}^3) \\ &= (rfh)^{-1}f' - r^{-2}(1 - h^{-1}), &(5.2.3) \\ 8\pi T_{22} = 8\pi P = G_{22} &= \frac{1}{2}(fh)^{-1/2}\frac{d}{dr}[(fh)^{-1/2}f'] + \frac{1}{2}(rfh)^{-1}f' - \frac{1}{2}(rh^2)^{-1}h'. &(5.2.4) \end{aligned}$$

Note that the first equation involves only h and can be written in the form

$$\frac{1}{r^2}\frac{d}{dr}[r(1 - h^{-1})] = 8\pi\rho, \tag{5.2.5}$$

and the solution for h is

$$h(r) = \left[1 - \frac{2m(r)}{r}\right]^{-1}, \text{ where } m(r) = 4\pi\int_0^r \rho(r')r'^2 dr' + a, \tag{5.2.6}$$

where a is a constant. For the metric on Σ to be smooth at $r = 0$, we require that $h(r) \to 1$ as $r \to 0$. Thus, in order to avoid a "conical singularity" in the metric at $r = 0$, we set $a = 0$. Since Σ must be spacelike for a static configuration, the necessary condition for staticity is $h \geq 0$, i.e. $r \geq 2m(r)$.

If $\rho = 0$ for $r > R$, the solution for h (Eq. 5.2.6) joins the vacuum Schwarzschild solution with total mass $M = m(R)$. This is formally identical to the expression for total mass in Newtonian gravity. However, we must note that the proper volume element on Σ is $\sqrt{{}^{(3)}g}d^3x = h^{1/2}r^2\sin\theta dr d\theta d\phi$, so the *proper mass* is

$$M_p = 4\pi\int_0^R \rho(r)r^2\left[1 - \frac{2m(r)}{r}\right]^{-1/2} dr. \tag{5.2.7}$$

We can interpret the difference between M and M_p as the gravitational binding energy of the configuration: $E_B = M_p - M$, which is always positive since $M_p > M$.

If we write $f = e^{2\phi}$, then Eq. 5.2.3 becomes

$$\frac{d\phi}{dr} = \frac{m(r) + 4\pi r^3 P}{r[r - 2m(r)]}. \tag{5.2.8}$$

In the Newtonian limit, $r^3P \ll m(r)$ and $m(r) \ll r$, so this reduces to the spherically symmetric Poisson's equation for the Newtonian gravitational potential: $\frac{d\phi}{dr} \approx \frac{m(r)}{r^2}$. Thus, in the static spherically symmetric case, $\phi = \frac{1}{2}\ln f$ is the general relativistic analog of the Newtonian potential. However, there is no known analog for nonstationary configurations.

Substituting Eq. 5.2.6 and 5.2.8 into Eq. 5.2.4, we will obtain an equation for dP/dr. I will skip the algebra here. The result is

$$\frac{dP}{dr} = -(P+\rho)\frac{m(r)+4\pi r^3 P}{r[r-2m(r)]}. \tag{5.2.9}$$

This is known as the **Tolman-Oppenheimer-Volkoff equation** of hydrostatic equilibrium. In the Newtonian limit ($P \ll \rho, m(r) \ll r$), it reduces to $\frac{dP}{dr} \approx -\frac{\rho m(r)}{r^2}$.

In summary, the interior metric of a static, spherical fluid star is

$$ds^2 = -e^{2\phi}dt^2 + \left(1-\frac{2m(r)}{r}\right)^{-1}dr^2 + r^2 d\Omega^2, \tag{5.2.10}$$

wher $m(r)$ is defined the same way as before and ϕ is determined from Eq. 5.2.8.

Thus, we can determine the equilibrium condition for fluid matter with a given equation of state $P = P(\rho)$ mostly in the same way as in Newtonian gravity: We choose a central density ρ_c and hence a central pressure P_c, and integrate Eq. 5.2.9 outward to $P = \rho = 0$, at which point we join this with the vacuum Schwarzschild solution and solve for ϕ using 5.2.8.

The most important difference between equilibrium configurations in GR and Newtonian gravity is that, assuming non-negative pressure, for a given density profile $\rho(r) \geq 0$ the right hand side of Eq. 5.2.9 is always larger in magnitude than the right hand side of the Newtonian equation. This means for a given density profile, the central pressure required for equilibrium is always higher in GR than in Newtonian gravity, so it is harder to maintain equilibrium in GR. Consider a star of uniform density ρ_0 and radius R. In both theories, $m(r) = \frac{4}{3}\pi r^3\rho_0$. In the Newtonian case, $P(r) = \frac{2}{3}\pi\rho_0^2(R^2-r^2)$, so the central pressure is $P_c = \frac{2}{3}\pi\rho_0^2R^2 = \left(\frac{\pi}{6}\right)^{1/3}M^{2/3}\rho_0^{4/3}$, which is finite for all values of ρ_0 and R. This means equilibrium can always be achieved. On the other hand, the pressure profile in general relativity is found to be

$$P(r) = \rho_0\left[\frac{(1-2M/R)^{1/2}-(1-2Mr^2/R^2)^{1/2}}{(1-2Mr^2/R^3)^{1/2}-3(1-2M/R)^{1/2}}\right], \tag{5.2.11}$$

and the central pressure is

$$P(r) = \rho_0\left[\frac{1-(1-2M/R)^{1/2}}{3(1-2M/R)^{1/2}-1}\right]. \tag{5.2.12}$$

This reduces to the Newtonian value for $R \gg M$. However, P_c could become infinite when $R = \frac{9}{4}M$, so the maximum possible mass for a star of uniform density in GR is

$$M_{\max} = \frac{4}{9(3\pi\rho_0)^{1/2}}. \tag{5.2.13}$$

This result does not only hold for the uniform density case, and we will derive the upper mass limit for static, spherical stars with a fixed radius R.

First we should point out that the staticity condition for h already implies an upper mass limit $M \leq R/2$, but this can be sharpened using the condition $f \geq 0$ which states that the Killing field ξ^a is timelike everywhere. Assuming only that $\rho \geq 0$ and $d\rho/dr \leq 0$, we can take the difference of Eq. 5.2.3 and 5.2.4 to obtain

$$0 = G_{11} - G_{22} = \frac{1}{2}(rfh)^{-1}f' - r^{-2}(1-h^{-1}) + \frac{1}{2}(rh^2)^{-1}h' - \frac{1}{2}(fh)^{-1/2}\frac{d}{dr}[(fh)^{-1/2}f']. \quad (5.2.14)$$

Substituting the solution for h in the second and third terms, and use the fact that the average density ($\propto m(r)/r^3$) decreases monotonically with r, we have

$$\frac{d}{dr}\left[r^{-1}h^{-1/2}\frac{df^{1/2}}{dr}\right] \leq 0. \quad (5.2.15)$$

Integrate this inequality inward from R to r, then multiply it by $rh^{1/2}$ and integrate inward again from R to 0 to get

$$f^{1/2}(0) \leq (1-2M/R)^{1/2} - \frac{M}{R^3}\int_0^R \left[1 - \frac{2m(r)}{r}\right]^{-1/2} r dr. \quad (5.2.16)$$

The condition $d\rho/dr \leq 0$ implies that $m(r)$ cannot be smaller than the value it would have for a uniform density star, so $m(r) \geq Mr^3/R^3$. Thus, we obatain

$$f^{1/2}(0) \leq \frac{3}{2}(1-2M/R)^{1/2} - \frac{1}{2}. \quad (5.2.17)$$

The condition $f^{1/2}(0) \geq 0$ then implies

$$M \leq 4R/9. \quad (5.2.18)$$

In addition to the upper mass limit at a fixed radius, we can also find an upper mass limit for a given equation of state below density ρ_0. Since $d\rho/dr \leq 0$, stars whose density fails to be less than ρ_0 must have a "core" of mass m_0 and radius r_0 where $\rho \geq \rho_0$, surrounded by an "envelope" where $\rho < \rho_0$. Given the equation of state for $\rho < \rho_0$, the total mass M is determined by the parameters m_0 and r_0. Since the the core density is at least ρ_0, the lower mass limit for the core is

$$m_0 \geq \frac{4}{3}\pi r_0^3 \rho_0. \quad (5.2.19)$$

Using the same argument as before, we can find an upper mass limit for the core:

$$m_0 \leq \frac{2}{9}r_0[1 - 6\pi r_0^2 P_0 + (1 + 6\pi r_0^2 P_0)^{1/2}], \quad (5.2.20)$$

where $P_0 = P(\rho_0)$ is the pressure at the core-envelope boundary. Eq. 5.2.19 and 5.2.20 restrict m_0 and r_0 to the compact region of the m_0-r_0 plane, so $M(m_0, r_0)$ is a continuous function defined on a compact set and therefore M is bounded. This upper mass limit for a given equation of state also exists in Newtonian gravity. The difference is that in GR, at sufficiently high densities, the limit is independent of of the equation of state.

For cold matter at densities much less than the nuclear density ($\sim 10^{14}$ g cm^{-3}), electron degeneracy pressure is the dominant source of pressure (the similar neutron degeneracy pressure

would dominate near the nuclear density). At "low" densities ($n \ll m_e^3 c^3/\hbar^3 \sim 10^{-31}$ cm^{-3}) and temperature $T = 0$, it provides a pressure of

$$P = \frac{\hbar^2 (3\pi^2)^{2/3}}{5 m_e} n^{5/3}. \tag{5.2.21}$$

At high densities ($n \gg m_e^3 c^3/\hbar^3$) the pressure is

$$P = \frac{\hbar c (3\pi^2)^{1/3}}{4} n^{4/3}. \tag{5.2.22}$$

Cold bodies that comprise stable equilibrium configurations supported by electron degeneracy pressure are know as **white dwarfs**, whose maximum mass is given by (Chandrasekhar 1939)

$$M_C \approx 1.4 \left(\frac{2}{\mu_N}\right)^2 M_\odot, \tag{5.2.23}$$

where μ_N is the number of nucleons per electron. In Newtonian theory with a fixed μ_N, the mass monotonically approaches the value given by Eq. 5.2.23 with $\rho_c \to \infty$ and $R \to 0$. In GR, however, the mass begins to decrease with ρ_c at some finite value of ρ_c, although M_C is not significantly changed. At this point, the configuration becomes unstable again, and μ_N increases as protons and electrons are converted into neutrons at high density. This leads to **neutron stars**, which are supported by neutron degeneracy pressure in a similar way. The precise upper mass limit for neutrons stars is uncertain.

5.3 Geodesics of Schwarzschild: Gravitational Redshift, Perihelion Precession, Bending of Light, and Time Delay of Radar Signals

In previous sections we have discussed the vacuum Schwarzschild solution and the interior of a static, spherically symmetric star. In this section, we will analyze the behavior of test bodies and light rays in the exterior ($r > 2M$) of the Schwarzschild solution. The geodesics in the weak field regime ($r \gg M$) are applicable to non-compact objects such as the sun.

We may use the invariance of the inner product $u^a \xi_a$ of a Killing field ξ^a and a geodesic tangent u^a along the geodesic to spare us the labor to solve the geodesic equation (Eq. 2.3.3). First of all, this allows us to derive a formula for the gravitational redshift. This derivation is similar to that of the cosmological redshift derived in section 4.3.1.

Consider two static observers (observers whose 4-velocity is tangent to the static Killing field $\xi^a = (\partial/\partial t)^a$) O_1 and O_2 with 4-velocities u_1^a and u_2^a. Suppose O_1 emits a signal at event P_1 which is received by O_2 at event O_2. In the geometrical optics approximation this signal travels on a null geodesic with tangent k^a. The frequency of emission is $\omega_1 = -(k_a u_1^a)|_{P_1}$ and the measured frequency is $\omega_2 = -(k_a u_2^a)|_{P_2}$. However, since both 4-velocities are unit vectors pointing in the direction of the timelike ξ^a, we have

$$u_1^a = [\xi^a/(-\xi^b \xi_b)^{1/2}]|_{P_1}, \tag{5.3.1}$$
$$u_2^a = [\xi^a/(-\xi^b \xi_b)^{1/2}]|_{P_2}. \tag{5.3.2}$$

By the invariance of the inner product, we have $(k_a \xi^a)|_{P_1} = (k_a \xi^a)|_{P_2}$, so

$$\frac{\omega_1}{\omega_2} = \frac{(-\xi^b \xi_b)^{1/2}|_{P_2}}{(-\xi^b \xi_b)^{1/2}|_{P_1}} = \frac{(1 - 2M/r_2)^{1/2}}{(1 - 2M/r_1)^{1/2}}, \tag{5.3.3}$$

where we used $\xi^b\xi_b = g_{tt} = -(1-2M/r)$ for Schwarzschild spacetime. Eq. 5.3.3 shows that for $r_2 > r_1$ (emitter closer to the center of gravitational attraction), $\omega_2 < \omega_1$. This makes sense because the photon would lose energy as it climbs out of the gravitational well. In the case of the exterior region of ordinary bodies ($M \ll r_1, r_2$), Eq. 5.3.3 becomes

$$\frac{\Delta\omega}{\omega} \approx -\frac{GM}{c^2 r_1} + \frac{GM}{c^2 r_2}. \tag{5.3.4}$$

We may interpret this as the change in the locally measured energy of a photon is equal to the change in its Newtonian gravitational energy.

We have shown in section 5.2 that the maximum value of M/R is 4/9. Thus, by Eq. 5.3.3 the maximum redshift of light emitted from the surface of a static star is

$$\frac{\omega_1}{\omega_2}\Big|_{\max} = \frac{\omega(r = 9M/4)}{\omega(r=\infty)} = 3; \text{ or } z_{\max} = \frac{\omega_1}{\omega_2}\Big|_{\max} - 1 = 2. \tag{5.3.5}$$

This means observed redshifts of greater than 2 cannot solely be attributed to this gravitational redshift.

To solve the timelike and null geodesics, we first note that because of the parity reflection symmetry $\theta \to \pi - \theta$ or the Schwarzschild metric, if the initial position and tangent vector of a geodesic lie in the equatorial plane $\theta = \pi/2$, then the entire geodesic should lie in this plane. Since every geodesic can be brought to an initially equatorial geodesic by a rotational isometry, we may restrict our discussion to the equatorial geodesics without loss of generality.

The coordinate basis components of the tangent u^a to a curve parameterized by τ are $u^\mu = \frac{dx^\mu}{d\tau} \equiv \dot{x}^\mu$. For timelike geodesics, we choose τ to be the proper time; for null geodesics, τ is an affine parameter. Thus, we have

$$-\kappa = g_{ab}u^a u^b = -(1-2M/r)\dot{t}^2 + (1-2M/r)^{-1}\dot{r}^2 + r^2\dot{\phi}^2, \tag{5.3.6}$$

where

$$\kappa = \begin{cases} 1 \text{ (timelike geodesics)} \\ 0 \text{ (null geodesics).} \end{cases} \tag{5.3.7}$$

In the derivation of the gravitational redshift, we used the fact that the quantity

$$E = -g_{ab}\xi^a u^b = (1-2M/r)\dot{t} \tag{5.3.8}$$

is a constant of motion. We may interpret E for timelike geodesics as representing the total energy per unit rest mass of a particle following the geodesic in question, relative to a static observer at infinity. This is the energy that would be required for such an observer to put a unit rest mass particle in the given orbit. Similarly, in the null case, $\hbar E$ represents the total energy of a photon.

The rotational Killing field $\psi^a = (\partial/\partial\phi)^a$ also yields a constant of motion:

$$L = g_{ab}\psi^a u^b = r^2\dot{\phi}. \tag{5.3.9}$$

We may interpret L as the angular momentum per unit rest mass of a particle in the timelike case, and $\hbar L$ as the angular momentum of a photon in the null case. In the Newtonian limit where the geometry is Euclidean, Eq. 5.3.9 is just Kepler's second law.

Substituting Eq. 5.3.8 and 5.3.9 in Eq. 5.3.6, we obtain the final equation for geodesics

$$\frac{1}{2}\dot{r}^2 + \frac{1}{2}\left(1 - \frac{2M}{r}\right)\left(\frac{L^2}{r^2} + \kappa\right) = \frac{1}{2}E^2. \tag{5.3.10}$$

This shows that the radial motion of a geodesic is the same as that of a unit mass particle of energy $E^2/2$ in ordinary 1-dimensional, nonrelativistic mechanics moving in the effective potential

$$V = \frac{1}{2}\kappa - \kappa\frac{M}{r} + \frac{L^2}{2r^2} - \frac{ML^2}{r^3}. \tag{5.3.11}$$

Eq. 5.3.8, 5.3.9, and 5.3.11 determine the time coordinate change, angular motion, and radial motion of the particle. It is worth noting that in addition to the Newtonian term $-\kappa M/r$ and the centrifugal term $L^2/2r^2$, we now have a new term $-ML^2/r^3$ which dominates over the centrifugal term at small r.

Let us first consider timelike geodesics ($\kappa = 1$). We will find that extremizing V yields the roots

$$R_\pm = \frac{L^2 \pm \sqrt{L^4 - 12L^2M^2}}{2M}. \tag{5.3.12}$$

If $L^2 < 12M^2$, there are no extrema of V. Such a particle heading toward the center of attraction ($\dot{r} \leq 0$) will fall directly to the $r = 2M$ surface and continue to fall into the spacetime singularity at $r = 0$.

For $L^2 > 12M^2$, R_+ is a minimum and R_- is a maximum. Thus, stable circular orbits ($\dot{r} = 0$) exist at $r = R_+$ and unstable circular orbits exist at $r = R_-$. For $L \ll M$, we have $R_+ \approx L^2/M$, which is just the Newtonian formula for the radius of a circular orbit. This justifies our earlier interpretation of the constant C as $-2M$ in Eq. 5.1.15. According to Eq. 5.3.12, R_+ is restricted to the range $R_+ > 6M$ (no stable circular orbits exist inside $6M$) and R_- is restricted to $3M < R_- < 6M$ (no circular orbits exist inside $3M$).

The energy of an ordinary particle in 1-dimensional motion which sits at the minimum/maximum of V is just the value of V at that point. Thus, from Eq. 5.3.10, the true energy per unit rest mass E of a particle in a circular orbit of radius R is

$$E(R) = \frac{R - 2M}{R^{1/2}(R - 3M)^{1/2}}. \tag{5.3.13}$$

Note that if $R \leq 4M$, we have $E \geq 1$ and $E \to \infty$ as $R \to 3M$. Thus, particles in the unstable orbits between $3M$ and $4M$ would escape to infinity if perturbed radially outward.

The binding energy E_B per unit rest mass of the last stable circular orbit at $R = 6M$ is $E_B = 1 - E = 1 - (8/9)^{1/2} \approx 0.06$. As discussed in section 3.4.2, a particle orbiting in the Schwarzschild geometry will emit gravitational radiation. Because of radiation reaction, it will deviate slightly from geodesic motion. A particle initially in a circular orbit with $R \gg M$ ($E \approx 1$) should slowly spiral in to smaller radii as it loses energy through radiation, remaining in a nearly circular orbit until it reaches $R = 6M$, after which the orbit becomes unstable and the particle falls rapidly to $r = 0$. According to the calculation of E_B above, about 6% of the original mass-energy of the particle will be radiated away as it spirals to $R = 6M$.

For sufficiently small displacements from the equilibirum radius R_+, a particle will oscillate in simple harmonic motion about R_+ with frequency ω_r

$$\omega_r^2 = k_{\text{eff}} = \frac{d^2V}{dr^2}\Big|_{R_+} = \frac{M(R_+ - 6M)}{R_+^3(R_+ - 3M)}. \tag{5.3.14}$$

Note that the time implicit in ω_r is the proper time τ measured by the particle. On the other hand, the angular frequency $\omega_\phi = \dot{\phi}$ of a circular orbit is given by

$$\omega_\phi^2 = \frac{L^2}{R_+^4} = \frac{M}{R_+^2(R_+ - 3M)}. \tag{5.3.15}$$

In the Newtonian limit ($R_+ \gg M$), $\omega_r \approx \omega_\phi$. If the two frequencies are the same, we will have closed orbits. The fact that the two do not match in general relativity means that the orbit is not closed and there is a precession of the angle at which the maximum and minimum values of r are achieved. For nearly circular orbits, this precession rate is given by

$$\omega_p = \omega_\phi - \omega_r = -[(1 - 6M/R_+)^{1/2} - 1]\omega_\phi. \tag{5.3.16}$$

In the Newtonian limit and to the lowest nonvanishing order, this reduces to

$$\omega_p \approx \frac{3M^{3/2}}{R_+^{5/2}} = \frac{3(GM)^{3/2}}{c^2 R_+^{5/2}}. \tag{5.3.17}$$

A more general analysis (Weinberg 1972) gives the precession rate of an arbitrary elliptical orbit

$$\omega_p \approx \frac{3(GM)^{3/2}}{c^2(1 - e^2)a^{5/2}}, \tag{5.3.18}$$

where a is the semimajor axis and e is the eccentricity.

Now let us consider the case of null geodesics ($\kappa = 0$). According to Eq. 5.3.10, the effective potential for null geodesics is

$$V = \frac{L^2}{2r^3}(r - 2M). \tag{5.3.19}$$

The shape of V is independent of L and it has only one maximum at $r = 3M$. Thus, photons have unstable circular orbits at $r = 3M$, but no stable circular orbits. The minimum energy E required to overcome the potential barrier is given by

$$\frac{1}{2}E^2 = V(R = 3M) = \frac{L^2 M}{2(3M)^3}, \text{ or } \frac{L^2}{E^2} = 27M^2. \tag{5.3.20}$$

For a light ray propagating in flat spacetime, the impact parameter (distance of closest approach to $r = 0$) is L/E. Since the Schwarzschild geometry is asymptotically flat, for a light ray initially in the $r \gg M$ region, we can define the **apparent impact parameter**

$$b =\equiv \frac{L}{E}, \tag{5.3.21}$$

although this no longer represents the distance of closest approach. Thus, in the Schwarzschild geometry, any photon with an apparent impact parameter smaller than the critical value $b_c = 3^{3/2}M$ will be captured. Hence the cross section for photons in the Schwarzschild geometry is

$$\sigma = \pi b_c^2 = 27\pi M^2. \tag{5.3.22}$$

To find the angle $\Delta\phi = \phi_{+\infty} - \phi_{-\infty}$ for photons that are not captured, we can use Eq. 5.3.9 and 5.3.10 to derive an expression for the light bending effects of the Schwarzschild geometry:

$$\frac{d\phi}{dr} = \frac{L}{r^2}\left[E^2 - \frac{L^2}{r^3}(r-2M)\right]^{-1/2}. \tag{5.3.23}$$

In order to not be captured, the impact parameter must be greater than the critical value b_c, and the orbit of the light must have a turning point at the largest radius R_0 for which $V(R_0) = E^2/2$, i.e. at the largest root of

$$R_0^3 - b^2(R_0 - 2M) = 0, \tag{5.3.24}$$

which gives

$$R_0 = \frac{2b}{\sqrt{3}}\cos\left[\frac{1}{3}\arccos\left(-\frac{3^{3/2}M}{b}\right)\right]. \tag{5.3.25}$$

By symmetry, the contributions to $\Delta\phi$ before and after the turning point should be equal, and we have

$$\Delta\phi = 2\int_{R_0}^{\infty} \frac{dr}{[r^4 b^{-2} - r(r-2M)]^{1/2}}. \tag{5.3.26}$$

With a change of variables $u = 1/r$, we can find that in the case of flat spacetime ($M = 0, R_0 = b$), we have $\Delta\phi|_{M=0} = 2\arcsin(b/R_0) = \pi$, i.e. a straight line. For the case of curved spacetime, if we want to find $\Delta\phi$ to first order in M, we may use M and R_0 as independent variables, i.e. we compare $Delta\phi$ for lights rays which have the same radial coordinate R_0 at the closest approach rather than with the same apparent impact parameter. Thus we obtain

$$\Delta\phi = 2\int_0^{1/R_0} \frac{du}{(R_0^{-2} - 2MR_0^{-3} - u^2 + 2Mu^3)^{1/2}}. \tag{5.3.27}$$

Differentiating this with respect to M at a fixed R_0 and evaluating the result at $M = 0$, we find, to first order in M, the deflection angle is

$$\delta\phi = \Delta\phi - \pi \approx M\frac{\partial(\Delta\phi)}{\partial M}\Big|_{M=0} = \frac{4GM}{bc^2}. \tag{5.3.28}$$

Another measurable effect concerning the null geodesics is the time delay of radar signals emitted from Earth. From Eq. 5.3.8 and 5.3.10 we can obtain

$$\frac{dt}{dr} = \left(1 - \frac{2M}{r}\right)^{-1}\left[1 - \left(1 - \frac{2M}{r}\right)\frac{b^2}{r^2}\right]^{-1/2}. \tag{5.3.29}$$

We can integrate this equation over the trajectory of a null geodesic to obtain the total change Δt in the Schwarzschild time coordinate along the trajectory. Consider a radar signal emitted from Earth, located at R_E. The signal passes near the sun with radius of closest approach R_0 and is reflected off a planet located at R_p and retraces its trajectory back to Earth. We wish to find, to first order in M, the time $\Delta\tau$ experienced by an observer on Earth between the emission and reception of the signal. Similar to the light bending analysis, we will obtain

$$\begin{aligned}\Delta t &= \frac{2}{c}[(R_E^2 - R_0^2)^{1/2} + (R_p^2 - R_0^2)^{1/2}] + \frac{2GM}{c^3}\left\{2\ln\left[\frac{R_E + (R_E^2 - R_0^2)^{1/2}}{R_0}\right]\right. \\ &+ \left. 2\ln\left[\frac{R_p + (R_p^2 - R_0^2)^{1/2}}{R_0}\right] + \left(\frac{R_E - R_0}{R_E + R_0}\right)^{1/2} + \left(\frac{R_p - R_0}{R_p + R_0}\right)^{1/2}\right\}.\end{aligned} \tag{5.3.30}$$

The proper time elapsed on Earth is given by $\Delta\tau = (1 - 2M/R_E)^{1/2}\Delta t$. Thus, to first order in M, we have

$$\Delta\tau = -\frac{2GM}{c^3 R_E}[(R_E^2 - R_0^2)^{1/2} + (R_p^2 + R_0^2)^{1/2}] + \Delta t. \tag{5.3.31}$$

5.4 The Kruskal Extension

As we have discussed in section 5.2, analysis of the singularities at $r = 2M$ and $r = 0$ for the vacuum Schwarzschild solution is irrelevant to the study of the gravitational field of a static star as they are well within the matter-filled interior. However, this becomes important when we seek to describe the endpoint of gravitational collapse.

Whenever the metric components in a coordinate basis are badly behaved for certain values of the coordinates, there are two possible causes: (1) the spacetime is singular; or (2) the spacetime is not singular but the coordinates fail to properly cover a region of spacetime. Normally, possibility (1) can be demonstrated by calculating a curvature scalar such as $R_{abcd}R^{abcd}$ and showing that it blows up at the singularity, although exceptions exist. This singularity also lies at a finite affine parameter along some geodesic, because a "singularity" at infinity is not really a singularity. We can demonstrate possibility (2) by displaying an extension to the nonsingular region of the original metric, i.e. a spacetime that includes the original spacetime as a proper subset, through a coordinate transformation. For the Schwarzschild metric, the coordinates are not well-behaved where the timelike Killing field ξ^a becomes collinear with $\nabla^a r$. This occurs at $r = 0$ and $r = 2M$.

Let us look at two examples. Consider the two dimensional metric

$$ds^2 = -\frac{1}{t^4}dt^2 + dx^2 \tag{5.4.1}$$

defined over the range $-\infty < x < \infty, 0 < t < \infty$. This metric appears to have a singularity at $t = 0$. However, if we make a coordinate transformation $t \to t' = 1/t$, we can obtain the flat metric $ds^2 = -(dt')^2 + dx^2$. The original spacetime is then just the $t' > 0$ portion of Minkowski spacetime, and $t = 0$ of the original metric just represents $t' \to \infty$ in Minkowski spacetime. It is the result of a covering of an infinite region of spacetime with a finite range of a coordinate. The spacetime geometry is geodesically complete as $t \to 0 (t' \to \infty)$, i.e. all the geodesics approaching $t = 0$ extend to arbitrarily large values of their affine parameter (extendible). On the other hand, the original metric is not geodesically complete as $t \to \infty (t' \to 0)$, but we can extend the original metric "beyond $t = 0$" by adding the portion $t' \leq 0$ of Minkowski spacetime. We can see from this example that coordinate labels may not be physically meaningful quantities.

For the second example, let us consider the **Rindler spacetime**,

$$ds^2 = -x^2 dt^2 + dx^2 \tag{5.4.2}$$

with ranges $-\infty < t < \infty, 0 < x < \infty$. This metric appears to have a singularity at $x = 0$. Geodesics terminate at $x = 0$ with finite length, but the curvature scalars are well behaved as $x \to 0$, suggesting a coordinate singularity. It is not easy to guess a coordinate transformation for this metric, so instead we start by introducing new coordinates which are linked closely with to the spacetime geometry. For example, we can choose a family of geodesics that head toward the singularity and use the affine parameter along the geodesics as one of the coordinates. This method, however, does not work at all times, and new coordinate singularities may appear whenever the geodesics cross. Nevertheless, in two-dimensional spacetimes, we have a foolproof method to

eliminate coordinate singularities because the null geodesics divide up into two classes - "ingoing" and "outgoing" - and within each class two distinct geodesics cannot cross, since their tangents would have to coincide at the intersection, implying that the geodesics coincide everywhere. We can introduce null coordinates, for which the two coordinates are constant along each "ingoing" or "outgoing" geodesic, respectively. Thus, the coordinate grid will be based on teh geometrical grid of null geodesics, and in this case the only coordinate singularities arise from bad parmeterization of the geodesics. This can be corrected by comparing with an affine parameterization.

The null geodesics of Rindler spacetime can be found from the null condition

$$0 = g_{ab}k^a k^b = -x^2\dot{t}^2 + \dot{x}^2, \tag{5.4.3}$$

where k^a is teh geodesic tangent and the dot denotes derivative with respect to the affine parameter. This equation implies $(dt/dx)^2 = 1/x^2$, so that along each geodesic, we have $t = \pm \ln x + \text{constant}$, where the plus/minus sign refers to the "outgoing"/"ingoing" geodesics. Thus, we can define the null coordinates (u, v) by

$$u = t - \ln x, \tag{5.4.4}$$
$$v = t + \ln x. \tag{5.4.5}$$

In these coordinates, the metric components are

$$ds^2 = -e^{v-u}dudv. \tag{5.4.6}$$

Since the coordinate ranges $-\infty < u < \infty$ and $-\infty < v < \infty$ still only cover the $x > 0$ region, we are finished yet. We can extend the spacetime beyond $x = 0$ (or, beyond "$u = \infty$" and "$v = -\infty$") by a reparameterization of coordinates $U = U(u), V = V(v)$. We calculate the affine parameter along the null geodesics. Note that the time translation vector $(\partial/\partial t)^a$ of Eq. 5.4.2 is a Killing field. Therefore,

$$E = -g_{ab}k^a(\partial/\partial t)^b = x^2 dt/d\lambda \tag{5.4.7}$$

is a constant of the motion, where λ is the affine parameter. For the outgoing null geodesics, setting u constant, we find

$$\lambda = \frac{1}{2E}\int e^{v-u}dv = C + (e^{-u}/2E)e^v, \tag{5.4.8}$$

wher C is a constant. Thus, $\lambda_{\text{out}} = e^v$ is an affine parameter along the outgoing geodesics. Similarly, we can show that $\lambda_{\text{in}} = e^{-u}$ is an affine parameter along the ingoing geodesics. Therefore, we make the transformation $U = -e^{-u}, V = e^v$, which results in

$$ds^2 = -dUdV. \tag{5.4.9}$$

The original Rindler spacetime corresponds to the ranges $U < 0, V > 0$, but there is no longer any singularity in the metric components at $U = 0$ or $V = 0$, so we can extend the spacetime by allowing $-\infty < U < \infty$ and $-\infty < V < \infty$. Finally, we make a transformation $T = (U+V)/2, X = (V-U)/2$ to convert the metric into the familiar form

$$ds^2 = -dT^2 + dX^2. \tag{5.4.10}$$

This shows the extended spacetime is just the Minkowski spacetime. The original coordinates (t, x) are related to the final Minkowski coordinates (T, X) by

$$x = \sqrt{X^2 - T^2}, \tag{5.4.11}$$
$$t = \tanh^{-1} T/X. \tag{5.4.12}$$

From these equations we see that Rindler spacetime is the wedge $X > |T|$ (region I) of Minkowski spacetime. The nature of the coordinate singularity is revealed: the null lines $X = \pm T$ are mislabeled by the original coordinates as $x = 0, t = \pm\infty$.

Note that the time translation symmetry of the Rindler metric (Eq. 5.4.2) corresponds to the boost symmetry of Minkowski spacetime. The observers at constant x undergo the uniform acceleration $a = 1/x$, which diverges as $x \to 0$. We can check that static observers in the Schwarzschild spacetime, in order to stand still in the gravitational field, must undergo a proper acceleration $a = (1 - 2M/r)^{-1/2} M/r^2$, which diverges as $r \to 2M$. This is analogous to the behavior of the Rindler time coordinate as $x \to 0$.

The Schwarzschild spacetime is four-dimensional, but because of spherical symmetry, we only need to look at the two-dimensional "r-t part" of the metric

$$ds^2 = -(1 - 2M/r)dt^2 + (1 - 2M/r)^{-1} dr^2. \tag{5.4.13}$$

The null condition, analogous to Eq. 5.4.3, is

$$0 = g_{ab} k^a k^b = -(1 - 2M/r)\dot{t}^2 + (1 - 2M/r)^{-1} \dot{r}^2, \tag{5.4.14}$$

which implies

$$\left(\frac{dt}{dr}\right)^2 = \left(\frac{r}{r - 2M}\right)^2. \tag{5.4.15}$$

Thus, the radial null geodesics of Schwarzschild satisfy $t = \pm r_* + \text{constant}$, where r_* is the "**Regge-Wheeler tortoise coordinate**," given by $r_* = r + 2M \ln(r/2M - 1)$ such that $dr_*/dr = (1 - 2M/r)^{-1}$. We define the null coordinates u, v by

$$u = t - r_*, \tag{5.4.16}$$
$$v = t + r_*. \tag{5.4.17}$$

In these coordinates, the metric is $ds^2 = -(1 - 2M/r)dudv$, where r is a function of u and v, defined implicitly by $r_*(r) = (v - u)/2$. Using this definition, we can rewrite the metric as

$$ds^2 = -\frac{2Me^{-r/2M}}{r} e^{(v-u)/4M} dudv, \tag{5.4.18}$$

where the metric components are factored into $e^{-r/2M}/r$, which is not singular as $r \to 2M$ ($u \to \infty$ or $v \to -\infty$). Similar to the Rindler case, we make the transformation

$$U = -e^{-u/4M}, \tag{5.4.19}$$
$$V = e^{v/4M}, \tag{5.4.20}$$

and the metric becomes

$$ds^2 = -\frac{32M^3 e^{-r/2M}}{r} dUdV. \tag{5.4.21}$$

Now the singularity at $r = 2M$ (at $U = 0$ or $V = 0$) is removed, and we can extend the Schwarzschild solution to all values of U and V that are compatible with $r > 0$. One final transformation $T = (U + V)/2, X = (V - U)/2$ gives the Schwarzschild metric in the final form given by Kruskal (1960)

$$ds^2 = \frac{32M^3 e^{-r/2M}}{r}(-dT^2 + dX^2) + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{5.4.22}$$

The relation between the old coordinates (t, r) and the new coordinates (T, X) is given by

$$\left(\frac{r}{2M} - 1\right) e^{r/2M} = X^2 - T^2, \tag{5.4.23}$$

$$\frac{t}{2M} = \ln\left(\frac{T + X}{X - T}\right) = 2\tanh^{-1}(T/X). \tag{5.4.24}$$

The allowed range of X and T is given by the condition $r > 0$, which yields $X^2 - T^2 > -1$.

By construction, the radial null geodesics are 45° in Kruskal coordinates, and the Kruskal extension is remarkably similar to the extension of the Rindler spacetime, with the major differences being (1) the Schwarzschild spacetime is four-dimensional, so each point in the X-T diagram represents a two-dimensional sphere of radius r; (2) there are physical singularities in the extended region at $X = \pm(T^2 - 1)^{1/2}$. Note that these singularities have a spacelike character and exist in the future of region II and the past of region III (Figure 6.9). Naively, we may think that the singularities correspond to a timelike lines at the origin of coordinates, as suggested by the Schwarzschild coordinates (t, r). From Eq. 5.4.23 we see that $\nabla_a r = 0$ at $X = T = 0$, and we can verify that the static Killing field ξ^a also vanishes there. $\nabla^a r$ and ξ^a become collinear along the null lines $X = \pm T$, and the vanishing of ξ^a at $X = T = 0$ leads to a mislabeling of the null lines as $t = \pm\infty$. We must point out that the Kruskal coordinates work well for the strong field region of the Schwarzschild geometry but not for the asymptotically flat region ($r \to \infty$).

For the extended Schwarzschild spacetime, region I corresponds to $r > 2M$ and can be interpreted physically as representing the exterior gravitational field of a spherical body. An observer in region I falling radially inward will cross $X = T$ and enter region II. Once he enters region II, he can never escape from it. He will fall into the singularity at $X = (T^2 - 1)^{1/2}$ within a finite proper time and any signal he sends in region II will stay in region II and fall into the singularity. Thus, region II is referred to as a **black hole**. Region III has the time reversed properties of region II and is referred to as a **white hole**. Any observer in region II must have originated from the spacetime singularity $X = -(T^2 - 1)^{1/2}$ and must leave region III within a finite time. Region IV has properties identical to region I and represents another asymptotically flat region of spacetime which lies "inside" the "radius" $r = 2M$. This is best illustrated in Figure 6.10 in Wald. However, no observer in region I can communicate with any observer in region IV and any light signal sent between the two regions will go into the black hole and be captured by the spacetime singularity.

The extended Schwarzschild solution is a valid solution of the vacuum Einstein equation and therefore represents a possible spacetime structure in GR, but there is no reason for us to believe that any region in the universe corresponds to this solution as it requires the initial condition that a singularity in region III connects two asymptotically flat regions. However, as we have discussed before in section 5.2, the metric outside a spherical body with sufficient mass will be the Schwarzschild metric at all stages of a gravitational collapse. Thus, as shown in Figure 6.11, all of regions III and IV, as well as parts of I and II, will be covered up by the collapsing matter, and part of region II (black hole) is produced when the radial coordinate of of the collapsing body becomes less than $2M$.

A Maps of Manifolds and Lie Derivatives

Let $\phi : M \to N$ be a C^∞ map, then this map carries along tangent vectors at a point $p \in M$ to tangent vectors at the point $\phi(p) \in N$. This defines a map $\phi^* : V_p \to V_{\phi(p)}$. We can also use ϕ to pull back a function $f : N \to \mathbb{R}$ on N to M to get $(f \circ \phi) : M \to \mathbb{R}$.

ϕ^* maps a vector $v \in V_p$ to a vector $\phi^* v \in V_{\phi(p)}$. Recall from section 1.2 that vectors map a collection of functions on a manifold to a number, so we define $\forall f : N \to \mathbb{R}$,

$$(\phi^* v)(f) = v(f \circ \phi). \tag{A.0.1}$$

We can think of ϕ^* as the Jacobian/derivative of ϕ at p. The map $\phi_* : V^*_{\phi(p)} \to V^*_p$ for dual vectors can be defined similarly. For all $v^a \in V_p$,

$$(\phi_* \mu)_a v^a = \mu_a (\phi^* v)^a, \tag{A.0.2}$$

where $\mu \in V^*_{\phi(p)}$. Now we can extend the definitions of ϕ^* (p to $\phi(p)$) and ϕ_* ($\phi(p)$ to p) to type $(k, 0)$ tensors and $(0, l)$ tensors, respectively.

If ϕ is a diffeomorphism, i.e. it is bijective and has C^∞ inverse, then we can extend the definition of ϕ^* to all types of tensors. Under this extended definition, ϕ^*_t with a parameter $t \in \mathbb{R}$ carries along a tensor field $T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l}$. If we choose a coordinate system on M such that the parameter t is one of coordinate x^1, then the tangent vector $v^a = (\partial/\partial x^1)^a$.

Diffeomorphisms lead to gauge freedom to any theory formulated in terms of spacetime manifolds and tensor fields because they require identical manifold structure for the manifolds they connect. We will show next that this gauge freedom is given by the Lie derivative of the tensor field.

Let ϕ_t be a one-parameter group of diffeomorphism in M, i.e. generated by v^a. Treating t as a coordinate as before, we can view $\phi^*_{-t} T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l}$ as the tensor field $T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l}$ being moved along the x^1 direction by t with all other coordinates fixed. The Lie derivative with respect to v^a is then

$$\mathcal{L}_v T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l} = \lim_{t \to \infty} \left\{ \frac{\phi^*_{-t} T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l} - T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l}}{t} \right\} \tag{A.0.3}$$

Note this is a map from a type (k, l) tensor to another type (k, l) tensor. The Lie derivative $\mathcal{L}_v$ has the following properties

(1) Leibnitz rule

(2) $\mathcal{L}_v(f) = v(f) = v^a \nabla_a f$

(3) $\mathcal{L}_v T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l} = 0$ everywhere if and only if ϕ_t is a symmetry for $T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l}$ for all t.

Under the coordinate system adapted to v^a, the components of the Lie derivative are

$$\mathcal{L}_v T^{\mu_1 \ldots \mu_k}{}_{\nu_1 \ldots \nu_l} = \frac{\partial T^{\mu_1 \ldots \mu_k}{}_{\nu_1 \ldots \nu_l}}{\partial x^1}. \tag{A.0.4}$$

From here we see that ϕ_t is a symmetry transformation of $T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l}$ if and only if the components are independent of x^1.

The intuition of the Lie derivative can be described as follows. Let the vector v^a that generates the Lie derivative $\mathcal{L}_v$ be a generator of rotation. When we have parallel transport a tensor ($\nabla_a T = 0$) around a "circle", we can view the the parallel transported tensors as being "parallel" to the original tensor. However, under the analogous condition $\mathcal{L}_v T = 0$, the "transported" tensors are "tangent" to the circle. The Lie derivative generated by the vector field v^a of a vector field w^a can written as the commutator between v^a and w^a:

$$\mathcal{L}_v w^a = [v, w]^a = v^b \nabla_b w^a - w^b \nabla_b v^a. \tag{A.0.5}$$

Using properties (1) and (2), this becomes

$$\mathcal{L}_v \mu_a = v^b \nabla_b \mu_a + \mu_b \nabla_a v^b. \tag{A.0.6}$$

And the general rule for applying the Lie derivative to a tensor is given by

$$\mathcal{L}_v T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l} = v^c \nabla_c T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l} - \sum_{i=1}^{k} T^{a_1 \ldots c \ldots a_k}{}_{b_1 \ldots b_l} \nabla_c v^{a_i} + \sum_{i=1}^{l} T^{a_1 \ldots a_k}{}_{b_1 \ldots c \ldots b_l} \nabla_{b_i} v^c \tag{A.0.7}$$

Now, consider the first order perturbation γ_{ab} of the spacetime metric g_{ab} with parameter λ around $\lambda = 0$. We have $\gamma_{ab} = dg_{ab}/d\lambda|_{\lambda=0}$ and $\gamma'_{ab} = d(\phi^*_\lambda g_{ab})/d\lambda|_{\lambda=0}$, where ϕ_λ is associated with a vector field v. From the definition of Lie derivative, we get

$$\mathcal{L}_v \ g_{ab} = \gamma_{ab} - \gamma'_{ab}. \tag{A.0.8}$$

By equation A.0.7, and assuming the derivative operator ∇_a is compatible with g_{ab},

$$\mathcal{L}_v \ g_{ab} = v^c \nabla_c g_{ab} + g_{cb} \nabla_a v^c + g_{ac} \nabla_b v^c = \nabla_a v_b + \nabla_b v_a. \tag{A.0.9}$$

Therefore the gauge transformation of γ_{ab} is given by

$$\gamma_{ab} \to \gamma'_{ab} = \gamma_{ab} - \nabla_a v_b - \nabla_b v_a. \tag{A.0.10}$$

For ϕ_λ to be a symmetry transformation of g_{ab}, the equation $\mathcal{L}_v \ g_{ab} = \nabla_a v_b + \nabla_b v_a = 0$ must be satisfied.

B Killing Vector Fields

Let $\phi_t : M \to M$ be a one-parameter group of isometries, i.e. $\phi_t^* g_{ab} = g_{ab}$, then the vector ξ^a generating ϕ_t is called a **Killing vector field**, and the following equivalence holds for ξ^a:

$$\phi_t^* g_{ab} = g_{ab} \Leftrightarrow \mathcal{L}_\xi g_{ab} = 0. \tag{B.0.1}$$

The Killing vector field solves Killing's equation

$$\nabla_a \xi_b + \nabla_b \xi_a = 0, \tag{B.0.2}$$

and has the following properties:

(1) For a geodesic γ with tangent u^a, $\xi_a u^a$ is constant along γ, i.e. $u^b \nabla_b (\xi_a u^a) = 0$.

(2) $\nabla_a \nabla_b \xi_c = -R_{bca}{}^d \xi_d$. This implies that values of ξ^a and $\nabla_a \xi_b$ at a point $p \in M$ completely determine ξ^a everywhere on M.

   (i) If $\xi_a = \nabla_a \xi_b = 0$ at a point, then $\xi_a = 0$ everywhere.

   (ii) On an n-dimensional manifold M, there can be *at most* $n + n(n-1)/2 = n(n+1)/2$ linearly independent Killing fields.

The Killing vector field describes the direction of time translation invariance.

References

[1] R. M. Wald, *General Relativity*, University of Chicago Press (1984).