BRAC UNIVERSITY

Inspiring Excellence

General Theory of Relativity (PHY413)

Lecture Notes

Preface

This series of lecture notes has been prepared for aiding students who took the BRAC University course **General Theory of Relativity (PHY413)** in Fall 2022 semester. The first part of these notes are from the course **MAT313 - Differential Geometry** that is offered by BRAC University. These particular notes are from the Summer 2022 Semester. The second part of these notes are a typeset of the notes of the course **PG512 - General Relativity and Cosmology** that is offered by Dhaka University. These particular notes are from the **MS Session 2016 - 2017**. These notes were typeset under the supervision of **Dr. Syed Hasibul Hassan Chowdhury**. The main goal of this typeset is to have an organized digital version of the notes, which is easier to share and handle. I would like to thank my friend Nian Ibne Nazrul for his contributions to the sections on gravitational waves. If you see any mistakes or typos, please send me an email at atonuroychowdhury@gmail.com.

Atonu Roy Chowdhury

References:

- GR Lecture Notes, by Harvey Reall
- *General Relativity*, by Robert M. Wald
- *Spacetime and Geometry: An Introduction to General Relativity* by Sean M. Carroll

Optional References:

- *An Introduction to Manifolds*, by Loring W. Tu
- *Differential Geometry*, by Loring W. Tu
- *Semi-Riemannian Geometry With Applications to Relativity*, by Barrett O'Neill
- *An Introduction to Differentiable Manifolds and Riemannian Geometry*, by William Boothby

Contents

1 Topology Overview

§1.1 Euclidean Space $\mathbb{R}^n$

Before embarking on the concept of general topological space, let us look at the Euclidean space $\mathbb{R}^n$. $\mathbb{R}^n$ is equipped with the notion of distance between 2 points p and q.

Definition 1.1 (Distance). Let the coordinates of p and q be $(p^1, p^2, ...p^n)$ and $(q^1, q^2, ...q^n)$, respectively. The distance between p and q is given by

$$d(p,q) = \left[\sum_{i=1}^{n}(p^i - q^i)^2\right]^{\frac{1}{2}} \tag{1.1}$$

Definition 1.2 (Open ball). An open ball $B(p,r)$ in $\mathbb{R}^n$ with center $p \in \mathbb{R}^n$ and radius $r > 0$ is defined as the set

$$B(p,r) = \{x \in \mathbb{R}^n \ : \ d(x,p) < r\} \tag{1.2}$$

A set equipped with the notion of distance between its elements is called a metric space. Thus the Euclidean space $\mathbb{R}^n$ is a metric space. And we can talk about open balls in $\mathbb{R}^n$ using this metric. We can define open sets in $\mathbb{R}^n$ using open balls $B(p,r)$ defined above.

Definition 1.3 (Open Set in $\mathbb{R}^n$). A set U in $\mathbb{R}^n$ is said to be open if for every p in U, there is an open ball $B(p,r)$ such that $B(p,r) \subseteq U$.

Proposition 1.1

The union of an arbitrary collection $\{U_\alpha\}$ of open sets is open. The intersection of finite collection of open sets is open.

Example 1.1

The intervals $\left(-\frac{1}{n}, \frac{1}{n}\right)$, $n = 1, 2, 3, ...$ are all open in $\mathbb{R}$ but their intersection

$$\bigcap_{n\in\mathbb{N}} \left(-\frac{1}{n}, \frac{1}{n}\right) = \{0\} \tag{1.3}$$

is not open.

The metric d in $\mathbb{R}^n$ allows us to define open sets in $\mathbb{R}^n$. In other words, given a subset of $\mathbb{R}^n$, we can tell if it is open or not. This situation is a special case called **metric topology in** $\mathbb{R}^n$.

§1.2 Topology

Definition 1.4 (Topology). A topology on a set S is a collection $\mathcal{T}$ of subsets of S containing both the empty set $\varnothing$ and the S such that $\mathcal{T}$ is closed under arbitrary union and finite intersection. In other words,

- If $U_\alpha \in \mathcal{T}$ for all α in an index set A, then $\bigcup_{\alpha\in A} U_\alpha \in \mathcal{T}$

- If $U_i \in \mathcal{T}$ for $i \in \{1, 2, ..., n\}$, then $\bigcap_{i=1}^{n} U_i \in \mathcal{T}$

The elements of $\mathcal{T}$ are called open sets.

Definition 1.5 (Topological Space)**.** The pair $(S, \mathcal{T})$ consisting of a set S together with a topology $\mathcal{T}$ on S is called a **topological space**.

Abuse of Notation. We shall often say "S is a topological space" in short. But there is always a topology $\mathcal{T}$ on S, which we recall when necessary.

Definition 1.6 (Neighborhood)**.** A **neighbourhood** of a point $p \in S$ is called an open set U containing p.

Definition 1.7 (Closed Set)**.** The complement of an open set is called a **closed set**.

Proposition 1.2

The union of a finite collection of closed sets is closed. The intersection of an arbitrary collection of closed sets is closed.

Proof. Let $\{F_i\}_{i=1}^{n}$ be a finite collection of closed sets. Then, $\{S \setminus F_i\}_{i=1}^{n}$ is a finite collection of open sets. The intersection of a finite collection of open sets is open, therefore $\bigcap_{i=1}^{n} (S \setminus F_i)$ is open. By De Morgan's law,

$$\bigcap_{i=1}^{n} (S \setminus F_i) = S \setminus \left(\bigcup_{i=1}^{n} F_i \right) \quad \text{is open} \implies \bigcup_{i=1}^{n} F_i \quad \text{is closed} \tag{1.4}$$

Therefore, the union of a finite collection of closed sets is closed.
Now, let $\{F_\alpha\}_{\alpha \in A}$ be an arbitrary collection of closed sets with A being an index set. Then $\{S \setminus F_\alpha\}_{\alpha \in A}$ is an arbitrary collection of open sets. We know that the union of an arbitrary collection of open sets is open, therefore $\bigcup_{\alpha \in A} (S \setminus F_\alpha)$ is open. By De Morgan's law,

$$\bigcup_{\alpha \in A} (S \setminus F_\alpha) = S \setminus \left(\bigcap_{\alpha \in A} F_\alpha \right) \quad \text{is open} \implies \bigcap_{\alpha \in A} F_\alpha \quad \text{is closed} \tag{1.5}$$

Therefore, the intersection of an arbitrary collection of closed sets is closed. ■

Definition 1.8 (Subspace Topology)**.** Let $(S, \mathcal{T})$ be a topological space and A a subset of S. Define $\mathcal{T}_A$ to be the collection of subsets

$$\mathcal{T}_A = \{U \cap A \mid U \in \mathcal{T}\} \tag{1.6}$$

$\mathcal{T}_A$ is called the **subspace topology** of A in S.

It is not hard to see that $\mathcal{T}_A$ satisfies the conditions of a Topology. Firstly, $\mathcal{T}_A$ contains both $\varnothing$ and A. For these, taking $U = \varnothing$ and $U = S$, respecitvely, suffices. By the distributive property of union and intersection

$$\bigcup_{\alpha} (U_\alpha \cap A) = \left(\bigcup_{\alpha} U_\alpha \right) \cap A \text{ and } \bigcap_{i=1}^{n} (U_i \cap A) = \left(\bigcap_{i=1}^{n} U_i \right) \cap A \tag{1.7}$$

which shows that $\mathcal{T}_A$ is closed under arbitrary union and finite intersection. So $\mathcal{T}_A$ is a Topology indeed.

Example 1.2

Consider the subset $A = [0, 1]$ of $\mathbb{R}$. In the subspace topology, the half-open interval $\left[0, \frac{1}{2}\right)$ is an open subset of A, because $\left[0, \frac{1}{2}\right) = \left(-\frac{1}{2}, \frac{1}{2}\right) \cap [0, 1]$

§1.3 Bases and Countability

Definition 1.9 (Basis and Basic Open Sets). A subcollection $\mathcal{B}$ of a topology $\mathcal{T}$ is a **basis** for $\mathcal{T}$ if given an open set U and a point p in U, there is an open set $\mathrm{B} \in \mathcal{B}$ such that $p \in \mathrm{B} \subseteq U$. An element of $\mathcal{B}$ is called a **basic open set**.

Example 1.3

The collection of all open balls $B(p, r)$ in $\mathbb{R}^n$ with $p \in \mathbb{R}^n$ and $r > 0$ is a basis for the standard topology (metric topology) on $\mathbb{R}^n$.

Proposition 1.3

A collection $\mathcal{B}$ of open sets of S is a basis if and only if every open set in S is a union of sets in $\mathcal{B}$.

Proof. ($\Rightarrow$) We are given a collection of $\mathcal{B}$ of open sets of S that is a basis. U is any open set in S. Also, let $p \in U$. Therefore, there is a basic open set $B_p \in \mathcal{B}$ such that $p \in B_p \subseteq U$. Hence, one can show that $U = \bigcup_{p \in U} B_p$.

($\Leftarrow$) Suppose, every open set in S is a union of open sets in $\mathcal{B}$. Now, given an open set U and a point $p \in U$, since $U = \bigcup_{B_\alpha \in \mathcal{B}} B_\alpha$, there is a $B_\alpha \in \mathcal{B}$, such that $p \in B_\alpha \subseteq U$. Hence $\mathcal{B}$ is a basis. ∎

We say that a point in $\mathbb{R}^n$ is rational if all of its coordinates are rational numbers. Let $\mathbb{Q}$ be the set of rational numbers and $\mathbb{Q}^+$ the set of positive rational numbers.

Lemma 1.4

Every open set in $\mathbb{R}^n$ contains a rational point.

Proof. An open set U in $\mathbb{R}^n$ contains an open ball $B(p, r)$ which, in turn, contains an open cube $\prod_{i=1}^{n} I_i$ where I_i is the open interval $\left(p^i - \frac{r}{\sqrt{n}}, p^i + \frac{r}{\sqrt{n}}\right)$. Here is a visual example for $n = 2$.

Now back to general n. For each i, let q^i be a rational number in I_i. Then $(q^1, q^2, ..., q^n)$ is a rational point in $\prod_{i=1}^{n} I_i \subseteq B(p, r)$. Therefore, every open set contains a rational point. ∎

Proposition 1.5

The collection $\mathcal{B}_{\mathbb{Q}}$ of all open balls in $\mathbb{R}^n$ with rational centers and rational radii is a basis for $\mathbb{R}^n$.

Proof. Given an open set U in $\mathbb{R}^n$ and $p \in U$, there is an open ball $B(p, r')$ with positive real radius r' such that $p \in B(p, r') \subseteq U$. Take a rational number $r \in (0, r')$. Then we have

$$p \in B(p, r) \subseteq B(p, r') \subseteq U \tag{1.8}$$

By Lemma 1.4 , there is a rational point in the smaller ball $B\left(p, \frac{r}{2}\right)$.

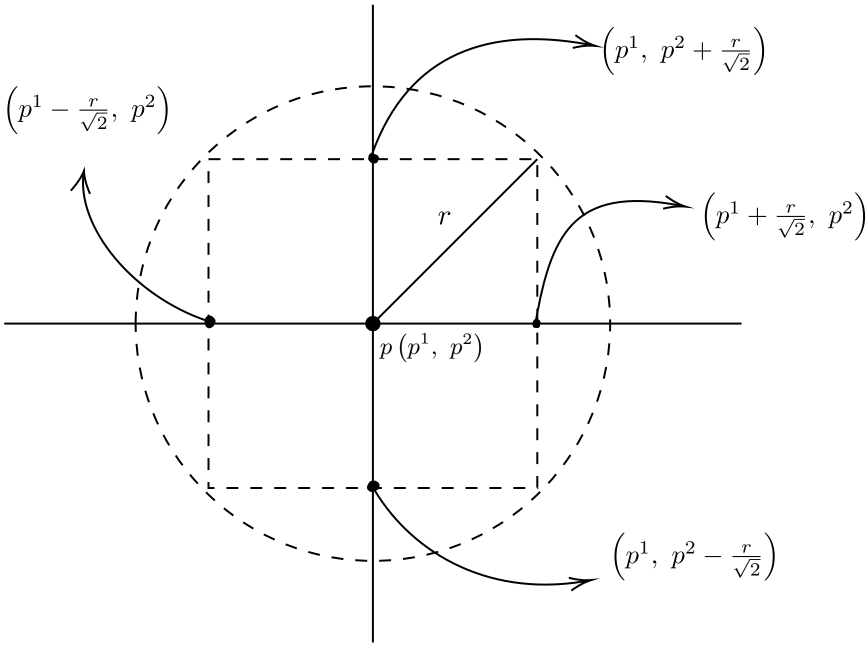


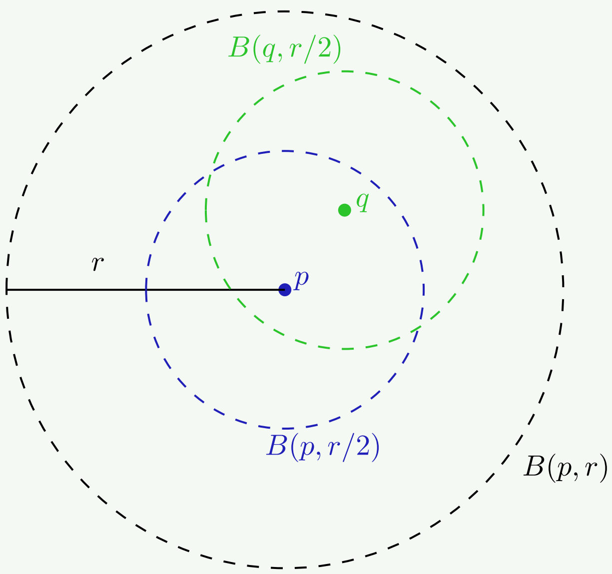
Figure 1.1: $B(p,r)$ contains $\left(p^1 - \frac{r}{\sqrt{n}}, p^1 + \frac{r}{\sqrt{n}}\right) \times \left(p^2 - \frac{r}{\sqrt{n}}, p^2 + \frac{r}{\sqrt{n}}\right)$

Claim — $p \in B\left(q, \frac{r}{2}\right) \subseteq B(p,r)$

Proof. Since $d(p,q) < \frac{r}{2}$, we have $p \in B\left(q, \frac{r}{2}\right)$. Next, if $x \in B\left(q, \frac{r}{2}\right)$, then by triangle inequality

$$d(x,p) \leq d(x,q) + d(q,p) < \frac{r}{2} + \frac{r}{2} = r \tag{1.9}$$

Therefore, $x \in B(p,r)$.



So, $p \in B\left(q, \frac{r}{2}\right)$ and $B\left(q, \frac{r}{2}\right) \subseteq B(p,r)$. □

As a result, $p \in B\left(q, \frac{r}{2}\right) \subseteq B(p,r) \subseteq B\left(p, r'\right) \subseteq U$. Hence we proved,

$$p \in B\left(q, \frac{r}{2}\right) \subseteq U \tag{1.10}$$

In other words, the collection $\mathcal{B}_{\mathbb{Q}}$ of open balls with rational centers and rational radii is a basis for $\mathbb{R}^n$. ■

Both the sets $\mathbb{Q}$ and $\mathbb{Q}^+$ are countable. Since the centers of the open balls in $\mathcal{B}_{\mathbb{Q}}$ are indexed by $\mathbb{Q}^n$, a countable set, and the radii are indexed by $\mathbb{Q}^+$, also a countable set, the collection $\mathcal{B}_{\mathbb{Q}}$ is countable.

Definition 1.10 (Second Countable). A topological space is said to be second countable if it has a countable basis.

Proposition 1.5 shows that $\mathbb{R}^n$ with its standard topology is second countable.

§1.4 Hausdorff Space

Definition 1.11 (Hausdorff Space). A topological space S is Hausdorff if given any 2 distinct points x, y in S there exist disjoint open sets U, V such that $x \in U$ and $y \in V$.

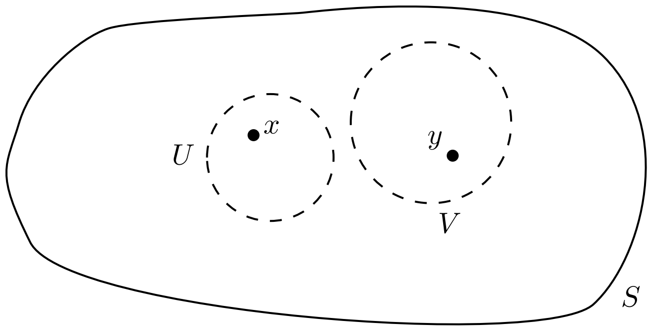


Figure 1.2: Here S is a Hausdorff space, U and V are disjoint open sets containing x and y respectively.

Proposition 1.6

Every singleton set (a one-point set) in a Hausdorff space S is closed.

Proof. Let $x \in S$. We want to prove that $\{x\}$ is closed, i.e. $S \setminus \{x\}$ is open.

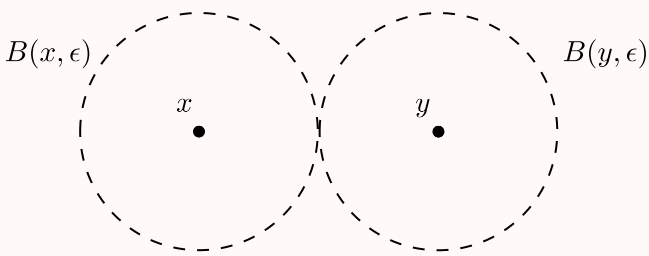
Let $y \in S \setminus \{x\}$. Since S is Hausdorff, we can find disjoint open sets U_y and V_y such that $x \in U_y$ and $y \in V_y$. No such V_y contains x. Therefore

$$S \setminus \{x\} = \bigcup_{y \in S \setminus \{x\}} V_y \tag{1.11}$$

So $S \setminus \{x\}$ is union of open sets, hence open. So $\{x\}$ is closed. ■

Example 1.4

The Euclidean space $\mathbb{R}^n$ (equipped with standard/ metric topology) is Hausdorff, for given distinct points x, y in $\mathbb{R}^n$, if $\epsilon = \frac{1}{2} d(x, y)$, then the open balls $B(x, \epsilon)$ and $B(y, \epsilon)$ will be disjoint.



§1.5 Continuity and Homeomorphism

Definition 1.12 (Continuous Maps)**.** Let $f : X \to Y$ be a map of topological spaces. f is said to be continuous if for each open subset V of Y, the set $f^{-1}(V)$ is an open subset of X.

Proposition 1.7

$f : X \to Y$ is continuous if and only if for every closed subset B of Y, the set $f^{-1}(B)$ will be closed in X.

Proof. ($\Rightarrow$) Suppose f is continuous. B is closed, so $Y \setminus B$ is open in Y. Therefore, by the continuity of f, $f^{-1}(Y \setminus B) = X \setminus f^{-1}(B)$ is open in X, so $f^{-1}(B)$ is closed.

($\Leftarrow$) Suppose $f^{-1}(B)$ is closed in X for any closed $B \subseteq Y$. Take any open set U in Y. Choose $B = Y \setminus U$. Then by the assumption $f^{-1}(Y \setminus U) = X \setminus f^{-1}(U)$ is closed in X. This gives us $f^{-1}(U)$ is open. So f is continuous. ■

Definition 1.13 (Homeomorphism)**.** Let X and Y be topological spaces; let $f : X \to Y$ be a bijection. If both f and the inverse function $f^{-1} : Y \to X$ are continuous, then f is called a **homeomorphism**.

Example 1.5

The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 3x + 1$ is a homeomorphism. We define $g : \mathbb{R} \to \mathbb{R}$ by $g(y) = \frac{1}{3}(y - 1)$. Then we have

$$f(g(y)) = y \quad \text{and} \quad g(f(x)) = x \qquad \forall\, x, y \in \mathbb{R} \tag{1.12}$$

This proves $g = f^{-1}$. It is easy to see that both f and g are continuous functions. Therefore f is a homeomorphism.

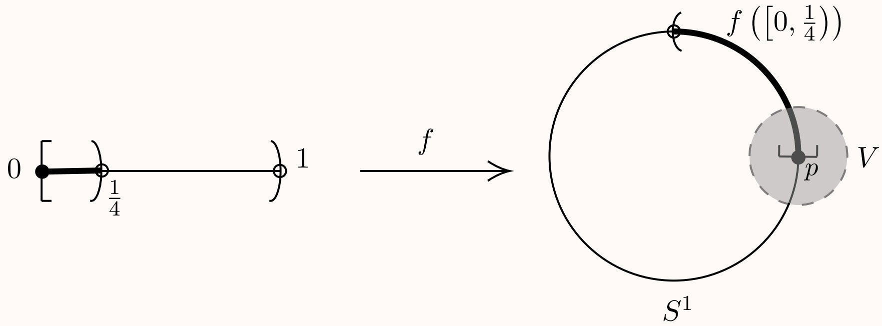
However, a bijective function can be continuous without being a homeomorphism.

Example 1.6

Let S^1 denote the unit circle in $\mathbb{R}^2$; that is $S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$, considered as a **subspace**[a] of the space $\mathbb{R}^2$. Let $f : [0, 1) \to S^1$ be the

$$f(t) = (\cos 2\pi t, \sin 2\pi t) \tag{1.13}$$

It is left as an exercise for the reader to show that f is a continuous bijective function. But the function f^{-1} is not continuous.



$U = \left[0, \frac{1}{4}\right)$ is an open set in $[0, 1)$ according to the subspace topology. We want to show that $f(U)$ is not open in S^1. That would prove the discontinuity of f^{-1}.

Let p be the point $f(0)$. And $p \in f(U)$. We need to find an open set of S^1 in subspace topology containing $p = f(0)$ and contained in $f(U)$ to show that $f(U)$ is open in S^1, i.e we have to find an open set in V of $\mathbb{R}^2$ such that $f(0) = p \in V \cap S^1 \subseteq f(U)$. But it is impossible as is evident from the figure above. No matter what V we choose, some part of $V \cap S^1$ would lie outside $f(U)$.

[a] Subset of $\mathbb{R}^2$ equipped with subspace topology.

2 Multivariable Calculus Review

§2.1 Differentiabiliy

Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined as follows:

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases} \tag{2.1}$$

For the piecewise defined function stated above, note that along the x-axis $y = 0$. So $f(x, 0) = 0$ for every $x \in \mathbb{R}$. In other words, f is constant and identically 0 on the x-axis. Therefore,

$$\left.\frac{\partial f}{\partial x}(x, y)\right|_{y=0} = 0\,. \tag{2.2}$$

Similarly, along the y-axis $x = 0$. So $f(0, y) = 0$ for every $y \in \mathbb{R}$. In other words, f is constant and identically 0 on the y-axis. Therefore,

$$\left.\frac{\partial f}{\partial y}(x, y)\right|_{x=0} = 0\,. \tag{2.3}$$

Therefore, both the partial derivatives exist at $(0, 0)$, and are equal to 0. We will now show that f is not even continuous at $(0, 0)$. Consider the line $y = x$, and we shall evaluate the limit of $f(x, y)$ as $(x, y) \to (0, 0)$ along this line.

$$\lim_{x\to 0} f(x, x) = \lim_{x\to 0} \frac{x \cdot x}{x^2 + x^2} = \frac{1}{2} \neq 0\,. \tag{2.4}$$

So we get,

$$\begin{aligned} &\lim_{(x,y)\to(0,0)} f(x, y) = 0\ ,\ \text{along } x\text{-axis;} \\ &\lim_{(x,y)\to(0,0)} f(x, y) = 0\ ,\ \text{along } y\text{-axis;} \\ &\lim_{(x,y)\to(0,0)} f(x, y) = \frac{1}{2}\ ,\ \text{along the line } y = x. \end{aligned} \tag{2.5}$$

Therefore, f is not even continuous at $(0, 0)$, let alone being differentiable. Therefore, mere existence of partial derivatives of order doesn't guarantee differentiability at a given point.

We will, first, consider functions whose domain is $U \subseteq \mathbb{R}^n$ and codomain is $\mathbb{R}$. If $f : U \to \mathbb{R}^n$ is such a function, then $f(\vec{x}) = f\left(x^1, x^2, \ldots, x^n\right)$ denotes its value at $\vec{x} \equiv \left(x^1, x^2, \ldots, x^n\right) \in U$. We also assume that the underlying domain of f is an open set $U \subseteq \mathbb{R}^n$. At each $\vec{a} \in U$, the partial derivative $\left.\frac{\partial f}{\partial x^j}\right|_{\vec{a}}$ of f with respect to x^j is the following limit, if it exists

$$\left.\frac{\partial f}{\partial x^j}\right|_{\vec{x}=\vec{a}} = \lim_{h\to 0} \frac{f\left(a^1, \ldots, a^j + h, \ldots, a^n\right) - f\left(a^1, \ldots, a^j, \ldots, a^n\right)}{h}\,. \tag{2.6}$$

If $\frac{\partial f}{\partial x^j}$ is defined, that is, the limit above exists at each point of U for $1 \leq j \leq n$, this defines n functions on U. Should these functions be continuous on U for $1 \leq j \leq n$, f is said to be continuously differentiable on U, denoted by $f \in C^1(U)$.

We shall say that f is differentiable at $\vec{a} \in U$ if there is a homogenous linear expression $\sum_{i=1}^{n} b_i \left(x^i - a^i\right)$ such that the inhomogenous expression $f(\vec{a}) + \sum_{i=1}^{n} b_i \left(x^i - a^i\right)$ approximates $f(\vec{x})$ near $\vec{a}$ in the following sense:

$$\lim_{\vec{x} \to \vec{a}} \frac{f(\vec{x}) - f(\vec{a}) - \sum_{i=1}^{n} b_i \left(x^i - a^i\right)}{\|\vec{x} - \vec{a}\|} = 0\,. \tag{2.7}$$

In other words, if there exist constants $b_1, b_2, \ldots, b_n$ and a real valued function $r(\vec{x}, \vec{a})$ defined on a neghborhood V of $\vec{a} \in U$ such that the following two conditions hold:

$$f(\vec{x}) = f(\vec{a}) + \sum_{i=1}^{n} b_i \left(x^i - a^i\right) + \|\vec{x} - \vec{a}\|\, r(\vec{x}, \vec{a}) \quad \text{and} \quad \lim_{\vec{x} \to \vec{a}} r(\vec{x}, \vec{a}) = 0\,. \tag{2.8}$$

b_i's are uniquely determined, and they are the partial derivatives at $\vec{a}$:

$$b_i = \left.\frac{\partial f}{\partial x^i}\right|_{\vec{x} = \vec{a}}\,. \tag{2.9}$$

In fact,

$$f(\vec{x}) = f(\vec{a}) + \sum_{i=1}^{n} \left.\frac{\partial f}{\partial x^i}\right|_{\vec{x} = \vec{a}} \left(x^i - a^i\right) + \|\vec{x} - \vec{a}\|\, r(\vec{x}, \vec{a})\,. \tag{2.10}$$

Actually, existence of partial derivatives and their continuity guarantees differentiability at a given point $\vec{a} \in U \subseteq \mathbb{R}^n$.

§2.2 Chain Rule

By a differentiable curve in $\mathbb{R}^n$, we mean $f : (a, b) \to \mathbb{R}^n$, with $f(t) = \left(x^1(t), x^2(t), \ldots, x^n(t)\right)$, where the n coordinate functions $x^i(t)$ are all differentiable on (a, b). Recall that, for a function of one variable, differentiability is equivalent to existence of derivative.

Here, $\left(x^i(t)\right)$ are real valued functions of one variable. And you must be familiar with the notion of C^r-differentiability of real valued functions of one variable. For example, $h(t) = t^{\frac{1}{3}}$ is not C^1, because its derivative does not exist at $t = 0$. Similarly, $k(t) = t^{\frac{4}{3}}$ is C^1, but not C^2.

Now, let's suppose $f : (a, b) \to \mathbb{R}^n$ is a C^r differentiable curve in the sense that all the n coordinate functions $x^i(t)$ are C^r differentiable. Take t_0 with $a < t_0 < b$, and $f : (a, b) \to U \subseteq \mathbb{R}^n$. Let g be a C^r-differentiable function from U to $\mathbb{R}$. In particular, $g : U \to \mathbb{R}$ is differentiable at $f(t_0) \in U$. Then $g \circ f : (a, b) \to \mathbb{R}$ is differentiable at t_0, and the derivative is given by:

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}(g \circ f)(t)\right|_{t = t_0} = \sum_{i=1}^{n} \left.\frac{\partial g(f(t))}{\partial x^i}\right|_{f(t_0)} \cdot \left.\frac{\mathrm{d}x^i(t)}{\mathrm{d}t}\right|_{t = t_0}\,. \tag{2.11}$$

This result is known as the chain rule for real-valued functions.

Now, we can generalize this idea to functions on subsets U of $\mathbb{R}^n$, whose range is not in $\mathbb{R}$, but in $\mathbb{R}^n$. In other words, we consider $F : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$.

$$\vec{x} \equiv \left(x^1, x^2, \ldots, x^n\right) \in U\ ;\ F(\vec{x}) = \left(F^1(\vec{x}), F^2(\vec{x}), \ldots, F^m(\vec{x})\right)\,. \tag{2.12}$$

Now take a point $\vec{p} \in U$ with coordinate $\left(p^1, p^2, \ldots, p^n\right)$. Then $F(\vec{p})$ is a point in V with coordinate $\left(F^1(\vec{p}), F^2(\vec{p}), \ldots, F^m(\vec{p})\right)$. Now let $G : V \subseteq R^m \to R^l$. Write a point $\vec{y} \equiv \left(y^1, y^2, \ldots, y^m\right) \in V \subseteq \mathbb{R}^m$. Then

$$G(\vec{y}) = \left(G^1(\vec{y}), G^2(\vec{y}), \ldots, G^l(\vec{y})\right)\,. \tag{2.13}$$

In other words, $G^i : V \to \mathbb{R}$. Then we have $G^i \circ F : U \subseteq \mathbb{R}^n \to \mathbb{R}$. In this case, the chain rule is

$$\frac{\partial \left(G^i \circ F\right)}{\partial x^j}(\vec{p}) = \sum_{k=1}^{m} \frac{\partial G^i}{\partial y^k}(F(\vec{p})) \cdot \frac{\partial F^k}{\partial x^j}(\vec{p})\,. \tag{2.14}$$

§2.3 Differential of a Map

Let $F : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$. Let $T_p\mathbb{R}^n$ denote the tangent space on $\mathbb{R}^n$ to the point $p \in \mathbb{R}^n$. (For convenience, we'll drop arrows in $\vec{p}$) The differential of F at p is a map $DF_p : T_p\mathbb{R}^n \to T_{F(p)}\mathbb{R}^m$. $T_p\mathbb{R}^n$ is clearly isomorphic to $\mathbb{R}^n$ as vector space. Hence, $DF_p : \mathbb{R}^n \to \mathbb{R}^m$. Let's try to see that DF_p is related to the Jacobian matrix of $F : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$.

$$\left\{ \left.\frac{\partial}{\partial x^1}\right|_p, \left.\frac{\partial}{\partial x^2}\right|_p, \ldots, \left.\frac{\partial}{\partial x^n}\right|_p \right\} \tag{2.15}$$

is a basis of $T_p\mathbb{R}^n$, which can be treated as $\mathbb{R}^n$ with origin at p. Similarly,

$$\left\{ \left.\frac{\partial}{\partial y^1}\right|_{F(p)}, \left.\frac{\partial}{\partial y^2}\right|_{F(p)}, \ldots, \left.\frac{\partial}{\partial y^m}\right|_{F(p)} \right\} \tag{2.16}$$

is a basis of $T_{F(p)}\mathbb{R}^m$, which can be treated as $\mathbb{R}^m$ with origin at $F(p)$.

Geometric tangent vectors like $\left.\frac{\partial}{\partial x^i}\right|_p$ or $\left.\frac{\partial}{\partial y^j}\right|_{F(p)}$ act on smooth functions of $\mathbb{R}^n$ or $\mathbb{R}^m$, respectively, and spit out real numbers.

$$\left.\frac{\partial}{\partial x^i}\right|_p f = \frac{\partial f}{\partial x^i}(p) \in \mathbb{R}\,. \tag{2.17}$$

Since DF_p is a linear map between two vector spaces, in order to express DF_p as a matrix, we need to find where the basis vectors are getting mapped. So we want to find $DF_p\left(\left.\frac{\partial}{\partial x^i}\right|_p\right)$. This is a vector in $T_{F(p)}\mathbb{R}^m$, and hence can be written as a linear combination of $\left.\frac{\partial}{\partial y^j}\right|_{F(p)}$'s. Now we wish to find the coefficients in the linear combination.

$DF_p\left(\left.\frac{\partial}{\partial x^i}\right|_p\right)$ acts on $f \in C^\infty(\mathbb{R}^m)$ and yields a real number.

$$DF_p\left(\left.\frac{\partial}{\partial x^i}\right|_p\right) f := \left.\frac{\partial}{\partial x^i}\right|_p (f \circ F)\,. \tag{2.18}$$

This makes perfect sense as $f \circ F : U \subseteq \mathbb{R}^n \to \mathbb{R}$. By chain rule,

$$\left.\frac{\partial}{\partial x^i}\right|_p (f \circ F) = \frac{\partial (f \circ F)}{\partial x^i}(p) = \sum_{j=1}^{m} \left.\frac{\partial f}{\partial y^j}\right|_{F(p)} \left.\frac{\partial F^j}{\partial x^i}\right|_p\,. \tag{2.19}$$

$$\therefore DF_p\left(\left.\frac{\partial}{\partial x^i}\right|_p\right) f = \sum_{j=1}^{m} \left.\frac{\partial F^j}{\partial x^i}\right|_p \left.\frac{\partial}{\partial y^j}\right|_{F(p)} f \implies \boxed{DF_p\left(\left.\frac{\partial}{\partial x^i}\right|_p\right) = \sum_{j=1}^{m} \frac{\partial F^j}{\partial x^i}(p) \cdot \left.\frac{\partial}{\partial y^j}\right|_{F(p)}} \tag{2.20}$$

Therefore, DF_p can be represented by the following $m \times n$ matrix:

$$\begin{bmatrix} \frac{\partial F^1}{\partial x^1}(p) & \frac{\partial F^1}{\partial x^2}(p) & \cdots & \frac{\partial F^1}{\partial x^n}(p) \\ \frac{\partial F^2}{\partial x^1}(p) & \frac{\partial F^2}{\partial x^2}(p) & \cdots & \frac{\partial F^2}{\partial x^n}(p) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F^m}{\partial x^1}(p) & \frac{\partial F^m}{\partial x^2}(p) & \cdots & \frac{\partial F^m}{\partial x^n}(p) \end{bmatrix} \tag{2.21}$$

F is differentiable at $p \in U \subseteq \mathbb{R}^n$ if all the entries in the $m \times n$ matrix DF exist and are continuous at p. If F is differentiable at every $p \in U$, we say that F is of class C^1. DF is called the total derivative in the language of multivariable calculus.

Similarly, if all the second order partial derivatives exist and are continuous at p, then we say F is twice differentiable at p. If F is twice differentiable at every $p \in U$, we say F is of class C^2. In a similar manner, we define maps of class C^r. If a map F is of class C^r for every $r \in \mathbb{N}$, we say F is **smooth** or **infinitely differentiable**, or F belongs in the class C^∞.

§2.4 Inverse Function Theorem

Definition 2.1. Let U and V be open subsets of $\mathbb{R}^n$. A map $F : U \to V$ is said to be a $\boldsymbol{C^r}$**-diffeomorphism** if F is a homeomorphism, and both F and F^{-1} are of class C^r. When $r = \infty$, we just say F is a **diffeomorphism**.

Theorem 2.1 (Inverse Function Theorem)

Let W be an open subset of $\mathbb{R}^n$ and $F : W \to \mathbb{R}^n$ a C^∞ mapping. If $p \in W$ and DF_p is nonsingular, then there exists a neighborhood U of p in W such that $V = F(U)$ is open and $F : U \to V$ is a diffomorphism. If $x \in U$, then

$$DF^{-1}_{F(x)} = (DF_x)^{-1}\ . \tag{2.22}$$

We are not going to prove it here. We will see an example now.

Example 2.1. Let's consider the conversion of polar to rectangular coordinate. $F : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by

$$F\begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}\ . \tag{2.23}$$

Then the differential DF is

$$DF = \begin{bmatrix} \dfrac{\partial F^1}{\partial r} & \dfrac{\partial F^1}{\partial \theta} \\ \dfrac{\partial F^2}{\partial r} & \dfrac{\partial F^2}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} \tag{2.24}$$

Hence, $\det DF = r$. So $DF_{(r,\theta)}$ is differentiable for $r \neq 0$. Choose $r = \sqrt{2}$ and $\theta = \frac{\pi}{4}$. Then

$$F\begin{pmatrix} \sqrt{2} \\ \frac{\pi}{4} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}\ . \tag{2.25}$$

$$DF_{(\sqrt{2},\frac{\pi}{4})} = \begin{bmatrix} \dfrac{1}{\sqrt{2}} & -1 \\ \dfrac{1}{\sqrt{2}} & 1 \end{bmatrix}\ . \tag{2.26}$$

By the Inverse Function Theorem, there is a local inverse

$$DF^{-1}_{(1,1)} = \left(DF_{(\sqrt{2},\frac{\pi}{4})}\right)^{-1}\ . \tag{2.27}$$

Now, F^{-1} is given by

$$F^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \sqrt{x^2+y^2} \\ \tan^{-1}\left(\frac{y}{x}\right) \end{pmatrix}\ . \tag{2.28}$$

Therefore,

$$DF^{-1} = \begin{bmatrix} \dfrac{2x}{2\sqrt{x^2+y^2}} & \dfrac{2y}{2\sqrt{x^2+y^2}} \\ \dfrac{-y}{x^2+y^2} & \dfrac{x}{x^2+y^2} \end{bmatrix} = \begin{bmatrix} \dfrac{x}{\sqrt{x^2+y^2}} & \dfrac{y}{\sqrt{x^2+y^2}} \\ \dfrac{-y}{x^2+y^2} & \dfrac{x}{x^2+y^2} \end{bmatrix}\ . \tag{2.29}$$

As a result,

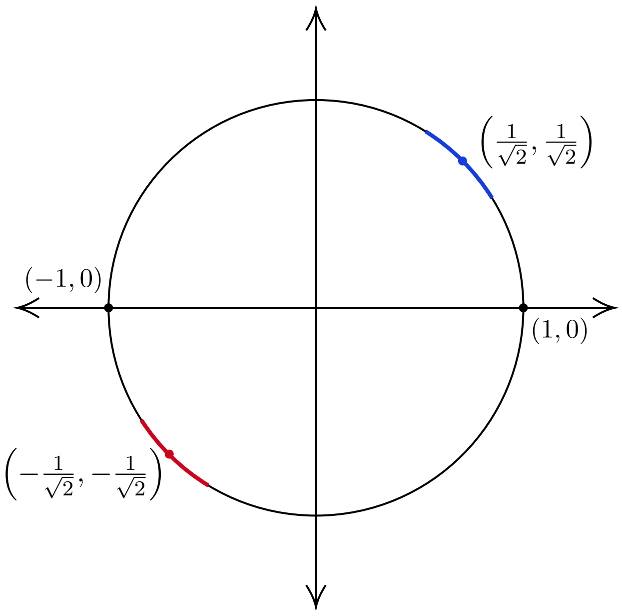
$$DF^{-1}_{(1,1)} = \begin{bmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\ \dfrac{-1}{2} & \dfrac{1}{2} \end{bmatrix}\ . \tag{2.30}$$

One can indeed check that

$$\begin{bmatrix} \dfrac{1}{\sqrt{2}} & -1 \\ \dfrac{1}{\sqrt{2}} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\ \dfrac{-1}{2} & \dfrac{1}{2} \end{bmatrix}\ . \tag{2.31}$$

§2.5 Implicit Function Theorem

Let us consider the equation of a unit circle in $\mathbb{R}^2$; $x^2 + y^2 = 1$.



The graph of the unit circle above does not represent a function. Because, for a given value of x, there are 2 values for y that satisfy the equation. Choose a point, say $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, on the unit circle. Then one can consider an arc (colored blue in the figure above) containing $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ that indeed represents a function given by $y = \sqrt{1 - x^2}$. Had we started with the point $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$, we could find an arc (colored red in the figure above) containing $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ that represents a function given by $y = -\sqrt{1 - x^2}$. The only problematic points are $(1, 0)$ and $(-1, 0)$. No matter how small an arc we choose about these points, it is not going to be represented by a function. Because, for those arcs, for a given x, there will be multiple values for y.

Now let us address the following 2-dimensional problem: Given an equation $F(x, y) = 0$, which is not globally a functional relationship (in the unit circle example, $F(x, y) = x^2 + y^2 - 1$), does there exist a point (x_0, y_0) satisfying $F(x_0, y_0) = 0$ so that there exists a neighborhood of (x_0, y_0) where y can be written as $y = f(x)$ for some real valued function f of one variable? In other words, $F(x, f(x)) = 0$ should hold for all values of x in that neighborhood. In the unit circle example, this f was given by $f(x) = \sqrt{1 - x^2}$ or $f(x) = -\sqrt{1 - x^2}$, depending on the choice of the point (x_0, y_0) in the upper or lower semicircle, respectively. The Implicit Function Theorem guarantees the **local existence** of such a function provided the initial point (x_0, y_0) was chosen *appropriately*. In the unit circle example, $(1, 0)$ and $(-1, 0)$ were two *inappropriate* points. As required by the Implicit Function Theorem, one must have

$$\frac{\partial F}{\partial y}(x_0, y_0) \neq 0 \,. \tag{2.32}$$

But in this case, for $F(x, y) = x^2 + y^2 - 1$,

$$\frac{\partial F}{\partial y} = 2y \implies \frac{\partial F}{\partial y}(1, 0) = 0 = \frac{\partial F}{\partial y}(1, 0) \,. \tag{2.33}$$

Therefore, in the light of Implicit Function Theorem, $(1, 0)$ and $(-1, 0)$ are not *appropriate* points on the unit circle around which we can construct a locally functional relationship. Now we state the most general form of Implicit Function Theorem.

Theorem 2.2 (Implicit Function Theorem)

Let U be an open set in $\mathbb{R}^n \times \mathbb{R}^m$ and $F : U \to \mathbb{R}^m$ a C^∞ map. Write $(x, y) = \left(x^1, \ldots, x^n, y^1, \ldots, y^m\right)$ for a point in U. Suppose the matrix

$$\left[\frac{\partial F^i}{\partial y^j}(x_0, y_0)\right]_{1 \leq i, j \leq m} \tag{2.34}$$

is non-singular for a point $(x_0, y_0) \in U$ satisfying $F(x_0, y_0) = 0$. Then there exists a neighborhood $X \times Y$ of (x_0, y_0) in U and a unique C^∞ map $f : X \to Y$ such that in $X \times Y \subseteq U \subseteq \mathbb{R}^n \times \mathbb{R}^m$,

$$F(x, y) = 0 \iff y = f(x) \ . \tag{2.35}$$

3 Differentiable Manifolds

§3.1 Manifolds

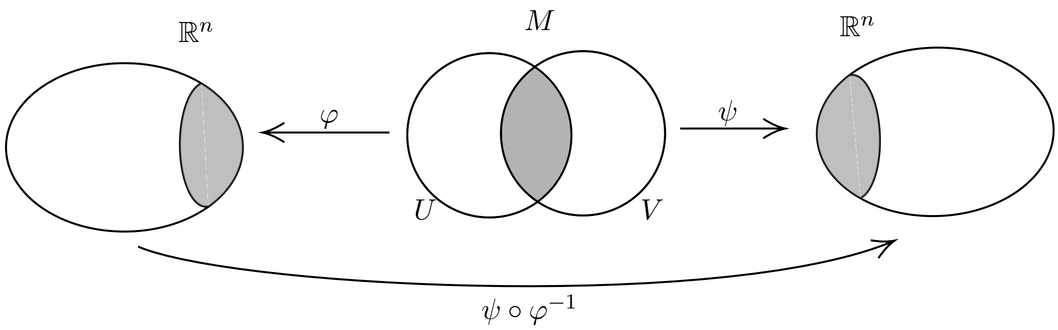
Definition 3.1 (Locally Euclidean Space). A topological space M is **locally Euclidean** of dimension n if every point in M has a neighborhood U such that there is a homeomorphism φ from U onto an open subset of $\mathbb{R}^n$. We call the pair $(U, \varphi : U \to \mathbb{R}^n)$ a **chart**, U a **coordinate neighborhood** and φ a **coordinate system** on U. We also say that a chart (U, φ) is centered at $p \in U$ if $\varphi(p) = \vec{0}$.

Definition 3.2 (Topological Manifold). A **topological manifold** of dimension n is a Hausdorff, second countable, locally Euclidean space of dimension n.

Definition 3.3 (Compatible Charts). Two charts $(U, \varphi : U \to \mathbb{R}^n)$ and $(V, \psi : V \to \mathbb{R}^n)$ of a topological manifold are **C^∞-compatible** if the two maps

$$\psi \circ \varphi^{-1} : \varphi(U \cap V) \to \psi(U \cap V) \quad \text{and} \quad \varphi \circ \psi^{-1} : \psi(U \cap V) \to \varphi(U \cap V) \tag{3.1}$$

are both C^∞. These two maps are called **transition functions** between the charts. If $U \cap V$ is empty, then the two charts are automatically compatible.



Definition 3.4 (Atlas). A **C^∞-atlas** or simply an **atlas** on a locally Euclidean space M is a collection $\mathscr{U} = \{(U_\alpha, \varphi_\alpha)\}$ of pairwise C^∞-compatible charts that cover M. In other words,

$$M = \bigcup_\alpha U_\alpha . \tag{3.2}$$

Definition 3.5 (Maximal Atlas). An atlas $\mathscr{M}$ on a locally Euclidean space is said to be **maximal** if it is not contained in a larger atlas. In other words, if $\mathscr{U}$ is any other atlas containing $\mathscr{M}$, then $\mathscr{U} = \mathscr{M}$.

Definition 3.6 (Smooth Manifold). A **smooth** or **C^∞ manifold** is a topological manifold M together with a maximal atlas $\mathscr{M}$. The maximal atlas is also called a *differentiable structure* on M.

In practice, to check that a topological manifold M is a smooth manifold, it is not necessary to exhibit a maximal atlas. The existence of *any* atlas on M will do, because of the following proposition.

Proposition 3.1

Any atlas $\mathscr{U} = \{(U_\alpha, \varphi_\alpha)\}$ on a locally Euclidean space is contained in a unique maximal atlas.

In summary, to show that a topological space M is a smooth manifold, it suffices to check that

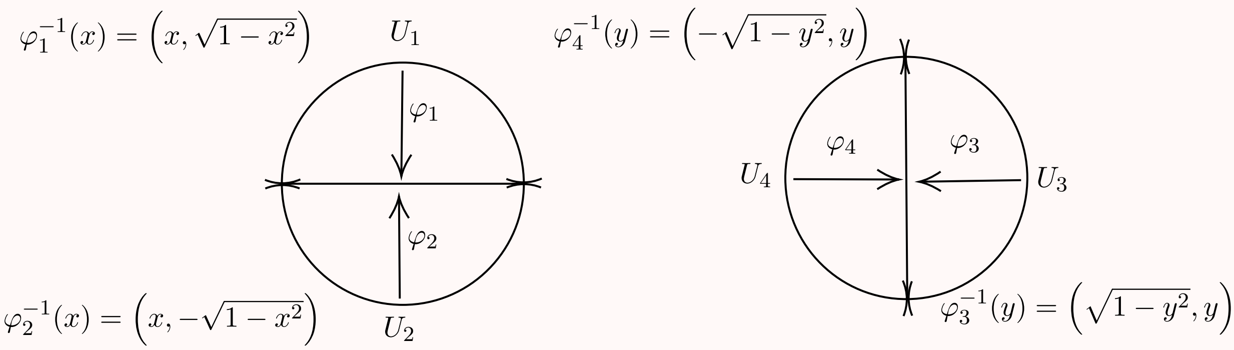
(i) M is Hausdorff and second countable,

(ii) M has a C^∞ atlas (not necessarily maximal).

Example 3.1 (Unit circle in the (x, y)-plane)

We'll view S^1 as the unit circle in $\mathbb{R}^2$ with defining equation $x^2 + y^2 = 1$. We can cover S^1 with 4 open sets: the upper and lower semicirles U_1 and U_2, the right and left semicircles U_3 and U_4. The homeomorphisms are:

$$\varphi_i : U_i \to (-1, 1)\ ,\ \varphi_i(x, y) = \begin{cases} x & \text{if } i = 1, 2 \\ y & \text{if } i = 3, 4 \end{cases} \tag{3.3}$$



Let us check that on $U_1 \cap U_3$,

$$\left(\varphi_3 \circ \varphi_1^{-1}\right)(\varphi_1(x, y)) = \left(\varphi_3 \circ \varphi_1^{-1}\right)(x) = \varphi_3\left(x, \sqrt{1-x^2}\right) = \sqrt{1-x^2}\,. \tag{3.4}$$

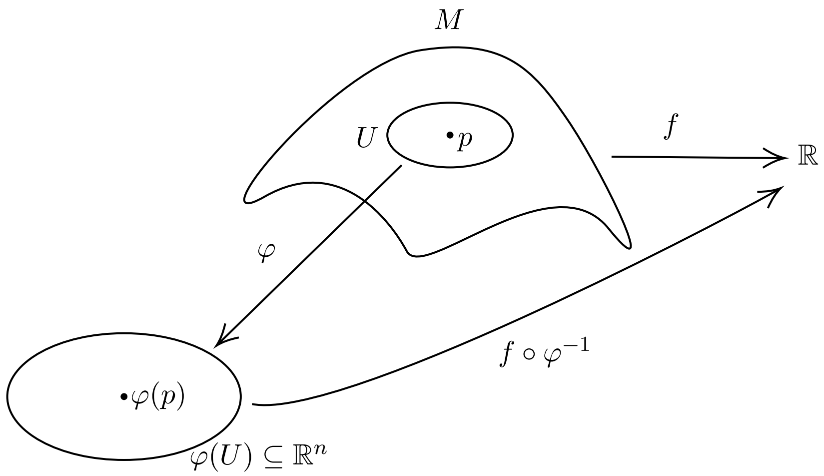
Since $(1, 0) \notin U_1 \cap U_3$, we can conclude that $\varphi_3 \circ \varphi_1^{-1}$ is C^∞. Also, on $U_2 \cap U_4$,

$$\left(\varphi_2 \circ \varphi_4^{-1}\right)(\varphi_4(x, y)) = \left(\varphi_2 \circ \varphi_4^{-1}\right)(y) = \varphi_2\left(-\sqrt{1-y^2}, y\right) = -\sqrt{1-y^2}\,. \tag{3.5}$$

Since $(0, -1) \notin U_2 \cap U_4$, we can conclude that $\varphi_2 \circ \varphi_4^{-1}$ is C^∞. In a similar manner, one can check that $\varphi_i \circ \varphi_j^{-1}$ is C^∞ for every i, j. Therefore, $\{(U_i, \varphi_i) \mid 1 \le i \le 4\}$ is indeed a C^∞ atlas on S^1.

§3.2 Smooth Maps on Manifold

Definition 3.7. Let M be a smooth manifold of dimension n. A function $f : M \to \mathbb{R}$ is said to be C^∞ or smooth at a point $p \in M$ if there is a chart (U, φ) about p in M such that $f \circ \varphi^{-1} : \varphi(U) \subseteq \mathbb{R}^n \to \mathbb{R}$ is C^∞ at $\varphi(p)$. The function f is said to be C^∞ on M if it is C^∞ at every point of M.



The tangent space T_pM at $p \in M$ is spanned by $\left\{ \frac{\partial}{\partial x^1}\Big|_p, \frac{\partial}{\partial x^2}\Big|_p, \dots, \frac{\partial}{\partial x^n}\Big|_p \right\}$. Any tangent vector X_p can be written as a linear combination of thee basis vectors,

$$X_p = \sum_{i=1}^{n} X^i(p) \left.\frac{\partial}{\partial x^i}\right|_p . \tag{3.6}$$

These vectors are maps $C^\infty(M, \mathbb{R}) \to \mathbb{R}$ given by

$$X_p f = \sum_{i=1}^{n} X^i(p) \left.\frac{\partial}{\partial x^i}\right|_p f = \sum_{i=1}^{n} X^i(p) \frac{\partial f}{\partial x^i}(p) . \tag{3.7}$$

So, it needs a well defined notion of $\frac{\partial}{\partial x^i}\Big|_p f$, for $f \in C^\infty(M, \mathbb{R})$, i.e. f is a smooth function defined in the neighborhood of p. This is defined as,

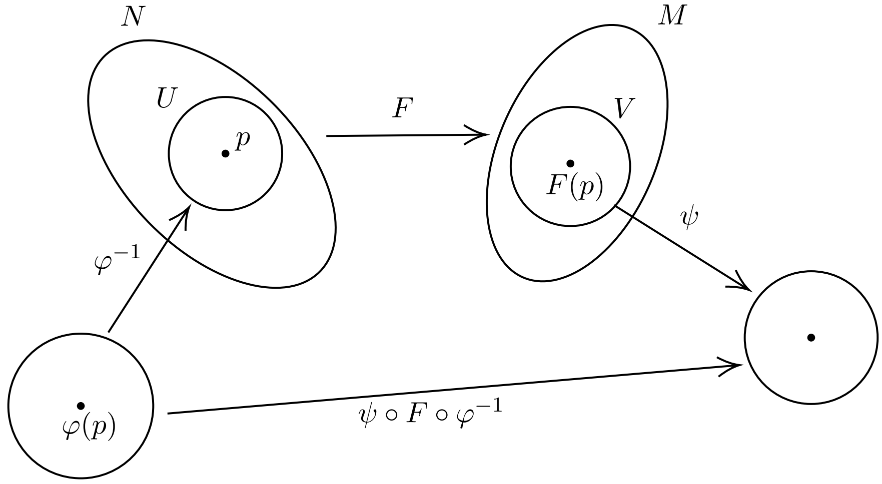
$$\left.\frac{\partial}{\partial x^i}\right|_p f = \frac{\partial \left(f \circ \psi^{-1}\right)}{\partial r^i}(\psi(p)), \tag{3.8}$$

where (U, ψ) is a chart, $p \in U$, and $\psi = \left(x^1, x^2, \dots, x^n\right)$, and $f \circ \psi^{-1} : \psi(U) \subset \mathbb{R}^n \to \mathbb{R}$.

Definition 3.8. Let N and M be manifolds of dimension n and m, respectively. A continuous map $F : N \to M$ is C^∞ at a point $p \in N$ if there are charts (V, ψ) about $F(p) \in M$ and (U, φ) about $p \in N$ such that the composition

$$\psi \circ F \circ \varphi^{-1} : \varphi\left(F^{-1}(V) \cap U\right) \subseteq \mathbb{R}^n \to \mathbb{R}^m,$$

which is a map from an open subset of $\mathbb{R}^n$ to an open subset of $\mathbb{R}^m$, is C^∞ at $\varphi(p)$. The continuous map $F : N \to M$ is said to be C^∞ if it is C^∞ at every point of N.



Let I be an open interval of $\mathbb{R}$ containing 0. Recall that

$$\left.\frac{\partial}{\partial x^\mu}\right|_p f = \left.\frac{\partial}{\partial r^\mu}\right|_{\varphi(p)} \left(f \circ \varphi^{-1}\right), \tag{3.9}$$

where p belongs to the chart (U, φ) of M. Let $\lambda : I \to U \subseteq M$ be a curve. WLOG assume $\lambda(0) = p$. We are interested in finding the tangent vector X_p to the curve λ at the point p. By definition of tangent vector at a point on the manifold, it is a map $C^\infty(U, \mathbb{R}) \to \mathbb{R}$.

Definition 3.9. Let $f \in C^\infty(U, \mathbb{R}) \to \mathbb{R}$, and $\lambda : I \to U \subseteq M$ a curve with $\lambda(0) = p$. Then the tangent vector X_p to the curve λ at the point p is defined as

$$X_p f := \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (f \circ \lambda) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(\lambda(t)). \tag{3.10}$$

Note that $f \circ \lambda$ is a map from I to $\mathbb{R}$, so its derivative is defined in the usual sense. Now, if we write $f \circ \lambda$ as $f \circ \varphi^{-1} \circ \varphi \circ \lambda$ and apply chain rule, then

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (f \circ \lambda) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \left(f \circ \varphi^{-1} \circ \varphi \circ \lambda\right) = \left.\frac{\partial}{\partial r^\mu}\right|_{\varphi(\lambda(0))} \left(f \circ \varphi^{-1}\right) \cdot \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} x^\mu(\lambda(t)). \tag{3.11}$$

Here Einstein summation convention is implied. Therefore,

$$X_p f = \left.\frac{\partial}{\partial x^\mu}\right|_p f \cdot \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} x^\mu(\lambda(t)).$$

$$\therefore \boxed{X_p = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} x^\mu(\lambda(t)) \cdot \left.\frac{\partial}{\partial x^\mu}\right|_p.} \tag{3.12}$$

§3.3 Relationship Between Coordinate Bases

Let $\varphi = \left(x^1, x^2, \ldots, x^n\right)$ and $\varphi' = \left(x'^1, x'^1, \ldots, x'^n\right)$ be two charts defined in neighborhoods U and V of p, respectively. Then for any smooth function f defined in a neighborhood of p,

$$\begin{aligned}
\left.\frac{\partial}{\partial x^\mu}\right|_p f &= \left.\frac{\partial}{\partial r^\mu}\right|_{\varphi(p)} \left(f \circ \varphi^{-1}\right) = \left.\frac{\partial}{\partial r^\mu}\right|_{\varphi(p)} \left(f \circ \varphi'^{-1} \circ \varphi' \circ \varphi^{-1}\right) \\
&= \left.\frac{\partial}{\partial r^\mu}\right|_{\varphi(p)} \left(r^\nu \circ \varphi' \circ \varphi^{-1}\right) \left.\frac{\partial}{\partial r^\nu}\right|_{\varphi'(\varphi^{-1}(\varphi(p)))} \left(f \circ \varphi'^{-1}\right) \\
&= \left.\frac{\partial}{\partial r^\mu}\right|_{\varphi(p)} \left(x'^\nu \circ \varphi^{-1}\right) \left.\frac{\partial}{\partial x'^\nu}\right|_p f \\
&= \frac{\partial x'^\nu}{\partial x^\mu}(p) \left.\frac{\partial}{\partial x'^\nu}\right|_p f.
\end{aligned} \tag{3.13}$$

Therefore,

$$\left.\frac{\partial}{\partial x^\mu}\right|_p = \frac{\partial x'^\nu}{\partial x^\mu}(p) \left.\frac{\partial}{\partial x'^\nu}\right|_p. \tag{3.14}$$

This is the change of basis formula for tangent vectors at a point. Now, let X_p be a tangent vector at p. Suppose its components with respect to the basis $\left\{\left.\frac{\partial}{\partial x^\mu}\right|_p\right\}$ are $X^\mu(p)$, and its components with respect to the basis $\left\{\left.\frac{\partial}{\partial x'^\nu}\right|_p\right\}$ are $X'^\nu(p)$. Then

$$X_p = X^\mu(p) \left.\frac{\partial}{\partial x^\mu}\right|_p = X'^\nu(p) \left.\frac{\partial}{\partial x'^\nu}\right|_p. \tag{3.15}$$

Using the change of basis formula (3.14), we get

$$X_p = X^\mu (p) \left.\frac{\partial}{\partial x^\mu}\right|_p = X^\mu (p) \frac{\partial x'^\nu}{\partial x^\mu} (p) \left.\frac{\partial}{\partial x'^\nu}\right|_p . \tag{3.16}$$

Equating this with $X'^\nu (p) \left.\frac{\partial}{\partial x'^\nu}\right|_p$, we get

$$X'^\nu (p) = \frac{\partial x'^\nu}{\partial x^\mu} (p) X^\mu (p) . \tag{3.17}$$

Therefore, in a change of coordinates, the components of a tangent vector transform in the abovementioned way. This is often called "contravariant transformation" in many GR texts.

§3.4 Covectors

Definition 3.10 (Dual Space). Let V be a vector space over the field $\mathbb{F}$. The **dual vector space** V^* of V is the space of all linear maps from V to $\mathbb{F}$.

If $\dim V = n$, then $\dim V^*$ is also n. If $\{e_1, \ldots, e_n\}$ is a basis of V, the "dual basis" of V^* is $\{f^1, f^2, \ldots, f^n\}$ such that $f^\mu (e_\nu) = \delta^\mu_\nu$. If $X = X^\mu e_\mu$ is a generic element of V,

$$f^\mu (X) = f^\mu (X^\nu e_\nu) = X^\nu f^\mu(e_\nu) = X^\nu \delta^\mu_\nu = X^\mu . \tag{3.18}$$

Since V and V^* have the same dimension, they are isomorphic vector spaces. The isomorphism is given by $e_\mu \mapsto f^\mu$. However, this isomorphism is basis dependent. There is a more natural isomorphism between V and $(V^*)^*$.

Theorem 3.2

If V is a finite dimensional vector space, then $(V^*)^*$ is naturally isomorphic to V with the isomorphism

$$\Phi : V \to (V^*)^* , \quad \text{with } \Phi (X) (\omega) = \omega (X) . \tag{3.19}$$

Now we get back to manifolds.

Definition 3.11. The dual space of T_pM, denoted by T^*_pM, is called the **cotangent space** of M at $p \in M$. An element of this space is called a covector.

Definition 3.12 (Vector field). A **vector field** is a map X which assigns any point $p \in M$ to a tangent vector X_p at p. Given a vector field X and a function f, we can define a new function $X(f) : M \to \mathbb{R}$ by $p \mapsto X_p f$. The vector field X is **smooth** if this map is a smooth function for any smooth f.

We shall mostly deal with smooth vector fields. So unless stated otherwise, a vector field will always mean a smooth vector field. One can also think of a vector field as a smooth map $X : C^\infty (M, \mathbb{R}) \to C^\infty (M, \mathbb{R})$, defined as

$$(Xf)(p) = X_p f, \tag{3.20}$$

for a given $f \in C^\infty (M, \mathbb{R})$.

Definition 3.13 (1-form). The dual notion of vector field on a smooth manifold is called 1-form. A 1-form ω assigns to each point $p \in M$ a covector ω_p at p.

Given a smooth real-valued function f on M, one can write down a 1-form $\mathrm{d}f$ such that $\mathrm{d}f\big|_p \in T_p^*M$ is given by

$$\mathrm{d}f\big|_p \left(X\big|_p\right) = Xf(p) = X\big|_p f. \tag{3.21}$$

Note that we denote vector field and 1-form by X and $\mathrm{d}f$, respectively, and vector and covector by $X\big|_p$ and $\mathrm{d}f\big|_p$, respectively.

Now, let $\left(x^1, \dots, x^n\right)$ be a chart defined in a neighborhood of p. Then a basis of T_pM is, as we know

$$\left\{ \frac{\partial}{\partial x^1}\bigg|_p, \dots, \frac{\partial}{\partial x^n}\bigg|_p \right\}.$$

The dual basis of T_p^*M is

$$\left\{ \mathrm{d}x^1\big|_p, \mathrm{d}x^2\big|_p, \dots, \mathrm{d}x^n\big|_p \right\}. \tag{3.22}$$

The action of $\mathrm{d}x^\mu\big|_p$ on $\frac{\partial}{\partial x^\nu}\big|_p$ is given by

$$\mathrm{d}x^\mu\big|_p \left(\frac{\partial}{\partial x^\nu}\bigg|_p \right) = \frac{\partial}{\partial x^\nu}\bigg|_p x^\mu = \frac{\partial x^\mu}{\partial x^\nu}(p) = \delta^\mu_\nu. \tag{3.23}$$

Now we want to know how do the covector components $\omega_\mu(p)$ of a given covector $\omega_p = \omega_\mu(p)\, \mathrm{d}x^\mu\big|_p$ transform under change of coordinates. We consider two different charts $\phi = \left(x^1, \dots, x^n\right)$ and $\phi' = \left(x', \dots, x'^n\right)$ defined in a neighbourhood of p. Relabeling x and x' in 3.14, we obtain

$$\frac{\partial}{\partial x'^\mu}\bigg|_p = \frac{\partial x^\nu}{\partial x'^\mu}(p)\, \frac{\partial}{\partial x^\nu}\bigg|_p. \tag{3.24}$$

Now, both $\left\{ dx^1\big|_p, dx^2\big|_p, \dots, dx^n\big|_p \right\}$ and $\left\{ dx'^1\big|_p, dx'^2\big|_p, \dots, dx'^n\big|_p \right\}$ are bases of T_p^*M. Therefore, $dx^\mu\big|_p$ can be written as a linear combination of $dx'^\nu\big|_p$'s.

$$dx^\mu\big|_p = a^\mu_\rho\, dx'^\rho\big|_p. \tag{3.25}$$

Now, applying $\frac{\partial}{\partial x'^\nu}\big|_p$ on both sides of (3.25), we get

$$dx^\mu\big|_p \left(\frac{\partial}{\partial x'^\nu}\bigg|_p \right) = a^\mu_\rho\ dx'^\rho\big|_p \left(\frac{\partial}{\partial x'^\nu}\bigg|_p \right) = a^\mu_\rho \delta^\rho_\nu = a^\mu_\nu. \tag{3.26}$$

Now, let's evaluate the LHS of (3.26) by expanding $\frac{\partial}{\partial x'^\nu}\big|_p$ using (3.24).

$$\begin{aligned} dx^\mu\big|_p \left(\frac{\partial}{\partial x'^\nu}\bigg|_p \right) &= dx^\mu\big|_p \left(\frac{\partial x^\sigma}{\partial x'^\nu}(p)\ \frac{\partial}{\partial x^\sigma}\bigg|_p \right) \\ &= \frac{\partial x^\sigma}{\partial x'^\nu}(p)\, dx^\mu\big|_p \left(\frac{\partial}{\partial x^\sigma}\bigg|_p \right) \\ &= \frac{\partial x^\sigma}{\partial x'^\nu}(p)\, \delta^\mu_\sigma \\ &= \frac{\partial x^\mu}{\partial x'^\nu}(p). \end{aligned} \tag{3.27}$$

Therefore, $a^\mu_\nu = \frac{\partial x^\mu}{\partial x'^\nu}(p)$. Hence,

$$dx^\mu\big|_p = a^\mu_\nu\, dx'^\nu\big|_p = \frac{\partial x^\mu}{\partial x'^\nu}(p)\ dx'^\nu\big|_p. \tag{3.28}$$

Consider a generic covector $\omega_p \in T_p^* M$. We write it in both the charts as follows:

$$\omega_p = \omega_\mu (p) \; dx^\mu \big|_p, \tag{3.29}$$

$$\omega_p = \omega'_\nu (p) \; dx'^\nu \big|_p. \tag{3.30}$$

Now, using (3.28),

$$\begin{aligned} \omega_p &= \omega_\mu (p) \; dx^\mu \big|_p \\ &= \omega_\mu (p) \; \frac{\partial x^\mu}{\partial x'^\nu} (p) \; dx'^\nu \big|_p \\ &= \frac{\partial x^\mu}{\partial x'^\nu} (p) \; \omega_\mu (p) \; dx'^\nu \big|_p. \end{aligned} \tag{3.31}$$

Since $\omega_p = \omega'_\nu (p) \; dx'^\nu \big|_p$, and basis decomposition is unique, from (3.30) and (3.31), we get

$$\omega'_\nu (p) = \frac{\partial x^\mu}{\partial x'^\nu} (p) \; \omega_\mu (p) \, . \tag{3.32}$$

4 Tensors and Tensor Field

§4.1 Tensor

Definition 4.1. A tensor of type (r,s) at p is a multilinear map

$$T\big|_p : T_p^*(M)^r \times T_p(M)^s := \underbrace{T_p^*(M) \times \cdots \times T_p^*(M)}_{r \text{ factors}} \times \underbrace{T_p(M) \times \cdots \times T_p(M)}_{s \text{ factors}} \to \mathbb{R}. \tag{4.1}$$

Multilinear means the map is linear in each argument.

Example 4.1

A tensor of type $(0,1)$ is a linear map $T_pM \to \mathbb{R}$, i.e. it's a covector. A tensor of type $(1,0)$ is a linear map $T_p^*M \to \mathbb{R}$, i.e. it is an element of $\left(T_p^*M\right)^*$. But $\left(T_p^*M\right)^*$ is naturally isomorphic to T_pM. Hence, a tensor of type $(1,0)$ is a tangent vector. Given a vector $X_p \in T_pM$, we define a linear map $T_p^*M \to \mathbb{R}$ by

$$\eta_p \mapsto \eta_p(X_p) \in \mathbb{R}, \tag{4.2}$$

for any $\eta_p \in T_p^*M$.

Example 4.2

We can define a $(1,1)$ tensor δ by

$$\delta_p : T_p^*M \times T_pM \to \mathbb{R}\,, \quad \text{with } \delta_p(\omega_p, X_p) = \omega_p(X_p)\,, \tag{4.3}$$

for any $\omega_p \in T_p^*M$ and $X_p \in T_pM$.

Definition 4.2. Let $T\big|_p$ be a tensor of type (r,s) at p. If $\left\{\frac{\partial}{\partial x^\mu}\big|_p\right\}$ is a basis for T_pM with dual basis $\left\{\mathrm{d}x^\mu\big|_p\right\}$ of T_p^*M, then the components of $T\big|_p$ in this basis are the numbers

$$T^{\mu_1\cdots\mu_r}{}_{\nu_1\ldots\nu_s}(p) = T\big|_p\left(\mathrm{d}x^{\mu_1}\big|_p, \ldots, \mathrm{d}x^{\mu_r}\big|_p, \frac{\partial}{\partial x^{\nu_1}}\bigg|_p, \ldots, \frac{\partial}{\partial x^{\nu_s}}\bigg|_p\right). \tag{4.4}$$

In the abstract index notation, we denote $T\big|_p$ by $T^{\mu_1\cdots\mu_r}{}_{\nu_1\ldots\nu_s}(p)$.

Remark 4.1. Tensors at p can be added together and multiplied by a scalar. Hence, $T\big|_p$'s form a vector space denoted by $\mathscr{T}_p(r,s)$. $T\big|_p$'s can be written as a linear combination of the following vectors

$$\frac{\partial}{\partial x^{\mu_1}}\bigg|_p \otimes \cdots \otimes \frac{\partial}{\partial x^{\mu_r}}\bigg|_p \otimes \mathrm{d}x^{\nu_1}\big|_p \otimes \cdots \otimes \mathrm{d}x^{\nu_s}\big|_p. \tag{4.5}$$

The coefficients of such linear combinations are precisely the numbers $T^{\mu_1\cdots\mu_r}{}_{\nu_1\ldots\nu_s}(p)$. The vectors of the form 4.5 form a basis of $\mathscr{T}_p(r,s)$. The action of these basis vectors on $T_p^*(M)^r \times T_p(M)^s$ is given by

$$\left(\frac{\partial}{\partial x^{\mu_1}}\bigg|_p \otimes \cdots \otimes \frac{\partial}{\partial x^{\mu_r}}\bigg|_p \otimes \mathrm{d}x^{\nu_1}\big|_p \otimes \cdots \otimes \mathrm{d}x^{\nu_s}\big|_p\right)\left(\mathrm{d}x^{\rho_1}\big|_p, \ldots, \mathrm{d}x^{\rho_r}\big|_p, \frac{\partial}{\partial x^{\sigma_1}}\bigg|_p, \ldots, \frac{\partial}{\partial x^{\sigma_s}}\bigg|_p\right)$$
$$= \delta^{\rho_1}{}_{\mu_1}\cdots\delta^{\rho_r}{}_{\mu_r}\delta^{\nu_1}{}_{\sigma_1}\cdots\delta^{\nu_s}{}_{\sigma_s}. \tag{4.6}$$

If $\dim T_pM = T_p^*M = n$, then there are n^{r+s} basis elements of $\mathscr{T}_p(r,s)$. Therefore, $\dim \mathscr{T}_p(r,s) = n^{r+s}$.

Remark 4.2. By "tensor" physicists mean the components $T^{\mu_1 \cdots \mu_r}{}_{\nu_1 \ldots \nu_s}(p)$ of $T\big|_p$.

Transformation of Tensors Under Change of Coordinates

Consider two charts $\phi = \left(x^1, x^2, \ldots, x^n\right)$ and $\phi' = \left(x'^1, \ldots, x'^n\right)$ about $p \in M$, and take an element $T\big|_p \in \mathscr{T}_p(r,s)$. Then

$$T\big|_p = T^{\mu_1 \cdots \mu_r}{}_{\nu_1 \ldots \nu_s}(p) \left.\frac{\partial}{\partial x^{\mu_1}}\right|_p \otimes \cdots \otimes \left.\frac{\partial}{\partial x^{\mu_r}}\right|_p \otimes \mathrm{d}x^{\nu_1}\big|_p \otimes \cdots \otimes \mathrm{d}x^{\nu_s}\big|_p. \tag{4.7}$$

This is the expression of $T\big|_p$ in the chart ϕ. In the chart ϕ', the expression for $T\big|_p$ is

$$T\big|_p = T'^{\sigma_1 \ldots \sigma_r}{}_{\rho_1 \ldots \rho_s} \left.\frac{\partial}{\partial x'^{\sigma_1}}\right|_p \otimes \cdots \otimes \left.\frac{\partial}{\partial x'^{\sigma_r}}\right|_p \otimes \mathrm{d}x'^{\rho_1}\big|_p \otimes \cdots \otimes \mathrm{d}x'^{\rho_s}\big|_p. \tag{4.8}$$

Using 3.14 and 3.28 in 4.7, we get

$$\begin{aligned} T\big|_p &= T^{\mu_1 \cdots \mu_r}{}_{\nu_1 \ldots \nu_s}(p) \left.\frac{\partial}{\partial x^{\mu_1}}\right|_p \otimes \cdots \otimes \left.\frac{\partial}{\partial x^{\mu_r}}\right|_p \otimes \mathrm{d}x^{\nu_1}\big|_p \otimes \cdots \otimes \mathrm{d}x^{\nu_s}\big|_p \\ &= T^{\mu_1 \cdots \mu_r}{}_{\nu_1 \ldots \nu_s}(p) \prod_{i=1}^{r} \frac{\partial x'^{\sigma_i}}{\partial x^{\mu_i}}(p) \prod_{j=1}^{s} \frac{\partial x^{\nu_j}}{\partial x'^{\rho_j}}(p) \left.\frac{\partial}{\partial x'^{\sigma_1}}\right|_p \otimes \cdots \otimes \left.\frac{\partial}{\partial x'^{\sigma_r}}\right|_p \otimes \mathrm{d}x'^{\rho_1}\big|_p \otimes \cdots \otimes \mathrm{d}x'^{\rho_s}\big|_p. \end{aligned} \tag{4.9}$$

Combining 4.8 and 4.9, we get

$$T'^{\sigma_1 \ldots \sigma_r}{}_{\rho_1 \ldots \rho_s} = T^{\mu_1 \cdots \mu_r}{}_{\nu_1 \ldots \nu_s}(p) \prod_{i=1}^{r} \frac{\partial x'^{\sigma_i}}{\partial x^{\mu_i}}(p) \prod_{j=1}^{s} \frac{\partial x^{\nu_j}}{\partial x'^{\rho_j}}(p). \tag{4.10}$$

§4.2 Tensor Field

Recall that a tangent vector at $p \in M$ is a map $X\big|_p : C^\infty(M,\mathbb{R}) \to \mathbb{R}$, i.e., $X\big|_p f$ is a number given $f \in C^\infty(M,\mathbb{R})$. A vector field X, on the other hand, is the map $X : C^\infty(M,\mathbb{R}) \to C^\infty(M,\mathbb{R})$ given by

$$(Xf)(p) = X\big|_p f. \tag{4.11}$$

So Xf is a smooth function on the manifold.

Definition 4.3. An (r,s) **tensor field** is a map T that maps any point $p \in M$ to an (r,s) tensor $T\big|_p$ at p. Given r 1-forms $\eta_1, \ldots, \eta_r$ and s vector fields $X_1, \ldots, X_s$, we can define a function $M \to \mathbb{R}$ by

$$p \mapsto T\big|_p \left(\eta_1\big|_p, \ldots, \eta_r\big|_p, X_1\big|_p, \ldots, X_s\big|_p\right). \tag{4.12}$$

The tensor field T is called **smooth** if this function is smooth for any r 1-forms $\eta_1, \ldots, \eta_r$ and s vector fields $X_1, \ldots, X_s$.

Commutator of two vector fields is again a vector field. In other words, given two vector fields X and Y, $[X,Y]$ defined by

$$[X,Y]f := X(Y(f)) - Y(X(f)) \tag{4.13}$$

is also a vector field. Let us now compute the components of this vector field in a given chart $\left(x^1, \ldots, x^n\right)$.

$$\begin{aligned}
[X, Y] f &= X\left(Y\left(f\right)\right) - Y\left(X\left(f\right)\right) \\
&= X^\mu \frac{\partial}{\partial x^\mu}\left(Y^\nu \frac{\partial f}{\partial x^\nu}\right) - Y^\nu \frac{\partial}{\partial x^\nu}\left(X^\mu \frac{\partial f}{\partial x^\mu}\right) \\
&= X^\mu \frac{\partial Y^\nu}{\partial x^\mu}\frac{\partial f}{\partial x^\nu} + X^\mu Y^\nu \frac{\partial^2 f}{\partial x^\mu \partial x^\nu} - Y^\nu \frac{\partial X^\mu}{\partial x^\nu}\frac{\partial f}{\partial x^\mu} - Y^\nu X^\mu \frac{\partial^2 f}{\partial x^\nu \partial x^\mu} \\
&= X^\mu \frac{\partial Y^\nu}{\partial x^\mu}\frac{\partial f}{\partial x^\nu} - Y^\nu \frac{\partial X^\mu}{\partial x^\nu}\frac{\partial f}{\partial x^\mu} \\
&= \left(X^\nu \frac{\partial Y^\mu}{\partial x^\nu} - Y^\nu \frac{\partial X^\mu}{\partial x^\nu}\right)\frac{\partial f}{\partial x^\mu}. \qquad (4.14)
\end{aligned}$$

In the component form,

$$[X, Y] f = [X, Y]^\mu \frac{\partial f}{\partial x^\mu}. \qquad (4.15)$$

Combining 4.14 and 4.15, we get

$$[X, Y]^\mu = X^\nu \frac{\partial Y^\mu}{\partial x^\nu} - Y^\nu \frac{\partial X^\mu}{\partial x^\nu}. \qquad (4.16)$$

Remark 4.3. Since the components of $\frac{\partial}{\partial x^\mu}$ in the coordinate basis are either 0 or 1, it follows that

$$\left[\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu}\right] = 0. \qquad (4.17)$$

§4.3 The Metric Tensor

Consider a curve $\mathbf{x}\left(t\right)$ in $\mathbb{R}^3$ with $a < t < b$. The length of this curve is

$$\int_a^b \mathrm{d}t \left\|\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}\right\| = \int_a^b \mathrm{d}t \sqrt{\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} \cdot \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}}. \qquad (4.18)$$

Motivated by the above scenario, we define the metric tensor as follows.

Definition 4.4 (Metric Tensor)**.** A metric tensor at $p \in M$ is a tensor of type $(0, 2)$, i.e. it is a map $g\big|_p : T_pM \times T_pM \to \mathbb{R}$ with the following properties:

1. (Symmetry) For $X_p, Y_p \in T_pM$, $g\big|_p\left(X_p, Y_p\right) = g\big|_p\left(Y_p, X_p\right)$.
2. (Non-degeneracy) If $g\big|_p\left(X_p, Y_p\right) = 0$ for all $Y_p \in T_pM$, then $X_p = 0$.

Remark 4.4. In the chart x^μ, $g\big|_p = g_{\mu\nu}\left(p\right) \mathrm{d}x^\mu\big|_p \otimes \mathrm{d}x^\nu\big|_p$. g, as a $(0, 2)$ tensor field, can also be written as $g = g_{\mu\nu}\mathrm{d}x^\mu \otimes \mathrm{d}x^\nu$. Since the components $g_{\mu\nu}\left(p\right)$ of $g\big|_p$ form a symmetric matrix (from the symmetry of metric tensor), such a matrix can be diagonalized following an appropriate choice of basis vectors. Non-degeneracy implies that none of the diagonal elements will be 0. Because, if $g_{ii} = 0$ after diagonalization, we take $X_p \in T_pM$ such that $X^i = 1$ and $X^j = 0$ for $j \neq i$ in that basis. Then

$$g\big|_p\left(X_p, Y_p\right) = g_{\mu\nu}X^\mu Y^\nu = g_{i\nu}X^i Y^\nu = g_{ii}X^i Y^i = 0, \qquad (4.19)$$

for any $Y_p \in T_pM$. This contradicts non-degenracy. So none of the diagonal elements are 0. Then after scaling the basis vectors appropriately, we can make the diagonal elements ± 1. Such a basis is then called an **orthonormal** basis.

In differential geometry, one is interested in Riemannian metrics. For such metrics, the signature is $+ + \cdots +$, i.e., all diagonal elements are $+1$ in an orthonormal basis. But in general relativity, we are interested in Lorentzian metrics, i.e., those are with signature $- + + \cdots +$.

Definition 4.5. A **Riemannian (Lorentzian) manifold** is a pair (M, g) where M is a Riemannian (Lorentzian) manifold and g is a Riemannian (Lorentzian) metric tensor field.

On a Riemannian manifold, one can now define the length of a curve as in $\mathbb{R}^3$: if $\lambda : (a, b) \to M$ is a smooth curve with tangent vector $X|_{\lambda(t)}$, then its length is

$$\int_a^b \mathrm{d}t \ g|_{\lambda(t)} \left(X|_{\lambda(t)}, X|_{\lambda(t)} \right) = \int_a^b \mathrm{d}t \ g\left(X, X\right)\left(\lambda\left(t\right)\right). \tag{4.20}$$

Example 4.3

In $\mathbb{R}^n$, the Euclidean metric is

$$g = \mathrm{d}x^1 \otimes \mathrm{d}x^1 + \mathrm{d}x^2 \otimes \mathrm{d}x^2 + \cdots + \mathrm{d}x^n \otimes \mathrm{d}x^n. \tag{4.21}$$

$(\mathbb{R}^n, g)$ is called the Euclidean space. A coordinate chart which covers all of $\mathbb{R}^n$ and in which the components of the metric are $\mathrm{diag}\left(1, 1, \ldots, 1\right)$ is called Cartesian.

Example 4.4

In $\mathbb{R}^4$, the **Minkowski metric** is

$$\eta = -\mathrm{d}x^0 \otimes \mathrm{d}x^0 + \mathrm{d}x^1 \otimes \mathrm{d}x^1 + \mathrm{d}x^2 \otimes \mathrm{d}x^2 + \mathrm{d}x^3 \otimes \mathrm{d}x^3. \tag{4.22}$$

$\left(\mathbb{R}^4, \eta\right)$ is called the Minkowski spacetime. A coordinate chart which covers all of $\mathbb{R}^4$ and in which the components of the metric are $\eta_{\mu\nu} = \mathrm{diag}\left(-1, 1, 1, 1\right)$ is called an **inertial frame**.

Example 4.5

On S^2, let (θ, ϕ) denote the spherical polar coordinate chart. The round unit metric on S^2 is

$$g = \mathrm{d}\theta \otimes \mathrm{d}\theta + \sin^2\theta \mathrm{d}\phi \otimes \mathrm{d}\phi. \tag{4.23}$$

In the chart (θ, ϕ), we have $g_{\mu\nu} = \mathrm{diag}\left(1, \sin^2\theta\right)$.

§4.4 Lorentzian Signature

On a Lorentzian manifold, we take basis indices μ, ν to run from 0 to $n-1$. At $p \in M$, we choose an orthonormal basis $\{e_\mu\}$ so that the matrix with entries of the metric components gets diagonalized, i.e.

$$g|_p \left(e_\mu|_p, e_\nu|_p \right) = \eta_{\mu\nu} = \mathrm{diag}\left(-1, 1, 1, \ldots, 1\right). \tag{4.24}$$

Such a basis is not unique. Let $\left\{ e'_\mu|_p \right\}$ be any other such basis with

$$e'_\mu|_p = \left(A^{-1}\right)^\nu{}_\mu e_\nu|_p. \tag{4.25}$$

Then we have

$$\begin{aligned}\eta_{\mu\nu} &= g|_p \left(e'_\mu|_p, e'_\nu|_p \right) = g|_p \left(\left(A^{-1}\right)^\sigma{}_\mu e_\sigma|_p, \left(A^{-1}\right)^\rho{}_\nu e_\rho|_p \right) \\ &= \left(A^{-1}\right)^\sigma{}_\mu \left(A^{-1}\right)^\rho{}_\nu g|_p \left(e_\sigma|_p, e_\rho|_p \right) = \left(A^{-1}\right)^\sigma{}_\mu \left(A^{-1}\right)^\rho{}_\nu \eta_{\sigma\rho}. \\ \therefore \eta_{\sigma\rho} &= \eta_{\mu\nu} A^\mu{}_\sigma A^\nu_\rho. \end{aligned} \tag{4.26}$$

4.26 is precisely the defining equations of a Lorentz transformation in special relativity. Hence, different orthonormal basis at $p \in M$ are related by Lorentz transformations. The A's are actually the Λ's we see in special relativity.

> **Definition 4.6.** On a Lorentzian manifold (M, g), a non-zero vector $X_p \in T_pM$ is **timelike** if $g\big|_p (X_p, X_p) < 0$, **null** (or **lightlike**) if $g\big|_p (X_p, X_p) = 0$, and **spacelike** if $g\big|_p (X_p, X_p) > 0$.

Remark 4.5. In an orthonormal basis at p, the metric has components $\eta_{\mu\nu}$. So, the tangent space at p has exactly the same structure as Minkowski spacetime, i.e., null vectors at p define a light cone that separates timelike vectors at p from spacelike vectors at p.

§4.5 Geodesic Equation

Let p and q be points connected by a timelike curve, i.e. the tangent vectors at all point of the curve are timelike. There are infinitely many timelike curves between p and q. The proper time between p and q will be different for different curves. It's a natural question to ask which curve extremizes proper time.

Consider timelike curves from p to q with parameter u such that $\lambda(0) = p$ and $\lambda(1) = q$. The proper time between p and q along such a curve is given by the functional

$$\tau[\lambda] = \int_0^1 \mathrm{d}u \, G(x(u), \dot{x}(u)), \tag{4.27}$$

where $G(x(u), \dot{x}(u)) = \sqrt{-g_{\mu\nu}(\lambda(u))\,\dot{x}^\mu(u)\,\dot{x}^\nu(u)}$. We are writing $x^\mu(u)$ as a shorthand for $x^\mu(\lambda(u))$. The curve that extremizes proper time must satisfy the Euler-Lagrange equation.

$$\frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{\partial G}{\partial \dot{x}^\mu}\right) - \frac{\partial G}{\partial x^\mu} = 0. \tag{4.28}$$

Recall that

$$\begin{aligned}
G(x(\lambda(u)), \dot{x}(\lambda(u))) &= \sqrt{-g_{\sigma\nu}(\lambda(u))\dot{x}^\sigma(\lambda(u))\dot{x}^\nu(\lambda(u))}. &(4.29)\\
\frac{\partial G}{\partial \dot{x}^\mu} &= \frac{1}{2G}\left[-g_{\sigma\nu}\delta^\sigma_\mu \dot{x}^\nu - g_{\sigma\nu}\dot{x}^\sigma\delta^\nu_\mu\right]\\
&= \frac{1}{2G}\left[-g_{\mu\nu}\dot{x}^\nu - g_{\sigma\mu}\dot{x}^\sigma\right]\\
&= \frac{1}{2G}\left[-g_{\mu\nu}\dot{x}^\nu - g_{\mu\sigma}\dot{x}^\sigma\right]\\
&= \frac{1}{2G}\left(-2g_{\mu\nu}\dot{x}^\nu\right) = -\frac{g_{\mu\nu}}{G}\dot{x}^\nu. &(4.30)\\
\frac{\partial G}{\partial x^\mu} &= \frac{1}{2G}\left[-g_{\sigma\nu,\mu}\dot{x}^\sigma\dot{x}^\nu\right]. &(4.31)
\end{aligned}$$

Now, recall

$$\frac{\mathrm{d}\tau}{\mathrm{d}u} = \sqrt{-g_{\mu\nu}(\lambda(u))\frac{\mathrm{d}x^\mu(\lambda(u))}{\mathrm{d}u}\frac{\mathrm{d}x^\nu(\lambda(u)}{\mathrm{d}u}} = G. \tag{4.32}$$

So $\frac{\mathrm{d}}{\mathrm{d}u} = G\frac{\mathrm{d}}{\mathrm{d}\tau}$. Now Euler-Lagrange equation reads

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}u}\left(-\frac{g_{\mu\nu}}{G}\dot{x}^\nu\right) + \frac{1}{2G}g_{\sigma\nu,\mu}\dot{x}^\sigma\dot{x}^\nu = 0\\
\implies\; & -G\frac{\mathrm{d}}{\mathrm{d}\tau}\left(g_{\mu\nu}\frac{1}{G}G\frac{\mathrm{d}}{\mathrm{d}\tau}x^\nu\right) + \frac{1}{2G}g_{\sigma\nu,\mu}G\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau}\cdot G\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} = 0\\
\implies\; & -G\frac{\mathrm{d}}{\mathrm{d}\tau}\left(g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right) + \frac{G}{2}g_{\sigma\nu,\mu}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} = 0\\
\therefore\; & \frac{\mathrm{d}}{\mathrm{d}\tau}\left(g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right) - \frac{1}{2}g_{\sigma\nu,\mu}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} = 0. \qquad (4.33)
\end{aligned}$$

Hence,

$$\begin{aligned} & g_{\mu\nu}\frac{\mathrm{d}^2x^\nu}{\mathrm{d}\tau^2}+\frac{\mathrm{d}}{\mathrm{d}\tau}g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}-\frac{1}{2}g_{\sigma\nu,\mu}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}=0 \\ \Longrightarrow\ & g_{\mu\nu}\frac{\mathrm{d}^2x^\nu}{\mathrm{d}\tau^2}+\frac{\partial g_{\mu\nu}}{\partial x^\rho}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}-\frac{1}{2}g_{\sigma\nu,\mu}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}=0 \\ \Longrightarrow\ & g_{\mu\nu}\frac{\mathrm{d}^2x^\nu}{\mathrm{d}\tau^2}+g_{\mu\nu,\rho}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}-\frac{1}{2}g_{\rho\nu,\mu}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}=0. \end{aligned} \tag{4.34}$$

Now, observe that

$$\frac{\partial g_{\mu\nu}}{\partial x^\rho}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}=\frac{\partial g_{\mu\rho}}{\partial x^\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}. \tag{4.35}$$

Therefore, we write

$$\begin{aligned} g_{\mu\nu,\rho}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} &= \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\rho}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}+\frac{1}{2}\frac{\partial g_{\mu\rho}}{\partial x^\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau} \\ &= \frac{1}{2}\left(g_{\mu\nu,\rho}+g_{\mu\rho,\nu}\right)\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}. \end{aligned} \tag{4.36}$$

Now 4.34 reads

$$\begin{aligned} & g_{\mu\nu}\frac{\mathrm{d}^2x^\nu}{\mathrm{d}\tau^2}+g_{\mu\nu,\rho}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}-\frac{1}{2}g_{\rho\nu,\mu}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}=0 \\ \Longrightarrow\ & g_{\mu\nu}\frac{\mathrm{d}^2x^\nu}{\mathrm{d}\tau^2}+\frac{1}{2}\left(g_{\mu\nu,\rho}+g_{\mu\rho,\nu}-g_{\nu\rho,\mu}\right)\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}=0. \end{aligned} \tag{4.37}$$

Contracting with $g^{\sigma\mu}$ on both sides, we get

$$\begin{aligned} & g^{\sigma\mu}g_{\mu\nu}\frac{\mathrm{d}^2x^\nu}{\mathrm{d}\tau^2}+\frac{1}{2}g^{\sigma\mu}\left(g_{\mu\nu,\rho}+g_{\mu\rho,\nu}-g_{\nu\rho,\mu}\right)\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}=0 \\ \Longrightarrow\ & \delta^\sigma_\nu\frac{\mathrm{d}^2x^\nu}{\mathrm{d}\tau^2}+\underbrace{\frac{1}{2}g^{\sigma\mu}\left(g_{\mu\nu,\rho}+g_{\mu\rho,\nu}-g_{\nu\rho,\mu}\right)}_{\Gamma^\sigma_{\nu\rho}}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}=0 \\ \therefore\ & \frac{\mathrm{d}^2x^\sigma}{\mathrm{d}\tau^2}+\Gamma^\sigma_{\nu\rho}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}=0. \end{aligned} \tag{4.38}$$

$\Gamma^\sigma_{\nu\rho}$'s are called **Christoffel symbols**, and 4.38 is called the **geodesic equation**.

Remark 4.6. In Minkowski spacetime, metric in an inertial reference frame is constant. Therefore, $\Gamma^\sigma_{\nu\rho}$'s are all 0. Therefore, the geodesic equation reduces to

$$\frac{\mathrm{d}^2x^\sigma}{\mathrm{d}\tau^2}=0, \tag{4.39}$$

which is the equation of motion of a free particle.

5 Connection

§5.1 Covariant Derivative

Motivation

Partial derivative of a function belonging to $C^\infty(M,\mathbb{R})$ is denoted by

$$f_{,\mu} := \frac{\partial f}{\partial x^\mu}. \tag{5.1}$$

This is the component of a 1-form $\mathrm{d}f$.

$$\mathrm{d}f\big|_p = \frac{\partial}{\partial x^\nu}\bigg|_p f \mathrm{d}x^\nu\big|_p. \tag{5.2}$$

So $\frac{\partial f}{\partial x^\nu}(p)$ are the components of a covector $\mathrm{d}f\big|_p$. We can restate this fact by saying that the gradient –the partial derivative– of a scalar is a $(0,1)$ tensor. However, the partial derivative of a 1-form does not transform like a tensor. Under a change of coordinate, the components of $T_{\nu\mu} = \frac{\partial \omega_\nu}{\partial x^\mu}$ transform as follows:

$$\begin{aligned} T'_{\nu\mu} &= \frac{\partial}{\partial x'^\mu}\omega'_\nu = \frac{\partial x^\sigma}{\partial x'^\mu}\frac{\partial}{\partial x^\sigma}\left(\frac{\partial x^\rho}{\partial x'^\nu}\omega_\rho\right) \\ &= \frac{\partial x^\sigma}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}\frac{\partial \omega_\rho}{\partial x^\sigma} + \omega_\rho \frac{\partial x^\sigma}{\partial x'^\mu}\frac{\partial^2 x^\rho}{\partial x^\sigma \partial x'^\nu} \\ &= \frac{\partial x^\sigma}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}T_{\rho\sigma} + \omega_\rho \frac{\partial x^\sigma}{\partial x'^\mu}\frac{\partial^2 x^\rho}{\partial x^\sigma \partial x'^\nu}. \end{aligned} \tag{5.3}$$

The second term in 5.3 confirms that $T_{\nu\mu}$ does not transform as a tensor components. In a similar manner, one can show that given a vector field V, $T^\mu{}_\nu = V^\mu_{,\nu}$ does not transform as tensor components.

Definition 5.1. Let $\mathfrak{X}(M)$ be the space of all vector fields on M. Then a **covariant derivative** on M is a map

$$\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)\,,\ (X,Y) \mapsto \nabla_X Y, \tag{5.4}$$

satisfying the following properties:

1. $\nabla_{fX+gY}Z = f\nabla_X Z + g\nabla_Y Z$, where $f,g \in C^\infty(M)$ and $X,Y,Z \in \mathfrak{X}(M)$.
2. $\nabla_X(Y+Z) = \nabla_X Y + \nabla_X Z$.
3. $\nabla_X(fY) = f\nabla_X Y + (\nabla_X f)Y$, where $\nabla_X f$ is defined as $\nabla_X f = X(f)$.

Remark 5.1. If Y is a $(1,0)$ tensor field (a vector field), then ∇Y can be considered a $(1,1)$ tensor field in the following way

$$(\nabla Y)(\eta, X) = \eta(\nabla_X Y) \in C^\infty(M,\mathbb{R})\,. \tag{5.5}$$

In other words, ∇ increases the covariant index by 1. The $(1,1)$ tensor field ∇Y has components $(\nabla Y)^\mu{}_\nu$, also denoted by $Y^\mu_{;\nu}$.

Consider 5.5 again.

$$\begin{aligned} (\nabla Y)(\eta, X) = \eta(\nabla_X Y) &\implies (\nabla Y)^\mu{}_\nu \eta_\mu X^\nu = \eta_\mu (\nabla_X Y)^\mu\,. \\ \therefore (\nabla_X Y)^\mu &= (\nabla Y)^\mu{}_\nu X^\nu = Y^\mu_{;\nu}X^\nu. \end{aligned} \tag{5.6}$$

∇ is also called the connection.

Definition 5.2. In a basis $\left\{\frac{\partial}{\partial x^\mu}\right\}$, the **connection components** $\Gamma^\mu_{\nu\rho}$ are defined by

$$\nabla_{\frac{\partial}{\partial x^\rho}} \frac{\partial}{\partial x^\nu} := \Gamma^\mu_{\nu\rho} \frac{\partial}{\partial x^\mu}. \tag{5.7}$$

Remark 5.2. In general, the connection components are not the same as Christoffel symbols. However, we shall see that the Christoffel symbols are the components of a very special kind of connection, called the Levi-Civita connection.

Now, consider two vector fields $X = X^\mu \frac{\partial}{\partial x^\mu}$ and $Y = Y^\nu \frac{\partial}{\partial x^\nu}$. Then

$$\begin{aligned}
\nabla_X Y &= \nabla_{X^\mu \frac{\partial}{\partial x^\mu}} \left(Y^\nu \frac{\partial}{\partial x^\nu} \right) = X^\mu \nabla_{\frac{\partial}{\partial x^\mu}} \left(Y^\nu \frac{\partial}{\partial x^\nu} \right) \\
&= X^\mu Y^\nu \nabla_{\frac{\partial}{\partial x^\mu}} \frac{\partial}{\partial x^\nu} + X^\mu \frac{\partial Y^\nu}{\partial x^\mu} \frac{\partial}{\partial x^\nu} \\
&= X^\mu Y^\nu \Gamma^\sigma_{\nu\mu} \frac{\partial}{\partial x^\sigma} + X^\mu \frac{\partial Y^\sigma}{\partial x^\mu} \frac{\partial}{\partial x^\sigma} \\
&= \left(X^\mu \frac{\partial Y^\sigma}{\partial x^\mu} + X^\mu Y^\nu \Gamma^\sigma_{\nu\mu} \right) \frac{\partial}{\partial x^\sigma}.
\end{aligned}$$

$$\therefore (\nabla_X Y)^\sigma = X^\mu \frac{\partial Y^\sigma}{\partial x^\mu} + X^\mu Y^\nu \Gamma^\sigma_{\nu\mu}. \tag{5.8}$$

Furthermore,

$$(\nabla_X Y)^\sigma = \left(\frac{\partial Y^\sigma}{\partial x^\mu} + \Gamma^\sigma_{\nu\mu} Y^\nu \right) X^\mu \implies Y^\sigma_{;\mu} = (\nabla Y)^\sigma{}_\mu = \frac{\partial Y^\sigma}{\partial x^\mu} + \Gamma^\sigma_{\nu\mu} Y^\nu. \tag{5.9}$$

In other words,

$$Y^\sigma_{;\mu} = Y^\sigma_{,\mu} + \Gamma^\sigma_{\nu\mu} Y^\nu. \tag{5.10}$$

Exercise 5.1. Show that under a change of coordinates, $\Gamma^\sigma_{\nu\mu}$ does not transform as a tensor field, but $Y^\sigma_{;\mu}$ transforms as a tensor field.

Now, along the same lines, ∇ can be defined on tensor fields by Leibniz property. If T is a tensor field of type (r, s), then ∇T is a tensor field of type $(r, s+1)$. For instance, if η is a 1-form, then for any vector fields X and Y, we define

$$(\nabla_X \eta)(Y) := \nabla_X (\eta(Y)) - \eta(\nabla_X Y). \tag{5.11}$$

$\nabla_X \eta$ is another 1-form. It acting on Y gives out a smooth function. It is not obvious that ∇Y is indeed a $(0, 2)$ tensor field. Let's have a closer look.

$$\begin{aligned}
(\nabla_X \eta)(Y) &= \nabla_X (\eta(Y)) - \eta(\nabla_X Y) = \nabla_X (\eta_\mu Y^\mu) - \eta_\mu (\nabla_X Y)^\mu \\
&= X(\eta_\mu) Y^\mu + \eta_\mu X(Y^\mu) - \eta_\mu \left[X^\nu \frac{\partial Y^\mu}{\partial x^\nu} + \Gamma^\mu_{\rho\nu} Y^\rho X^\nu \right] \\
&= X(\eta_\mu) Y^\mu - \Gamma^\mu_{\rho\nu} \eta_\mu Y^\rho X^\nu \\
&= \left(X^\nu \frac{\partial \eta_\mu}{\partial x^\nu} - \Gamma^\rho_{\mu\nu} \eta_\rho X^\nu \right) Y^\mu.
\end{aligned}$$

Therefore,

$$\begin{aligned}
(\nabla_X \eta)_\mu &= X^\nu \frac{\partial \eta_\mu}{\partial x^\nu} - \Gamma^\rho_{\mu\nu} \eta_\rho X^\nu \qquad (5.12) \\
&= \left(\frac{\partial \eta_\mu}{\partial x^\nu} - \Gamma^\rho_{\mu\nu} \eta_\rho \right) X^\nu.
\end{aligned}$$

$$\therefore (\nabla \eta)_{\mu\nu} = \eta_{\mu;\nu} = \eta_{\mu,\nu} - \Gamma^\rho_{\mu\nu} \eta_\rho. \tag{5.13}$$

For an (r,s) tensor field T, the Leibniz rule for covariant derivative is

$$\begin{aligned}\nabla_X\left(T\left(\eta_1,\cdots,\eta_r,Y_1,\cdots,Y_s\right)\right)=&\left(\nabla_XT\right)\left(\eta_1,\cdots,\eta_r,Y_1,\cdots,Y_s\right)\\&+T\left(\nabla_X\eta_1,\cdots,\eta^r,Y_1,\cdots,Y_s\right)+\cdots+T\left(\eta_1,\cdots,\nabla_X\eta_r,Y_1,\cdots,Y_s\right)\\&+T\left(\eta_1,\cdots,\eta_r,\nabla_XY_1,\cdots,Y_s\right)+\cdots+T\left(\eta_1,\cdots,\eta_r,Y_1,\cdots,\nabla_XY_s\right),\end{aligned}\tag{5.14}$$

where η_i are 1-forms and Y_j are vector fields. 5.14 allows us to define ∇ on tensor fields. Therefore, $(\nabla_XT)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s}$ is given by

$$\begin{aligned}(\nabla_XT)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s}=&\left(\nabla_XT\right)\left(\mathrm{d}x^{\mu_1},\cdots,\mathrm{d}x^{\mu_r},\frac{\partial}{\partial x^{\nu_1}},\cdots,\frac{\partial}{\partial x^{\nu_s}}\right)\\=&\nabla_X\left(T\left(\mathrm{d}x^{\mu_1},\cdots,\mathrm{d}x^{\mu_r},\frac{\partial}{\partial x^{\nu_1}},\cdots,\frac{\partial}{\partial x^{\nu_s}}\right)\right)\\&-\sum_{i=1}^{r}T\left(\mathrm{d}x^{\mu_1},\cdots,\mathrm{d}x^{\mu_{i-1}},\nabla_X\mathrm{d}x^{\mu_i},\mathrm{d}x^{\mu_{i+1}},\cdots,\mathrm{d}x^{\mu_r},\frac{\partial}{\partial x^{\nu_1}},\cdots,\frac{\partial}{\partial x^{\nu_s}}\right)\\&-\sum_{j=1}^{s}T\left(\mathrm{d}x^{\mu_1},\cdots,\mathrm{d}x^{\mu_r},\frac{\partial}{\partial x^{\nu_1}},\cdots,\frac{\partial}{\partial x^{\nu_{j-1}}},\nabla_X\frac{\partial}{\partial x^{\nu_j}},\frac{\partial}{\partial x^{\nu_{j+1}}},\cdots,\frac{\partial}{\partial x^{\nu_s}}\right).\end{aligned}\tag{5.15}$$

Now,

$$\begin{aligned}&\nabla_X\left(T\left(\mathrm{d}x^{\mu_1},\cdots,\mathrm{d}x^{\mu_r},\frac{\partial}{\partial x^{\nu_1}},\cdots,\frac{\partial}{\partial x^{\nu_s}}\right)\right)\\&=\nabla_{X^\rho\frac{\partial}{\partial x^\rho}}\left(T^{\mu_1\cdots\mu_r}_{\nu_1\cdots\nu_s}\right)=X^\rho\frac{\partial}{\partial x^\rho}\left(T^{\mu_1\cdots\mu_r}_{\nu_1\cdots\nu_s}\right)\\&=T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s,\rho}\,X^\rho.\end{aligned}\tag{5.16}$$

We have seen earlier that

$$\nabla_X\eta=\left(\nabla_X\eta\right)_\mu\mathrm{d}x^\mu=X\left(\eta_\mu\right)\mathrm{d}x^\mu-\Gamma^\tau_{\mu\rho}X^\rho\eta_\tau\mathrm{d}x^\mu.\tag{5.17}$$

If we take $\eta=\mathrm{d}x^{\mu_i}$, $\eta_\mu=0$ unless $\mu=\mu_i$. Plugging it into 5.17, we get

$$\nabla_X\mathrm{d}x^{\mu_i}=X\left(\eta_\mu\right)\;\mathrm{d}x^\mu-\Gamma^\tau_{\mu\rho}X^\rho\eta_\tau\;\mathrm{d}x^\mu=-\Gamma^\tau_{\mu\rho}X^\rho\delta^{\mu_i}_\tau\;\mathrm{d}x^\mu=-\Gamma^{\mu_i}_{\sigma\rho}X^\rho\;\mathrm{d}x^\sigma.\tag{5.18}$$

Therefore,

$$\begin{aligned}&T\left(\mathrm{d}x^{\mu_1},\cdots,\mathrm{d}x^{\mu_{i-1}},\nabla_X\mathrm{d}x^{\mu_i},\mathrm{d}x^{\mu_{i+1}},\cdots,\mathrm{d}x^{\mu_r},\frac{\partial}{\partial x^{\nu_1}},\cdots,\frac{\partial}{\partial x^{\nu_s}}\right)\\&=T\left(\mathrm{d}x^{\mu_1},\cdots,\mathrm{d}x^{\mu_{i-1}},-\Gamma^{\mu_i}_{\sigma\rho}X^\rho\;\mathrm{d}x^\sigma,\mathrm{d}x^{\mu_{i+1}},\cdots,\mathrm{d}x^{\mu_r},\frac{\partial}{\partial x^{\nu_1}},\cdots,\frac{\partial}{\partial x^{\nu_s}}\right)\\&=-\Gamma^{\mu_i}_{\sigma\rho}X^\rho\;T\left(\mathrm{d}x^{\mu_1},\cdots,\mathrm{d}x^{\mu_{i-1}},\mathrm{d}x^\sigma,\mathrm{d}x^{\mu_{i+1}},\cdots,\mathrm{d}x^{\mu_r},\frac{\partial}{\partial x^{\nu_1}},\cdots,\frac{\partial}{\partial x^{\nu_s}}\right)\\&=-\Gamma^{\mu_i}_{\sigma\rho}T^{\mu_1\cdots\mu_{i-1}\sigma\mu_{i+1}\cdots\mu_r}{}_{\nu_1\cdots\nu_s}X^\rho.\end{aligned}\tag{5.19}$$

Moreover, we have

$$\nabla_X\frac{\partial}{\partial x^{\nu_j}}=\nabla_{X^\rho\frac{\partial}{\partial x^\rho}}\frac{\partial}{\partial x^{\nu_j}}=\Gamma^\sigma_{\nu_j\rho}X^\rho\frac{\partial}{\partial x^\sigma}.\tag{5.20}$$

Therefore,

$$\begin{aligned}&T\left(\mathrm{d}x^{\mu_1},\cdots,\mathrm{d}x^{\mu_r},\frac{\partial}{\partial x^{\nu_1}},\cdots,\frac{\partial}{\partial x^{\nu_{j-1}}},\nabla_X\frac{\partial}{\partial x^{\nu_j}},\frac{\partial}{\partial x^{\nu_{j+1}}},\cdots,\frac{\partial}{\partial x^{\nu_s}}\right)\\&=T\left(\mathrm{d}x^{\mu_1},\cdots,\mathrm{d}x^{\mu_r},\frac{\partial}{\partial x^{\nu_1}},\cdots,\frac{\partial}{\partial x^{\nu_{j-1}}},\Gamma^\sigma_{\nu_j\rho}X^\rho\frac{\partial}{\partial x^\sigma},\frac{\partial}{\partial x^{\nu_{j+1}}},\cdots,\frac{\partial}{\partial x^{\nu_s}}\right)\\&=\Gamma^\sigma_{\nu_j\rho}X^\rho\;T\left(\mathrm{d}x^{\mu_1},\cdots,\mathrm{d}x^{\mu_r},\frac{\partial}{\partial x^{\nu_1}},\cdots,\frac{\partial}{\partial x^{\nu_{j-1}}},\frac{\partial}{\partial x^\sigma},\frac{\partial}{\partial x^{\nu_{j+1}}},\cdots,\frac{\partial}{\partial x^{\nu_s}}\right)\\&=\Gamma^\sigma_{\nu_j\rho}T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{j-1}\sigma\nu_{j+1}\cdots\nu_s}X^\rho.\end{aligned}\tag{5.21}$$

Plugging 5.16, 5.19, 6.43 into 5.15, we obtain

$$(\nabla_X T)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} = T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s,\rho}\, X^\rho + \sum_{i=1}^{r} \Gamma^{\mu_i}_{\sigma\rho} T^{\mu_1\cdots\mu_{i-1}\sigma\mu_{i+1}\cdots\mu_r}{}_{\nu_1\cdots\nu_s}\, X^\rho - \sum_{j=1}^{s} \Gamma^{\sigma}_{\nu_j\rho} T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{j-1}\sigma\nu_{j+1}\cdots\nu_s}\, X^\rho.$$

$$\therefore T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s;\rho} = T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s,\rho} + \sum_{i=1}^{r} \Gamma^{\mu_i}_{\sigma\rho} T^{\mu_1\cdots\mu_{i-1}\sigma\mu_{i+1}\cdots\mu_r}{}_{\nu_1\cdots\nu_s} - \sum_{j=1}^{s} \Gamma^{\sigma}_{\nu_j\rho} T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{j-1}\sigma\nu_{j+1}\cdots\nu_s}. \tag{5.22}$$

§5.2 Torsion Freeness

Recall that $(\nabla_X Y)^\mu = Y^\mu_{;\nu} X^\nu$.[1] In a coordinate basis, the components of $(\nabla_X Y - \nabla_Y X)$ are

$$\begin{aligned} X^\nu Y^\mu_{;\nu} - Y^\nu X^\mu_{;\nu} &= X^\nu \left(Y^\mu_{,\nu} + \Gamma^\mu_{\rho\nu} Y^\rho\right) - Y^\nu \left(X^\mu_{,\nu} + \Gamma^\mu_{\nu\rho} X^\rho\right) \\ &= X^\nu Y^\mu_{,\nu} - Y^\nu X^\mu_{,\nu} + \Gamma^\mu_{\rho\nu} X^\nu Y^\rho - \Gamma^\mu_{\nu\rho} X^\rho Y^\nu \\ &= [X,Y]^\mu + \left(\Gamma^\mu_{\rho\nu} - \Gamma^\mu_{\nu\rho}\right) X^\nu Y^\rho \\ &= [X,Y]^\mu + 2\Gamma^\mu_{[\rho\nu]} X^\nu Y^\rho. \end{aligned} \tag{5.23}$$

Now, consider 5.13 again.

$$(\nabla\eta)_{\mu\nu} \equiv \nabla_\nu \eta_\mu = \eta_{\mu;\nu} = \eta_{\mu,\nu} - \Gamma^\rho_{\mu\nu}\eta_\rho$$

Take $\nabla = \mathrm{d}f$ here. Then the LHS is

$$\nabla_\nu (\mathrm{d}f)_\mu = \nabla_\nu \left(\frac{\partial f}{\partial x^\mu}\right) = \nabla_\nu \nabla_\mu f = f_{\ ;\mu\nu}. \tag{5.24}$$

On the other hand, the first term on the RHS is

$$\eta_{\mu,\nu} = \frac{\partial}{\partial x^\nu}\left((\mathrm{d}f)_\mu\right) = \frac{\partial}{\partial x^\nu}\left(\frac{\partial f}{\partial x^\mu}\right) = f_{\ ,\mu\nu}. \tag{5.25}$$

Therefore, we have

$$f_{\ ;\mu\nu} = f_{\ ,\mu\nu} - \Gamma^\rho_{\mu\nu} f_{\ ,\rho}. \tag{5.26}$$

Although $f_{\ ,\mu\nu} = f_{\ ,\nu\mu}$ holds, $f_{\ ;\mu\nu} \neq f_{\ ;\nu\mu}$ in general. In fact, from 5.26, we get

$$f_{;[\mu\nu]} = -\Gamma^\rho_{[\mu\nu]} f_{,\rho}. \tag{5.27}$$

5.27 guarantees that $f_{\ ;\mu\nu} = f_{\ ;\nu\mu}$ if and only if $\Gamma^\rho_{\mu\nu} = \Gamma^\rho_{\nu\mu}$. This condition is called the torsion freeness of the connection.

Definition 5.3. A connection ∇ is torsion-free if $\nabla_b\nabla_a f = \nabla_a \nabla_b f$ for any function f. From 5.27, it is equivalent to

$$\Gamma^\rho_{\mu\nu} = \Gamma^\rho_{\nu\mu} \tag{5.28}$$

in a coordinate basis.

Lemma 5.1

For a torsion free connection ∇, given two vector fields X and Y, $\nabla_X Y - \nabla_Y X = [X, Y]$.

Proof. From 5.23, we have

$$(\nabla_X Y - \nabla_Y X)^\mu - [X,Y]^\mu = 2\Gamma^\mu_{[\rho\nu]} X^\nu Y^\rho, \tag{5.29}$$

which is 0 if ∇ is torsion free. Therefore, $\nabla_X Y - \nabla_Y X = [X, Y]$. ■

Remark 5.3. We've just seen that if the connection is torsion free, the second covariant derivatives of a scalar function commute. However, the second covariant derivatives of a tensor field, in general, do not commute even though the coonnection is torsion free.

[1] Note that these formulations don't require a metric on the manifold.

§5.3 The Levi-Civita Connection

Theorem 5.2 (Fundamental Theorem of Riemannian Geometry)
Let M be a manifold with a metric g. There exists a unique torsion-free connection ∇ such that the metric is covariantly constant, i.e. $\nabla g = 0$. This connection is called the Levi-Civita connection.

Proof. Assume such a connection exists. Let X, Y, Z be vector fields. Then we have

$$\begin{aligned} X\left(g\left(Y, Z\right)\right) = \nabla_X\left(g\left(Y, Z\right)\right) &= \left(\nabla_X g\right)\left(Y, Z\right) + g\left(\nabla_X Y, Z\right) + g\left(Y, \nabla_X Z\right) \\ &= g\left(\nabla_X Y, Z\right) + g\left(Y, \nabla_X Z\right), \end{aligned} \tag{5.30}$$

because $\nabla g = 0$. Similarly,

$$Z\left(g\left(X, Y\right)\right) = g\left(\nabla_Z X, Y\right) + g\left(X, \nabla_Z Y\right), \tag{5.31}$$
$$Y\left(g\left(Z, X\right)\right) = g\left(\nabla_Y Z, X\right) + g\left(Z, \nabla_Y X\right). \tag{5.32}$$

5.30−5.31+5.32 gives us

$$\begin{aligned} &X\left(g\left(Y, Z\right)\right) + Y\left(g\left(Z, X\right)\right) - Z\left(g\left(X, Y\right)\right) \\ &= g\left(\nabla_X Y, Z\right) + g\left(Y, \nabla_X Z\right) + g\left(\nabla_Y Z, X\right) + g\left(Z, \nabla_Y X\right) - g\left(\nabla_Z X, Y\right) - g\left(X, \nabla_Z Y\right) \\ &= g\left(\nabla_X Y + \nabla_Y X, Z\right) + g\left(\nabla_X Z - \nabla_Z X, Y\right) + g\left(\nabla_Y Z - \nabla_Z Y\right). \end{aligned} \tag{5.33}$$

Torsion-free condition implies $\nabla_X Y - \nabla_Y X = [X, Y]$. Therefore,

$$\begin{aligned} &X\left(g\left(Y, Z\right)\right) + Y\left(g\left(Z, X\right)\right) - Z\left(g\left(X, Y\right)\right) \\ &= g\left(2\nabla_X Y - \left[X, Y\right], Z\right) + g\left(\left[X, Z\right], Y\right) + g\left(\left[Y, Z\right], X\right) \\ &= 2g\left(\nabla_X Y, Z\right) - g\left(\left[X, Y\right], Z\right) - g\left(\left[Z, X\right], Y\right) + g\left(\left[Y, Z\right], X\right). \end{aligned} \tag{5.34}$$

$$\begin{aligned} \therefore g\left(\nabla_X Y, Z\right) = \frac{1}{2}\Big[&X\left(g\left(Y, Z\right)\right) + Y\left(g\left(Z, X\right)\right) - Z\left(g\left(X, Y\right)\right) \\ &+ g\left(\left[X, Y\right], Z\right) + g\left(\left[Z, X\right], Y\right) - g\left(\left[Y, Z\right], X\right)\Big]. \end{aligned} \tag{5.35}$$

5.35 is known as the Koszul formula. If there is another such connection $\widetilde{\nabla}$, then it also satisfies 5.35. Therefore,

$$g\left(\nabla_X Y, Z\right) = g\left(\widetilde{\nabla}_X Y, Z\right) \implies g\left(\nabla_X Y - \widetilde{\nabla}_X Y, Z\right) = 0, \tag{5.36}$$

for every $Z \in \mathfrak{X}(M)$. Therefore, by the non-degeneracy of the metric, $\nabla_X Y = \widetilde{\nabla}_X Y$, so the connection is unique (if it exists). Now we need to verify that such a connection exists. For that purpose, it suffices to show that any $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ satisfying 5.35 is indeed a connection. First, we shall show that $\nabla_{fX} Y = f \nabla_X Y$ for a smooth function f.

$$\begin{aligned} 2g\left(\nabla_{fX} Y, Z\right) &= fX(g(Y,Z)) + Y(g(Z, fX)) - Z(g(fX, Y)) \\ &\qquad + g([fX, Y], Z) + g([Z, fX], Y) - g([Y, Z], fX) \\ &= fX(g(Y,Z)) + Y(f\,g(Z, X)) - Z(f\,g(X, Y)) \\ &\qquad + g([fX, Y], Z) + g([Z, fX], Y) - g([Y, Z], fX) \\ &= fX(g(Y,Z)) + f\,Y(g(Z,X)) + Y(f)\,g(Z,X) - f\,Z(g(X,Y)) - Z(f)\,g(X,Y) \\ &\qquad + g([fX, Y], Z) + g([Z, fX], Y) - f\,g([Y, Z], X) \end{aligned} \tag{5.37}$$

$[fX, Y] = f\,[X, Y] - Y\,(f)\,X$, and $[Z, fX] = f\,[Z, X] + Z\,(f)\,X$. Therefore,

$$\begin{aligned}2g\left(\nabla_{fX}Y, Z\right) &= fX(g(Y,Z)) + f\,Y(g(Z,X)) + Y\,(f)\,g(Z,X) - f\,Z(g(X,Y)) - Z\,(f)\,g(X,Y)\\ &\quad + g(f\,[X,Y] - Y\,(f)\,X, Z) + g(f\,[Z,X] + Z\,(f)\,X, Y) - f\,g([Y,Z],X)\\ &= fX(g(Y,Z)) + f\,Y(g(Z,X)) + Y\,(f)\,g(Z,X) - f\,Z(g(X,Y)) - Z\,(f)\,g(X,Y)\\ &\quad + fg([X,Y]\,,Z) - Y\,(f)\,g(X,Z) + fg([Z,X]\,,Y) + Z\,(f)\,g(X,Y) - f\,g([Y,Z],X)\\ &= fX(g(Y,Z)) + f\,Y(g(Z,X)) - f\,Z(g(X,Y))\\ &\quad + fg([X,Y]\,,Z) + fg([Z,X]\,,Y) - f\,g([Y,Z],X)\\ &= 2f\,g\left(\nabla_X Y, Z\right) = 2g\left(f\nabla_X Y, Z\right).\end{aligned} \tag{5.38}$$

So, $g\left(\nabla_{fX}Y - f\nabla_X Y, Z\right) = 0$ for every $Z \in \mathfrak{X}(M)$. Hence, by the non-degeneracy of g, $\nabla_{fX}Y = f\nabla_X Y$.

$$\begin{aligned}2g\left(\nabla_{X_1+X_2}Y, Z\right) &= (X_1 + X_2)\,(g(Y,Z)) + Y(g(Z, X_1 + X_2)) - Z(g(X_1 + X_2, Y))\\ &\quad + g([X_1 + X_2, Y], Z) + g([Z, X_1 + X_2], Y) - g([Y,Z], X_1 + X_2)\\ &= X_1 g(Y,Z) + X_2 g(Y,Z) + Y(g(Z,X_1)) + Y(g(Z,X_2))\\ &\quad - Z(g(X_1,Y)) - Z(g(X_2,Y)) + g([X_1,Y],Z) + g([X_2,Y],Z)\\ &\quad + g([Z,X_1],Y) + g([Z,X_2],Y) - g([Y,Z],X_2) - g([Y,Z],X_2)\\ &= 2g\left(\nabla_{X_1}Y, Z\right) + 2g\left(\nabla_{X_2}Y, Z\right)\\ &= 2g\left(\nabla_{X_1}Y + \nabla_{X_2}Y, Z\right).\end{aligned} \tag{5.39}$$

So, $g\left(\nabla_{X_1+X_2}Y - \nabla_{X_1}Y - \nabla_{X_2}Y, Z\right) = 0$ for every $Z \in \mathfrak{X}(M)$. Hence, by the non-degeneracy of g, $\nabla_{X_1+X_2}Y = \nabla_{X_1}Y + \nabla_{X_2}Y$. By combining this with $\nabla_{fX}Y = f\nabla_X Y$, we get

$$\nabla_{f_1X_1+f_2X_2}Y = f_1\nabla_{X_1}Y + f_2\nabla_{X_2}Y. \tag{5.40}$$

Now, we shall show that $\nabla_X\left(Y_1 + Y_2\right) = \nabla_X Y_1 + \nabla_X Y_2$.

$$\begin{aligned}2g\left(\nabla_X\left(Y_1 + Y_2\right), Z\right) &= X(g(Y_1 + Y_2, Z)) + (Y_1 + Y_2)\,(g(Z,X)) - Z(g(X, Y_1 + Y_2))\\ &\quad + g([X, Y_1 + Y_2], Z) + g([Z,X], Y_1 + Y_2) - g([Y_1 + Y_2, Z], X)\\ &= X(g(Y_1,Z)) + X(g(Y_2,Z)) + Y_1(g(Z,X)) + Y_2\,(g(Z,X))\\ &\quad - Z(g(X,Y_1)) - Z(g(X,Y_2)) + g([X,Y_1],Z) + g([X,Y_2],Z)\\ &\quad + g([Z,X],Y_1) + g([Z,X],Y_2) - g([Y_1,Z],X) - g([Y_2,Z],X)\\ &= 2g\left(\nabla_X Y_1, Z\right) + 2g\left(\nabla_X Y_2, Z\right)\\ &= 2g\left(\nabla_X Y_1 + \nabla_X Y_2, Z\right).\end{aligned} \tag{5.41}$$

So, $g\left(\nabla_X\left(Y_1 + Y_2\right) - \nabla_X Y_1 - \nabla_X Y_2, Z\right) = 0$ for every $Z \in \mathfrak{X}(M)$. Hence, by the non-degeneracy of g, $\nabla_X\left(Y_1 + Y_2\right) = \nabla_X Y_1 + \nabla_X Y_2$. We are only left to show that $\nabla_X\left(fY\right) = f\nabla_X Y + X\,(f)\,Y$.

$$\begin{aligned}2g\left(\nabla_X\left(fY\right), Z\right) &= X(g(fY,Z)) + fY(g(Z,X)) - Z(g(X,fY))\\ &\quad + g([X,fY],Z) + g([Z,X],fY) - g([fY,Z],X)\\ &= X(f\,g(Y,Z)) + fY(g(Z,X)) - Z(f\,g(X,Y))\\ &\quad + g(f[X,Y] + X\,(f)\,Y, Z) + f\,g([Z,X],Y) - g(f[Y,Z] - Z\,(f)\,Y, X)\\ &= f\,X(g(Y,Z)) + X\,(f)\,g(Y,Z) + fY(g(Z,X)) - f\,Z(g(X,Y)) - Z\,(f)\,g(X,Y)\\ &\quad + f\,g([X,Y],Z) + X\,(f)\,g(Y,Z) + f\,g([Z,X],Y) - f\,g([Y,Z],X) + Z\,(f)\,g(Y,X)\\ &= 2f\,g\left(\nabla_X Y, Z\right) + 2X\,(f)\,g(Y,Z)\\ &= 2g\left(f\nabla_X Y + X\,(f)\,Y, Z\right).\end{aligned} \tag{5.42}$$

So, $g\left(\nabla_X\left(fY\right) - f\nabla_X Y - X\,(f)\,Y, Z\right) = 0$ for every $Z \in \mathfrak{X}(M)$. Hence, by the non-degeneracy of g, $\nabla_X\left(fY\right) - f\nabla_X Y = X\,(f)\,Y$. Therefore, ∇ is indeed a connection. ■

Let us now compute ∇ in a coordinate basis. From 5.35, we get

$$g\left(\nabla_{\frac{\partial}{\partial x^\rho}}\frac{\partial}{\partial x^\nu},\frac{\partial}{\partial x^\sigma}\right)=\frac{1}{2}\left[\frac{\partial}{\partial x^\rho}\left(g\left(\frac{\partial}{\partial x^\nu},\frac{\partial}{\partial x^\sigma}\right)\right)+\frac{\partial}{\partial x^\nu}\left(g\left(\frac{\partial}{\partial x^\sigma},\frac{\partial}{\partial x^\rho}\right)\right)-\frac{\partial}{\partial x^\sigma}\left(g\left(\frac{\partial}{\partial x^\rho},\frac{\partial}{\partial x^\nu}\right)\right)\right]$$

$$\therefore g\left(\Gamma^\tau_{\nu\rho}\frac{\partial}{\partial x^\tau},\frac{\partial}{\partial x^\sigma}\right)=\frac{1}{2}\left(g_{\nu\sigma,\rho}+g_{\sigma\rho,\nu}-g_{\rho\nu,\sigma}\right). \tag{5.43}$$

The LHS is nothing but $\Gamma^\tau_{\nu\rho}g_{\tau\sigma}$. So, contracting with $g^{\mu\sigma}$, we get

$$g^{\mu\sigma}\Gamma^\tau_{\nu\rho}g_{\tau\sigma}=\frac{1}{2}g^{\mu\sigma}\left(g_{\nu\sigma,\rho}+g_{\sigma\rho,\nu}-g_{\rho\nu,\sigma}\right). \tag{5.44}$$

The LHS is $\delta^\mu_\tau\Gamma^\tau_{\nu\rho}=\Gamma^\mu_{\nu\rho}$. Therefore,

$$\Gamma^\mu_{\nu\rho}=\frac{1}{2}g^{\mu\sigma}\left(g_{\nu\sigma,\rho}+g_{\sigma\rho,\nu}-g_{\rho\nu,\sigma}\right). \tag{5.45}$$

This expression is exactly identical to the expression of Christoffel symbols we have seen earlier. Therefore, the Christoffel symbols are the components of the Levi-Civita connection.

§5.4 Geodesic and Parallel Transport

Previously we considered curves that extremize proper time between points of spacetime, and showed that this gives the equation

$$\frac{\mathrm{d}^2x^\mu}{\mathrm{d}\tau^2}+\Gamma^\mu_{\nu\rho}\left(x\left(\tau\right)\right)\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}=0, \tag{5.46}$$

where τ is the proper time along the curve. The tangent vector to the curve has components $X^\mu=\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}$. This is defined along the curve. We extend X^μ in the neighborhood of the curve so that X^μ becomes a vector field and the curve is an integral curve of this vector field.

$$\frac{\mathrm{d}^2x^\mu}{\mathrm{d}\tau^2}=\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}\right)=\frac{\mathrm{d}X^\mu}{\mathrm{d}\tau}\left(x\left(\tau\right)\right)=\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\frac{\partial X^\mu}{\partial x^\nu}=X^\nu X^\mu_{,\nu}. \tag{5.47}$$

Then 5.46 becomes

$$\begin{aligned}&X^\nu X^\mu_{,\nu}+\Gamma^\mu_{\nu\rho}X^\nu X^\rho=0\\ \implies &X^\nu\left(X^\mu_{,\nu}+\Gamma^\mu_{\nu\rho}X^\rho\right)=0\\ \implies &X^\nu X^\mu_{;\nu}=0\implies\nabla_XX=0.\end{aligned} \tag{5.48}$$

Here ∇ is the Levi-Civita connection.

Remark 5.4. (On parallel transport) Let X^μ be the tangent vector to a curve $\lambda\left(t\right)$. A tensor field T is parallely transported along the curve if $\nabla_XT=0$. In particular, a vector field Y is parallely transported along a curve $\lambda\left(t\right)$ (which is the integral curve to a vector field X) if $\nabla_XY=0$. When a curve $\lambda\left(t\right)$ is a geodesic, the tangent vector field to $\lambda\left(t\right)$ is parallely transported along $\lambda\left(t\right)$, i.e. $\nabla_XX=0$.

> **Definition 5.4** (Affine parametrized geodesic)**.** Let M be a manifold with a connection ∇. An **affinely paramtrized geodesic** is an interal curve of a vector field X satisfying $\nabla_XX=0$.

Remark 5.5. Consider a curve λ with parameter t whose tangent vector field satisfies $\nabla_XX=0$. Let u be some other parameter so that $t=t(u)$ and $\frac{\mathrm{d}t}{\mathrm{d}u}>0$. Then the tangent vector becomes

$$Y^\mu=\frac{\mathrm{d}x^\mu}{\mathrm{d}u}=\frac{\mathrm{d}t}{\mathrm{d}u}\frac{\mathrm{d}x^\mu}{\mathrm{d}t}=hX^\mu, \tag{5.49}$$

where $h = \frac{\mathrm{d}t}{\mathrm{d}u}$. So $Y = hX$. Now,

$$\nabla_Y Y = \nabla_{hX} (hX) = h \nabla_X (hX) = X(h)\, hX + h^2 \nabla_X X = X(h)\, Y. \tag{5.50}$$

$X = X^\mu \frac{\partial}{\partial x^\mu}$, so

$$X(h) = X^\mu \frac{\partial h}{\partial x^\mu} = \frac{\mathrm{d}x^\mu}{\mathrm{d}t} \frac{\partial h}{\partial x^\mu} = \frac{\mathrm{d}h}{\mathrm{d}t}. \tag{5.51}$$

The new parameter is affine, i.e., $\nabla_Y Y = 0$, if $X(h) = 0$. Then h is a constant, so u and t are related by $u = at + b$ where a and b are constants and $a > 0$.

6 Curvature

§6.1 Riemann Curvature Tensor

Definition 6.1. The **Riemann curvature tensor** $R^a{}_{bcd}$ of a connection ∇ is defined by

$$R^a{}_{bcd} Z^b X^c Y^d = (R(X,Y)Z)^a , \tag{6.1}$$

where X, Y, Z are vector fields, and $R(X,Y)Z$ is the vector field

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z. \tag{6.2}$$

Recall that tensor fields are multilinear maps. Therefore,

$$R : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M) \tag{6.3}$$

is linear in all 3 vector fields X, Y, Z. Let us verify this.

1. We shall show that $R(fX,Y)Z = fR(X,Y)Z$.

$$\begin{aligned} R(fX,Y)Z &= \nabla_{fX} \nabla_Y Z - \nabla_Y \nabla_{fX} Z - \nabla_{[fX,Y]} Z \\ &= f \nabla_X \nabla_Y Z - \nabla_Y (f \nabla_X Z) - \nabla_{f[X,Y] - Y(f)X} Z \\ &= f \nabla_X \nabla_Y Z - f \nabla_Y \nabla_X Z - Y(f) \nabla_X Z - f \nabla_{[X,Y]} Z + Y(f) \nabla_X Z \\ &= f \left(\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \right) \\ &= fR(X,Y)Z. \end{aligned} \tag{6.4}$$

2. From the definition (6.2) of the Riemann curvature tensor, one immediately finds that $R(X,Y)Z = -R(Y,X)Z$. Therefore,

$$R(X,fY)Z = -R(fY,X)Z = -fR(Y,X)Z = fR(X,Y)Z. \tag{6.5}$$

3. It remains to show that $R(X,Y)(fZ) = fR(X,Y)Z$.

$$\begin{aligned} R(X,Y)(fZ) &= \nabla_X \nabla_Y (fZ) - \nabla_Y \nabla_X (fZ) - \nabla_{[X,Y]} (fZ) \\ &= \nabla_X (f \nabla_Y Z + Y(f) Z) - \nabla_Y (f \nabla_X Z + X(f) Z) - f \nabla_{[X,Y]} Z - [X,Y](f) Z \\ &= f \nabla_X \nabla_Y Z + X(f) \nabla_Y Z + X(Y(f)) Z - f \nabla_Y \nabla_X Z - Y(f) \nabla_X Z \\ &\qquad - Y(X(f)) Z - f \nabla_{[X,Y]} Z - [X,Y](f) Z \\ &= f \left(\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \right) \\ &= fR(X,Y)Z. \end{aligned} \tag{6.6}$$

Since $R(X,Y)Z = -R(Y,X)Z$, we have

$$\begin{aligned} (R(X,Y)Z)^a = (-R(Y,X)Z)^a &\implies R^a{}_{bcd} Z^b X^c Y^d = -R^a{}_{bcd} Z^b Y^c X^d \\ \implies R^a{}_{bcd} Z^b X^c Y^d = -R^a{}_{bdc} Z^b X^c Y^d. \end{aligned} \tag{6.7}$$

Since this is true for all $X, Y, Z \in \mathfrak{X}(M)$,

$$R^a{}_{bcd} = -R^a{}_{bdc}. \tag{6.8}$$

In other words,

$$R^a{}_{b(cd)} = 0. \tag{6.9}$$

Now let us compute Riemann curvature tensor in a coordinate basis. Choosing $X = \frac{\partial}{\partial x^\rho}, Y = \frac{\partial}{\partial x^\sigma}, Z = \frac{\partial}{\partial x^\nu}$, we get

$$\begin{aligned}
R^\mu{}_{\nu\rho\sigma}\frac{\partial}{\partial x^\mu} &= R\left(X,Y\right)Z = R\left(\frac{\partial}{\partial x^\rho},\frac{\partial}{\partial x^\sigma}\right)\frac{\partial}{\partial x^\nu}\\
&= \nabla_{\frac{\partial}{\partial x^\rho}}\nabla_{\frac{\partial}{\partial x^\sigma}}\frac{\partial}{\partial x^\nu} - \nabla_{\frac{\partial}{\partial x^\sigma}}\nabla_{\frac{\partial}{\partial x^\rho}}\frac{\partial}{\partial x^\nu} - \nabla_{\left[\frac{\partial}{\partial x^\rho},\frac{\partial}{\partial x^\sigma}\right]}\frac{\partial}{\partial x^\nu}\\
&= \nabla_{\frac{\partial}{\partial x^\rho}}\left(\Gamma^\tau_{\nu\sigma}\frac{\partial}{\partial x^\tau}\right) - \nabla_{\frac{\partial}{\partial x^\sigma}}\left(\Gamma^\tau_{\nu\rho}\frac{\partial}{\partial x^\tau}\right)\\
&= \frac{\partial\Gamma^\tau_{\nu\sigma}}{\partial x^\rho}\frac{\partial}{\partial x^\tau} + \Gamma^\tau_{\nu\sigma}\Gamma^\mu_{\tau\rho}\frac{\partial}{\partial x^\mu} - \frac{\partial\Gamma^\tau_{\nu\rho}}{\partial x^\sigma}\frac{\partial}{\partial x^\tau} - \Gamma^\tau_{\nu\rho}\Gamma^\mu_{\tau\sigma}\frac{\partial}{\partial x^\mu}\\
&= \left(\frac{\partial\Gamma^\mu_{\nu\sigma}}{\partial x^\rho} + \Gamma^\tau_{\nu\sigma}\Gamma^\mu_{\tau\rho} - \frac{\partial\Gamma^\mu_{\nu\rho}}{\partial x^\sigma} - \Gamma^\tau_{\nu\rho}\Gamma^\mu_{\tau\sigma}\right)\frac{\partial}{\partial x^\mu}.
\end{aligned} \tag{6.10}$$

Therefore,

$$R^\mu{}_{\nu\rho\sigma} = \frac{\partial\Gamma^\mu_{\nu\sigma}}{\partial x^\rho} - \frac{\partial\Gamma^\mu_{\nu\rho}}{\partial x^\sigma} + \Gamma^\tau_{\nu\sigma}\Gamma^\mu_{\tau\rho} - \Gamma^\tau_{\nu\rho}\Gamma^\mu_{\tau\sigma}. \tag{6.11}$$

§6.2 Normal Coordinates

Theorem 6.1

Let M be a manifold with a connection ∇. Let $p \in M$ and $X_p \in T_pM$. Then there exists a unique affinely parametrized geodesic through p and tangent vector X_p at p.

Proof. We choose a coordinate chart x^μ in a neighborhood of p. Let X_p^μ be the components of X_p in this coordinate basis. Consider a curve λ parametrized by τ. Then the components of the tangent vector to the curve are

$$X^\mu = \frac{\mathrm{d}\left(x^\mu\circ\lambda\right)}{\mathrm{d}\tau}. \tag{6.12}$$

The geodesic equation is

$$\frac{\mathrm{d}^2\left(x^\mu\circ\lambda\right)}{\mathrm{d}\tau^2} + \Gamma^\mu_{\nu\rho}\left(x\left(\lambda\left(\tau\right)\right)\right)\frac{\mathrm{d}\left(x^\nu\circ\lambda\right)}{\mathrm{d}\tau}\frac{\mathrm{d}\left(x^\rho\circ\lambda\right)}{\mathrm{d}\tau} = 0. \tag{6.13}$$

There are n differential equations, each with the initial condition

$$\left(x^\mu\circ\lambda\right)(0) = x^\mu\left(p\right),\quad \left.\frac{\mathrm{d}\left(x^\mu\circ\lambda\right)}{\mathrm{d}\tau}\right|_{\tau=0} = X_p^\mu. \tag{6.14}$$

Then existence and uniqueness of $x^\mu\circ\lambda$ is guaranteed by the theory of ordinary differential equations. ■

Definition 6.2 (Exponential Map)**.** Let M be a manifold with a connection ∇. Let $p \in M$. The exponential map $\exp : T_pM \to M$ is defined as the map which sends X_p to the point unit affine parameter distance along the geodesic through p with tangent X_p at p. In other words, if λ is the unique geodesic such that $\lambda\left(0\right) = p$ and $\lambda'\left(0\right) = X_p$ (existence and uniqueness of λ is guaranteed by Theorem 6.1.), then $exp\left(X_p\right) = \lambda\left(1\right)$.

It can be shown that exp is one-to-one and onto locally, i.e. for X_p in a neighborhood of the origin of T_pM.

Lemma 6.2

If $\exp X_p = \lambda\left(1\right)$ (as in Definition 6.2), then $\exp\left(tX_p\right) = \lambda\left(t\right)$ for $0 \le t \le 1$.

Proof. Clearly, it is true for $t = 0$. If λ_t is the unique geodesic through p that has tangent vector tX_p at p, then let $\lambda(\tau) = \lambda_t\left(\frac{1}{t}\tau\right)$. Then $\lambda(0) = p$, and

$$\left.\frac{\mathrm{d}\left(x^\mu \circ \lambda\right)}{\mathrm{d}\tau}\right|_{t=0} = \frac{1}{t}\left.\frac{\mathrm{d}\left(x^\mu \circ \lambda_t\right)}{\mathrm{d}\tau}\right|_{t=0} = \frac{1}{t} t X_p^\mu = X_p^\mu. \tag{6.15}$$

λ clearly satisfies the geodesic equation. Therefore, λ is the unique geodesic through p with tangent vector X_p at p. Therefore,

$$\exp\left(tX_p\right) = \lambda_t(1) = \lambda(t). \tag{6.16}$$

■

Definition 6.3 (Normal Coordinates)**.** Let $\{e_\mu\}$ be a basis for T_pM. **Normal coordinates** at p are defined in a neighborhood of p as follows: pick q near p. Suppose $q = \exp\left(X_p\right)$. Then the coordinates of q are X_p^μ.

Lemma 6.3

In normal coordinates at p, $\Gamma^\mu_{(\nu\rho)}(p) = 0$. For a torsion-free connection, $\Gamma^\mu_{\nu\rho}(p) = 0$.

Proof. Let λ be the unique geodesic through p with tangent vector X_p at p. By Lemma 6.2, $\lambda(t) = \exp\left(tX_p\right)$. In normal coordinates, the coordinates of $\lambda(t)$ are $\left(tX_p^1, tX_p^2, \ldots, tX_p^n\right)$. We write this by $x^\mu(t) = tX^\mu\big|_p$. So $\frac{\mathrm{d}x^\mu}{\mathrm{d}t} = X^\mu\big|_p$.

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}t^2} = \frac{\mathrm{d}}{\mathrm{d}t}\left(X^\mu\big|_p\right) = 0. \tag{6.17}$$

Then the geodesic equation reduces to

$$\begin{aligned} \frac{\mathrm{d}^2 x^\mu}{\mathrm{d}t^2} + \Gamma^\mu_{\nu\rho}\frac{\mathrm{d}x^\nu}{\mathrm{d}t}\frac{\mathrm{d}x^\rho}{\mathrm{d}t} &= 0 \\ \implies \Gamma^\mu_{\nu\rho}\left(\lambda(t)\right) X^\nu\big|_p X^\rho\big|_p &= 0. \end{aligned} \tag{6.18}$$

At $t = 0$, $\lambda(0) = p$, so we have

$$\Gamma^\mu_{\nu\rho}(p)\, X^\nu\big|_p X^\rho\big|_p = 0. \tag{6.19}$$

We can interchange ν and ρ to get

$$\Gamma^\mu_{\rho\nu}(p)\, X^\nu\big|_p X^\rho\big|_p = 0. \tag{6.20}$$

Combining 6.19 and 6.20,

$$\Gamma^\mu_{(\nu\rho)}(p)\, X^\nu\big|_p X^\rho\big|_p = 0, \tag{6.21}$$

which is true for any arbitrary $X_p \in T_pM$. Therefore, $\Gamma^\mu_{(\nu\rho)} = 0$ at p. For a torsion-free connection, $\Gamma^\mu_{[\nu\rho]} = 0$ everywhere. Hence, $\Gamma^\mu_{\nu\rho}(p) = 0$. ■

Remark 6.1. In general, $\Gamma^\mu_{\nu\rho}$ is not 0 away from the origin of normal coordinates. From 6.18, we can't conclude $\Gamma^\mu_{\nu\rho}(\lambda(t))$ is 0 for all t. Because, when we are writing 6.18, the tangent vector X_p is fixed. For q near p, if $q = \exp\left(Y_p\right)$ for some $Y_p \in T_pM$, then $q = \gamma(1)$, where γ is the unique geodesic through p with tangent vector Y_p at p. Then the analogouos equation of 6.18 is

$$\Gamma^\mu_{\nu\rho}\left(\gamma(t)\right) Y^\nu\big|_p Y^\rho\big|_p = 0. \tag{6.22}$$

In particular, $\Gamma^\mu_{\nu\rho}(q)\, Y^\nu\big|_p Y^\rho\big|_p = 0$. This equation is not necessarily true for every $Y_p \in T_pM$. That's why we can't conclude $\Gamma^\mu_{\nu\rho}(q)$. However, for **any** $Y_p \in T_pM$, the unique γ geodesic through p with tangent vector Y_p at p starts at 0, i.e. $\gamma(0) = p$. T hat's why we can take $t = 0$ in 6.18 and get 6.19, for **any** tangent vector, and hence we can conclude Lemma 6.3.

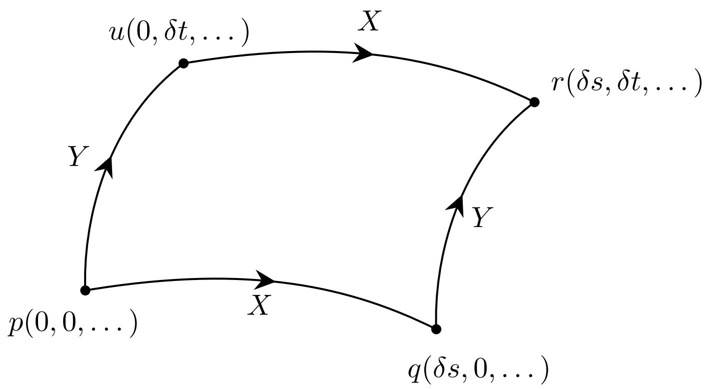
Lemma 6.4

On a manifold with metric, if the Levi-Civita connection is used to define normal coordinates at $p \in M$, then
$$g_{\mu\nu,\rho}(p) = 0. \tag{6.23}$$

Proof. Levi-Civita connection is torsion-free, so we have $\Gamma^{\sigma}_{\nu\rho}(p) = 0$ in normal coordinates. Now, using 5.45,
$$\Gamma^{\sigma}_{\nu\rho} = \frac{1}{2} g^{\sigma\tau}\left(g_{\nu\tau,\rho} + g_{\tau\rho,\nu} - g_{\rho\nu,\tau}\right). \tag{6.24}$$
Contracting by $2g_{\mu\sigma}$ yields
$$2g_{\mu\sigma}\Gamma^{\sigma}_{\nu\rho} = g_{\nu\mu,\rho} + g_{\mu\rho,\nu} - g_{\rho\nu,\mu}. \tag{6.25}$$
Now,
$$g_{\mu\nu,\rho} + g_{\mu\rho,\nu} - g_{\rho\nu,\mu} = 0 \tag{6.26}$$
at $p \in M$. Interchanging μ and ν, we get
$$g_{\nu\mu,\rho} + g_{\nu\rho,\mu} - g_{\rho\mu,\nu} = 0 \tag{6.27}$$
at $p \in M$. Adding 6.26 and 6.27, we get
$$2g_{\nu\mu,\rho} = 0 \text{ at } p \in M. \tag{6.28}$$
Therefore, $g_{\mu\nu,\rho}(p) = 0$. ■

§6.3 Parallel Transport and Curvature

Let X and Y be vector fields that are linearly independent everywhere, with $[X, Y] = 0$. Suppose we are dealing with a torsion-free connection. We can choose coordinates $(s, t, \ldots)$ such that
$$X = \frac{\partial}{\partial s}, \text{ and } Y = \frac{\partial}{\partial t}. \tag{6.29}$$
Let $p, q, r, u \in M$ along integral curves of X, Y with coordinates $p = (0, 0, \ldots)$, $q = (\delta s, 0, \ldots)$, $r = (\delta s, \delta t, \ldots)$, $u = (0, \delta t, \ldots)$. Let $Z_p \in T_pM$. We first parallel transport Z_p along pqr to obtain $Z_r \in T_rM$. Then we parallel transport Z_p along pur to obtain $Z'_r \in T_rM$. Then we shall compute $Z'_r - Z_r$ assuming a torsion-free connection. We shall use normal coordinates at p.



From p to q : pq is an integral curve of X, and Z is parallely transported along this curve. Therefore, $\nabla_X Z = 0$.

$$\begin{aligned} X^\sigma Z^\mu_{;\sigma} = 0 &\implies X^\sigma \left(\frac{\partial Z^\mu}{\partial x^\lambda} + \Gamma^\mu_{\rho\sigma} Z^\rho \right) = 0 \\ &\implies \frac{\mathrm{d}x^\sigma}{\mathrm{d}s} \frac{\partial Z^\mu}{\partial x^\lambda} + \Gamma^\mu_{\rho\sigma} Z^\rho X^\sigma = 0 \\ &\implies \frac{\mathrm{d}Z^\mu}{\mathrm{d}s} + \Gamma^\mu_{\rho\sigma} Z^\rho X^\sigma = 0 \qquad (6.30) \\ &\implies \frac{\mathrm{d}^2 Z^\mu}{\mathrm{d}s^2} = -\frac{\mathrm{d}x^\lambda}{\mathrm{d}s} \frac{\partial}{\partial x^\lambda} \left(\Gamma^\mu_{\rho\sigma} Z^\rho X^\sigma \right) \\ &\implies \frac{\mathrm{d}^2 Z^\mu}{\mathrm{d}s^2} = -X^\lambda \frac{\partial}{\partial x^\lambda} \left(\Gamma^\mu_{\rho\sigma} Z^\rho X^\sigma \right) \qquad (6.31) \end{aligned}$$

Now, using 6.30 and 6.31, we expand Z^μ in Taylor series around p and set $\Gamma^\mu_{\rho\sigma}(p) = 0$ since we are working in normal coordinates at p.

$$\begin{aligned} Z^\mu_q - Z^\mu_p &= \frac{\mathrm{d}Z^\mu}{\mathrm{d}s}(p)\,\delta s + \frac{1}{2}\frac{\mathrm{d}^2 Z^\mu}{\mathrm{d}s^2}(p)\,\delta s^2 + \mathcal{O}\left(\delta s^3\right) \\ &= -\Gamma^\mu_{\rho\sigma}(p)\, Z^\rho_p X^\sigma_p \delta s - \frac{1}{2} X^\lambda_p \left.\frac{\partial}{\partial x^\lambda}\right|_p \left(\Gamma^\mu_{\rho\sigma} Z^\rho X^\sigma \right) \delta s^2 + \mathcal{O}\left(\delta s^3\right) \\ &= -\frac{1}{2} X^\lambda_p Z^\rho_p X^\sigma_p \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta s^2 + \mathcal{O}\left(\delta s^3\right). \qquad (6.32) \end{aligned}$$

From q to r: qr is an integral curve of Y, and Z is parallely transported along this curve. Therefore, $\nabla_Y Z = 0$. Then following a similar procedure as above will lead us to

$$\frac{\mathrm{d}Z^\mu}{\mathrm{d}t} = -\Gamma^\mu_{\rho\sigma} Z^\rho Y^\sigma \quad \text{and} \quad \frac{\mathrm{d}^2 Z^\mu}{\mathrm{d}t^2} = -Y^\lambda \frac{\partial}{\partial x^\lambda}\left(\Gamma^\mu_{\rho\sigma} Z^\rho Y^\sigma \right). \qquad (6.33)$$

Using 6.33, we expand Z^μ in Taylor series again, this time around q.

$$\begin{aligned} Z^\mu_r - Z^\mu_q &= \frac{\mathrm{d}Z^\mu}{\mathrm{d}t}(q)\,\delta t + \frac{1}{2}\frac{\mathrm{d}^2 Z^\mu}{\mathrm{d}t^2}(q)\,\delta t^2 + \mathcal{O}\left(\delta t^3\right) \\ &= -\Gamma^\mu_{\rho\sigma}(q)\, Z^\rho_q Y^\sigma_q \delta t - \frac{1}{2} Y^\lambda_q \left.\frac{\partial}{\partial x^\lambda}\right|_q \left(\Gamma^\mu_{\rho\sigma} Z^\rho Y^\sigma \right) \delta t^2 + \mathcal{O}\left(\delta t^3\right) \qquad (6.34) \end{aligned}$$

Now, we shall expand $\Gamma^\mu_{\rho\sigma}(q)\, Z^\rho_q Y^\sigma_q$ and $Y^\lambda_q \left.\frac{\partial}{\partial x^\lambda}\right|_q \left(\Gamma^\mu_{\rho\sigma} Z^\rho Y^\sigma \right)$ in Taylor series around p. Since we are only interested in up to second order terms in 6.34, we shall expand the former one up to first order and the latter one up to zeroth order.

$$\begin{aligned} \Gamma^\mu_{\rho\sigma}(q)\, Z^\rho_q Y^\sigma_q &= \Gamma^\mu_{\rho\sigma}(p)\, Z^\rho_p Y^\sigma_p + \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_p \left(\Gamma^\mu_{\rho\sigma} Z^\rho Y^\sigma \right) \delta s + \mathcal{O}\left(\delta s^2\right) \\ &= \frac{\partial x^\lambda}{\partial s} \left.\frac{\partial}{\partial x^\lambda}\right|_p \left(\Gamma^\mu_{\rho\sigma} Z^\rho Y^\sigma \right) \delta s + \mathcal{O}\left(\delta s^2\right) \\ &= X^\lambda_p Z^\rho_p Y^\sigma_p \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta s + \mathcal{O}\left(\delta s^2\right). \qquad (6.35) \end{aligned}$$

$$Y^\lambda_q \left.\frac{\partial}{\partial x^\lambda}\right|_q \left(\Gamma^\mu_{\rho\sigma} Z^\rho Y^\sigma \right) = Y^\lambda_p \left.\frac{\partial}{\partial x^\lambda}\right|_p \left(\Gamma^\mu_{\rho\sigma} Z^\rho Y^\sigma \right) + \mathcal{O}\left(\delta s\right) = Y^\lambda_p Z^\rho_p Y^\sigma_p \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p) + \mathcal{O}\left(\delta s\right). \qquad (6.36)$$

Using 6.35 and 6.36, 6.34 reads

$$Z^\mu_r - Z^\mu_q = -X^\lambda_p Z^\rho_p Y^\sigma_p \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta s \delta t - \frac{1}{2} Y^\lambda_p Z^\rho_p Y^\sigma_p \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta t^2 + \mathcal{O}\left(\delta^3\right). \qquad (6.37)$$

Adding 6.32 and 6.37, we get

$$
\begin{aligned}
Z_r^\mu - Z_p^\mu = &-\frac{1}{2} X_p^\lambda Z_p^\rho X_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta s^2 - X_p^\lambda Z_p^\rho Y_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta s \delta t \\
&-\frac{1}{2} Y_p^\lambda Z_p^\rho Y_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta t^2 + \mathcal{O}\left(\delta^3\right).
\end{aligned}
\tag{6.38}
$$

From p to u: pu is an integral curve of Y, and Z is parallely transported along this curve. Therefore, $\nabla_Y Z = 0$. This leads us to 6.33. Now we expand Z^μ in Taylor seris around p.

$$
\begin{aligned}
Z_u^\mu - Z_p^\mu &= \frac{\mathrm{d}Z^\mu}{\mathrm{d}t}(p)\,\delta t + \frac{1}{2}\frac{\mathrm{d}^2 Z^\mu}{\mathrm{d}t^2}(p)\,\delta t^2 + \mathcal{O}\left(\delta t^3\right) \\
&= -\Gamma^\mu_{\rho\sigma}(p)\, Z_p^\rho Y_p^\sigma \delta t - \frac{1}{2} Y_p^\lambda \left.\frac{\partial}{\partial x^\lambda}\right|_p \left(\Gamma^\mu_{\rho\sigma} Z^\rho Y^\sigma\right)\delta t^2 + \mathcal{O}\left(\delta t^3\right) \\
&= -\frac{1}{2} Y_p^\lambda Z_p^\rho Y_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta t^2 + \mathcal{O}\left(\delta t^3\right)
\end{aligned}
\tag{6.39}
$$

From u to r: ur is an integral curve of X, and Z is parallely transported along this curve. Therefore, $\nabla_X Z = 0$. This leads us to 6.30 and 6.31. Now, we expand Z^μ in Taylor series around u.

$$
\begin{aligned}
Z_r'^\mu - Z_u^\mu &= \frac{\mathrm{d}Z^\mu}{\mathrm{d}s}(u)\,\delta s + \frac{1}{2}\frac{\mathrm{d}^2 Z^\mu}{\mathrm{d}s^2}(u)\,\delta s^2 + \mathcal{O}\left(\delta s^3\right) \\
&= -\Gamma^\mu_{\rho\sigma}(u)\, Z_u^\rho X_u^\sigma \delta s - \frac{1}{2} X_u^\lambda \left.\frac{\partial}{\partial x^\lambda}\right|_u \left(\Gamma^\mu_{\rho\sigma} Z^\rho X^\sigma\right)\delta s^2 + \mathcal{O}\left(\delta s^3\right)
\end{aligned}
\tag{6.40}
$$

$$
\begin{aligned}
\Gamma^\mu_{\rho\sigma}(u)\, Z_u^\rho X_u^\sigma &= \Gamma^\mu_{\rho\sigma}(p)\, Z_p^\rho X_p^\sigma + \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_p \left(\Gamma^\mu_{\rho\sigma} Z^\rho X^\sigma\right)\delta t + \mathcal{O}\left(\delta t^2\right) \\
&= \frac{\partial x^\lambda}{\partial t}\left.\frac{\partial}{\partial x^\lambda}\right|_p \left(\Gamma^\mu_{\rho\sigma} Z^\rho X^\sigma\right)\delta t + \mathcal{O}\left(\delta t^2\right) \\
&= Y_p^\lambda Z_p^\rho X_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta t + \mathcal{O}\left(\delta t^2\right).
\end{aligned}
\tag{6.41}
$$

$$
X_u^\lambda \left.\frac{\partial}{\partial x^\lambda}\right|_u \left(\Gamma^\mu_{\rho\sigma} Z^\rho X^\sigma\right) = X_p^\lambda \left.\frac{\partial}{\partial x^\lambda}\right|_p \left(\Gamma^\mu_{\rho\sigma} Z^\rho X^\sigma\right) + \mathcal{O}(\delta t) = X_p^\lambda Z_p^\rho X_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p) + \mathcal{O}(\delta s). \tag{6.42}
$$

Using 6.41 and 6.42, 6.40 reads

$$
Z_r'^\mu - Z_u^\mu = -Y_p^\lambda Z_p^\rho X_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta t \delta s - \frac{1}{2} X_p^\lambda Z_p^\rho X_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta s^2 + \mathcal{O}\left(\delta^3\right). \tag{6.43}
$$

Adding 6.39 and 6.43, we get

$$
\begin{aligned}
Z_r'^\mu - Z_p^\mu = &-\frac{1}{2} Y_p^\lambda Z_p^\rho Y_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta t^2 - Y_p^\lambda Z_p^\rho X_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta t \delta s \\
&-\frac{1}{2} X_p^\lambda Z_p^\rho X_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\,\delta s^2 + \mathcal{O}\left(\delta^3\right).
\end{aligned}
\tag{6.44}
$$

Subtracting 6.38 from 6.44 yields

$$
\begin{aligned}
Z_r'^\mu - Z_r^\mu &= \left(X_p^\lambda Z_p^\rho Y_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p) - Y_p^\lambda Z_p^\rho X_p^\sigma \frac{\partial \Gamma^\mu_{\rho\sigma}}{\partial x^\lambda}(p)\right)\delta s \delta t + \mathcal{O}\left(\delta^3\right) \\
&= \left.\left(\Gamma^\mu_{\rho\sigma,\lambda} Z^\rho \left(X^\lambda Y^\sigma - Y^\lambda X^\sigma\right)\right)\right|_p \delta s \delta t + \mathcal{O}\left(\delta^3\right) \\
&= \left.\left(\left(\Gamma^\mu_{\rho\sigma,\lambda} - \Gamma^\mu_{\rho\lambda,\sigma}\right) Z^\rho X^\lambda Y^\sigma\right)\right|_p \delta s \delta t + \mathcal{O}\left(\delta^3\right).
\end{aligned}
\tag{6.45}
$$

We have derived that

$$
R^\mu{}_{\nu\rho\sigma} = \Gamma^\mu_{\nu\sigma,\rho} - \Gamma^\mu_{\nu\rho,\sigma} + \Gamma^\tau_{\nu\sigma}\Gamma^\mu_{\tau\rho} - \Gamma^\tau_{\nu\rho}\Gamma^\mu_{\tau\sigma}. \tag{6.46}
$$

Since the components of Γ are 0 at p, $R^{\mu}{}_{\nu\rho\sigma}\big|_p = \Gamma^{\mu}_{\nu\sigma,\rho}\big|_p - \Gamma^{\mu}_{\nu\rho,\sigma}\big|_p$. Therefore, from 6.45,

$$Z'^{\mu}_r - Z^{\mu}_r = \left(R^{\mu}{}_{\rho\lambda\sigma} Z^{\rho} X^{\lambda} Y^{\sigma}\right)\Big|_p \delta s \delta t + \mathcal{O}\left(\delta^3\right). \tag{6.47}$$

Now,

$$\begin{aligned}
\left(R^{\mu}{}_{\rho\lambda\sigma} Z^{\rho} X^{\lambda} Y^{\sigma}\right)\Big|_r &= \left(R^{\mu}{}_{\rho\lambda\sigma} Z^{\rho} X^{\lambda} Y^{\sigma}\right)\Big|_q + \mathcal{O}(\delta t) \\
&= \left(R^{\mu}{}_{\rho\lambda\sigma} Z^{\rho} X^{\lambda} Y^{\sigma}\right)\Big|_p + \mathcal{O}(\delta s) + \mathcal{O}(\delta t) \\
\therefore \left(R^{\mu}{}_{\rho\lambda\sigma} Z^{\rho} X^{\lambda} Y^{\sigma}\right)\Big|_p &= \left(R^{\mu}{}_{\rho\lambda\sigma} Z^{\rho} X^{\lambda} Y^{\sigma}\right)\Big|_r + \mathcal{O}(\delta). && (6.48)
\end{aligned}$$

Combining 6.47 and 6.48, we get

$$\begin{aligned}
\frac{Z'^{\mu}_r - Z^{\mu}_r}{\delta s \delta t} &= \left(R^{\mu}{}_{\rho\lambda\sigma} Z^{\rho} X^{\lambda} Y^{\sigma}\right)\Big|_r + \mathcal{O}(\delta) && (6.49) \\
\therefore \lim_{\delta \to 0} \frac{Z'^{\mu}_r - Z^{\mu}_r}{\delta s \delta t} &= \left(R^{\mu}{}_{\rho\lambda\sigma} Z^{\rho} X^{\lambda} Y^{\sigma}\right)\Big|_r. && (6.50)
\end{aligned}$$

§6.4 Properties of Riemann Curvature Tensor

Proposition 6.5

If ∇ is torsion-free, then

$$R^{\mu}{}_{[\nu\rho\sigma]} = 0. \tag{6.51}$$

Proof. Let $p \in M$ and choose normal coordinates at p. Then at p

$$\begin{aligned}
R^{\mu}{}_{\nu\rho\sigma} &= \partial_{\rho}\Gamma^{\mu}_{\nu\sigma} - \partial_{\sigma}\Gamma^{\mu}_{\nu\rho}, && (6.52) \\
R^{\mu}{}_{\sigma\nu\rho} &= \partial_{\nu}\Gamma^{\mu}_{\sigma\rho} - \partial_{\rho}\Gamma^{\mu}_{\sigma\nu}, && (6.53) \\
R^{\mu}{}_{\rho\sigma\nu} &= \partial_{\sigma}\Gamma^{\mu}_{\rho\nu} - \partial_{\nu}\Gamma^{\mu}_{\rho\sigma}. && (6.54)
\end{aligned}$$

For a torsion-free connection, $\Gamma^{\mu}_{\nu\rho} = \Gamma^{\mu}_{\rho\nu}$. Using this, we can add 6.52, 6.53, 6.54 to get

$$R^{\mu}{}_{\nu\rho\sigma} + R^{\mu}{}_{\sigma\nu\rho} + R^{\mu}{}_{\rho\sigma\nu} = 0. \tag{6.55}$$

From 6.8, $R^{\mu}{}_{\nu\rho\sigma} = -R^{\mu}{}_{\nu\sigma\rho}$. Therefore,

$$\begin{aligned}
R^{\mu}{}_{[\nu\rho\sigma]} &= \frac{1}{3!}\left[R^{\mu}{}_{\nu\rho\sigma} + R^{\mu}{}_{\sigma\nu\rho} + R^{\mu}{}_{\rho\sigma\nu} - R^{\mu}{}_{\nu\sigma\rho} + R^{\mu}{}_{\sigma\rho\nu} + R^{\mu}{}_{\rho\nu\sigma}\right] \\
&= \frac{1}{3}\left[R^{\mu}{}_{\nu\rho\sigma} + R^{\mu}{}_{\sigma\nu\rho} + R^{\mu}{}_{\rho\sigma\nu}\right] \\
&= 0. && (6.56)
\end{aligned}$$

Therefore, $R^{\mu}{}_{[\nu\rho\sigma]} = 0$ at p in normal coordinates. If this is true in one basis then it is true in any basis. Therefore, $R^{\mu}{}_{[\nu\rho\sigma]}(p) = 0$ in any basis. Furthermore, p is arbitrary, so $R^{\mu}{}_{[\nu\rho\sigma]} = 0$ at all $p \in M$. ■

Proposition 6.5 is also known as the first Bianchi idntity.

Proposition 6.6 (Bianchi Identity)

If ∇ is torsion-free, then

$$R^{a}{}_{b[cd;e]} = 0. \tag{6.57}$$

Proof. Observe that

$$R^{a}{}_{bcd;e} = R^{a}{}_{bcd,e} + \Gamma^{a}_{ge} R^{g}{}_{bcd} - \Gamma^{g}_{be} R^{a}{}_{gcd} - \Gamma^{g}_{ce} R^{a}{}_{bgd} - \Gamma^{g}_{de} R^{a}{}_{bcg}. \tag{6.58}$$

Now, we use normal coordinates at $p \in M$. So $\Gamma^{a}_{bc}(p) = 0$. Therefore,

$$R^{a}{}_{bcd;e} = R^{a}{}_{bcd,e} = \partial_e R^{a}{}_{bcd} \tag{6.59}$$

at $p \in M$. Now, from 6.11, we get that $R^{\mu}{}_{\nu\rho\sigma} = \partial_\rho \Gamma^{\mu}_{\nu\sigma} - \partial_\sigma \Gamma^{\mu}_{\nu\rho}$ at $p \in M$ in normal coordinates. Therefore,

$$R^{\mu}{}_{\nu\rho\sigma;\tau}(p) = R^{\mu}{}_{\nu\rho\sigma,\tau}(p) = \partial_\tau \partial_\rho \Gamma^{\mu}_{\nu\sigma}\Big|_p - \partial_\tau \partial_\sigma \Gamma^{\mu}_{\nu\rho}\Big|_p. \tag{6.60}$$

Similarly,

$$R^{\mu}{}_{\nu\tau\rho;\sigma}(p) = \partial_\sigma \partial_\tau \Gamma^{\mu}_{\nu\rho}\Big|_p - \partial_\sigma \partial_\rho \Gamma^{\mu}_{\nu\tau}\Big|_p, \tag{6.61}$$

$$R^{\mu}{}_{\nu\sigma\tau;\rho}(p) = \partial_\rho \partial_\sigma \Gamma^{\mu}_{\nu\tau}\Big|_p - \partial_\rho \partial_\tau \Gamma^{\mu}_{\nu\sigma}\Big|_p. \tag{6.62}$$

Since partial derivatives commute, adding 6.60, 6.61, 6.62, we get

$$R^{\mu}{}_{\nu\rho\sigma;\tau}(p) + R^{\mu}{}_{\nu\tau\rho;\sigma}(p) + R^{\mu}{}_{\nu\sigma\tau;\rho}(p) = 0. \tag{6.63}$$

From 6.8, $R^{\mu}{}_{\nu\rho\sigma} = -R^{\mu}{}_{\nu\sigma\rho}$. Therefore,

$$\begin{aligned} R^{\mu}{}_{\nu[\rho\sigma;\tau]}(p) &= \frac{1}{3!}\left[R^{\mu}{}_{\nu\rho\sigma;\tau}(p) + R^{\mu}{}_{\nu\tau\rho;\sigma}(p) + R^{\mu}{}_{\nu\sigma\tau;\rho}(p) - R^{\mu}{}_{\nu\sigma\rho;\tau}(p) - R^{\mu}{}_{\nu\rho\tau;\sigma}(p) - R^{\mu}{}_{\nu\tau\sigma;\rho}(p)\right] \\ &= \frac{1}{3}\left[R^{\mu}{}_{\nu\rho\sigma;\tau}(p) + R^{\mu}{}_{\nu\tau\rho;\sigma}(p) + R^{\mu}{}_{\nu\sigma\tau;\rho}(p)\right] \\ &= 0. \end{aligned} \tag{6.64}$$

Therefore, $R^{\mu}{}_{\nu[\rho\sigma;\tau]} = 0$ at p in normal coordinates. If this is true in one basis then it is true in any basis. Therefore, $R^{\mu}{}_{\nu[\rho\sigma;\tau]}(p) = 0$ in any basis. Furthermore, p is arbitrary, so $R^{\mu}{}_{\nu[\rho\sigma;\tau]} = 0$ at all $p \in M$. ■

Proposition 6.7 (Ricci Identity)

For torsion-free connection,

$$\nabla_c \nabla_d Z^a - \nabla_d \nabla_c Z^a = R^{a}{}_{bcd} Z^b. \tag{6.65}$$

Proof. We shall prove that

$$X^c Y^d \nabla_c \nabla_d Z^a - X^c Y^d \nabla_d \nabla_c Z^a = R^{a}{}_{bcd} Z^b X^c Y^d. \tag{6.66}$$

The LHS is

$$\begin{aligned} &X^c Y^d \nabla_c \nabla_d Z^a - X^c Y^d \nabla_d \nabla_c Z^a \\ &= \left[X^c \nabla_c \left(Y^d \nabla_d Z^a\right) - X^c \left(\nabla_c Y^d\right) \nabla_d Z^a\right] - \left[Y^d \nabla_d \left(X^c \nabla_c Z^a\right) - Y^d \left(\nabla_d X^c\right) \nabla_c Z^a\right]. \end{aligned} \tag{6.67}$$

Now, $Y^d \nabla_d Z^a = (\nabla_Y Z)^a$, so $X^c \nabla_c \left(Y^d \nabla_d Z^a\right) = (\nabla_X \nabla_Y Z)^a$. Similarly, $Y^d \nabla_d \left(X^c \nabla_c Z^a\right) = (\nabla_Y \nabla_X Z)^a$. Also, for a torsion-free connection, $[X, Y] = \nabla_X Y - \nabla_Y X$.

$$\begin{aligned} \left(\nabla_{[X,Y]} Z\right)^a &= [X, Y]^b \nabla_b Z^a = (\nabla_X Y - \nabla_Y X)^b \nabla_b Z^a \\ &= (\nabla_X Y)^d \nabla_d Z^a - (\nabla_Y X)^c \nabla_c Z^a \\ &= X^c \left(\nabla_c Y^d\right) - Y^d \left(\nabla_d X^c\right) \nabla_c Z^a. \end{aligned} \tag{6.68}$$

Plugging 6.68 into 6.67, we get

$$\begin{aligned} & X^c Y^d \nabla_c \nabla_d Z^a - X^c Y^d \nabla_d \nabla_c Z^a \\ = & \left(\nabla_X \nabla_Y Z\right)^a - \left(\nabla_Y \nabla_X Z\right)^a - \left(\nabla_{[X,Y]} Z\right)^a \\ = & \left(R\left(X,Y\right)Z\right)^a = R^a{}_{bcd} Z^b X^c Y^d. \end{aligned} \tag{6.69}$$

Hence, 6.66 is proved. Since $X, Y \in \mathfrak{X}\left(M\right)$ are arbitrary, 6.65 holds and we are done. ■

Ricci Identity tells us that the second covariant derivative of a vector field commutes if and only if the Riemann curvature tensor $R^a{}_{bcd}$ vanishes.

§6.5 Geodesic Deviation

Consider a family γ_s of geodesics each labeled by $s \in \mathbb{R}$, and each of these geodesics is affinely parametrized by t. The map $(s,t) \mapsto \gamma_s\left(t\right)$ from $\mathbb{R} \to M$ is smooth with a smooth inverse. This implies that the family of geodesics form a 2d surface Σ imbedded in M.

Let T be the vector field tangent to the geodesics. One moves along a geodesic by varying t (s has to be fixed since one sticks to a certain geodesic). Similarly, one can move across the geodesics (for constant t) by varying s smoothly and obtain curves due to constant t.

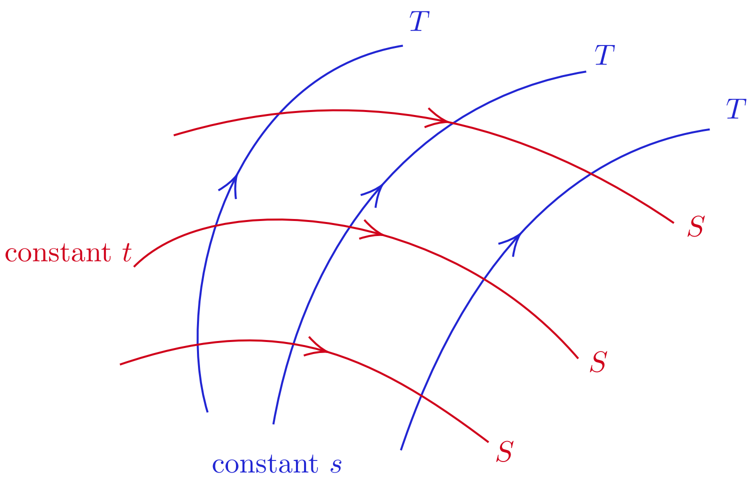


Figure 6.1: $\Sigma \subset M$.

Now, consider a chart x^μ and express the geodesics in this chart by $x^\mu\left(s,t\right)$. In this chart, the geodesics are specified by $S^\mu = \frac{\partial x^\mu}{\partial s}$ and $T^\mu = \frac{\partial x^\mu}{\partial t}$. In other words,

$$S = S^\mu \frac{\partial}{\partial x^\mu} = \frac{\partial x^\mu}{\partial s}\frac{\partial}{\partial x^\mu} = \frac{\partial}{\partial s}, \tag{6.70}$$

and similarly, $T = \frac{\partial}{\partial t}$. Now we taylor expand the geodesic $x^\mu\left(s,t\right)$ around s with small perturbation δs.

$$x^\mu\left(s+\delta s, t\right) = x^\mu\left(s,t\right) + \delta s\, S^\mu + \mathcal{O}\left(\delta s^2\right). \tag{6.71}$$

Therefore, S^μ points towards neighboring geodesic. S^μ is known as **deviation vector**. Now, we want to know how S changes as we move along geodesics. We quantify this change by $\nabla_T S$. This quantifies if two nearby geodesics are moving towards or away from each other.

The vector field $V = \nabla_T S$ captures the or **relative velocity** between infinitesimally close geodesics. One can also define **relative acceletation** between neighboring geodesics as $A = \nabla_T V = \nabla_T\left(\nabla_T S\right)$.

Since S and T are basis vector fields for chart $(s, t, \ldots)$, i.e. $S = \frac{\partial}{\partial s}$ and $T = \frac{\partial}{\partial t}$, we have

$$[S, T] = 0. \tag{6.72}$$

We further impose torsion-free condition to ∇. So

$$\nabla_T S - \nabla_S T = [T, S] = 0. \tag{6.73}$$

Hence, $\nabla_T S = \nabla_S T$. Therefore,

$$\nabla_T \nabla_T S = \nabla_T \nabla_S T = \nabla_S \nabla_T T + \nabla_{[T,S]} T + R(T, S) T. \tag{6.74}$$

Since the integral curve of T is a geodesic, $\nabla_T T = 0$. Also, the commutator of T and S is 0. Therefore,

$$A = \nabla_T \nabla_T S = R(T, S) T. \tag{6.75}$$

In a coordinate chart,

$$A^\mu = R^\mu{}_{\nu\rho\sigma} T^\nu T^\rho T^\sigma. \tag{6.76}$$

6.76 is known as the **geodesic deviation equation**.

Remark 6.2. Two initially parallel geodesics, i.e. $V^\mu = (\nabla_T S)^\mu = 0$ initially (meaning they neither move toward nor move apart from each other), will fail to remain parallel (meaning there will be non-zero accelerationo) if and only if $R^\mu{}_{\nu\rho\sigma} \neq 0$.

§6.6 Curvature of Levi-Civita connection

From now on, we shall restrict attention to a manifold with metric, and use the Levi-Civita connection. The Riemann tensor then enjoys additional symmetries. We can lower an index using the metric:

$$R_{\tau\nu\rho\sigma} = g_{\tau\mu} R^\mu{}_{\nu\rho\sigma}. \tag{6.77}$$

Proposition 6.8

The Riemann tensor satisfies

$$R_{abcd} = R_{cdab}\ , \quad \text{and } R_{(ab)cd} = 0. \tag{6.78}$$

Proof. By Lemma 6.4, $g_{\mu\nu,\rho}(p) = 0$ in normal coordinates at $p \in M$. Now, $\delta^\tau{}_\nu = g^{\tau\mu} g_{\mu\nu}$, so taking ∂_ρ gives us

$$\partial_\rho (g^{\tau\mu} g_{\mu\nu}) = 0 \implies g_{\mu\nu} \partial_\rho g^{\tau\mu} = 0 \implies \partial_\rho g^{\tau\mu} = 0, \tag{6.79}$$

at $p \in M$. Now, by 5.45,

$$\Gamma^\tau_{\nu\sigma} = \frac{1}{2} g^{\tau\mu} (g_{\mu\nu,\sigma} + g_{\mu\sigma,\nu} - g_{\nu\sigma,\mu}). \tag{6.80}$$

Taking ∂_ρ and using 6.79, we get

$$\partial_\rho \Gamma^\tau_{\nu\sigma} = \frac{1}{2} g^{\tau\mu} (g_{\mu\nu,\sigma\rho} + g_{\mu\sigma,\nu\rho} - g_{\nu\sigma,\mu\rho}). \tag{6.81}$$

Therefore, at $p \in M$,

$$\begin{aligned} R^\tau{}_{\nu\rho\sigma} &= \partial_\rho \Gamma^\tau_{\nu\sigma} - \partial_\sigma \Gamma^\tau_{\nu\rho} \\ &= \frac{1}{2} g^{\tau\mu} (g_{\mu\sigma,\nu\rho} - g_{\nu\sigma,\mu\rho} - g_{\mu\rho,\nu\sigma} + g_{\nu\rho,\mu\sigma}). \end{aligned} \tag{6.82}$$

As a result, at $p \in M$,

$$\begin{aligned} R_{\alpha\nu\rho\sigma} &= g_{\alpha\tau} R^\tau{}_{\nu\rho\sigma} \\ &= \frac{1}{2} g_{\alpha\tau} g^{\tau\mu} (g_{\mu\sigma,\nu\rho} - g_{\nu\sigma,\mu\rho} - g_{\mu\rho,\nu\sigma} + g_{\nu\rho,\mu\sigma}) \\ &= \frac{1}{2} (g_{\alpha\sigma,\nu\rho} - g_{\nu\sigma,\alpha\rho} - g_{\alpha\rho,\nu\sigma} + g_{\nu\rho,\alpha\sigma}) \\ \therefore R_{\mu\nu\rho\sigma} &= \frac{1}{2} (g_{\mu\sigma,\nu\rho} - g_{\nu\sigma,\mu\rho} - g_{\mu\rho,\nu\sigma} + g_{\nu\rho,\mu\sigma}). \end{aligned} \tag{6.83}$$

Similarly,

$$R_{\rho\sigma\mu\nu} = \frac{1}{2}\left(g_{\rho\nu,\sigma\mu} - g_{\sigma\nu,\rho\mu} - g_{\rho\mu,\sigma\nu} + g_{\sigma\mu,\rho\nu}\right). \tag{6.84}$$

Comparing 6.83 and 6.84 keeping in mind that g is symmetric and partial derivatives commute, we get that $R_{\mu\nu\rho\sigma} = R_{\rho\sigma\mu\nu}$ at p. Since it is a tensorial equation and it holds at $p \in M$ in a particular basis, it holds at $p \in M$ in any other basis. Since $p \in M$ was arbitrary, $R_{\mu\nu\rho\sigma} = R_{\rho\sigma\mu\nu}$ everywhere.

For the second part, recall 6.9 that $R^{\mu}{}_{\nu(\rho\sigma)} = 0$. Therefore, $R^{\mu}{}_{\nu\rho\sigma} = -R^{\mu}{}_{\nu\sigma\rho}$. Lowering the index μ gives us

$$g_{\mu\tau}R^{\mu}{}_{\nu\rho\sigma} = -g_{\mu\tau}R^{\mu}{}_{\nu\sigma\rho} \implies R_{\tau\nu\rho\sigma} = -R_{\tau\nu\sigma\rho}. \tag{6.85}$$

We have just proved that $R_{\tau\nu\rho\sigma} = R_{\rho\sigma\tau\nu}$. Therefore,

$$R_{\rho\sigma\tau\nu} = R_{\tau\nu\rho\sigma} = -R_{\tau\nu\sigma\rho} = -R_{\sigma\rho\tau\nu}. \tag{6.86}$$

Hence, $R_{(\sigma\rho)\tau\nu} = 0$. ■

Now we shall compute the number of independent components of Riemann tensor. We have proved that $R_{abcd} = -R_{bacd}$, and $R_{abcd} = R_{cdab}$. Since a and b are antisymmetric in R_{abcd}, the number of independent components in the first two indices is $N = \frac{n(n-1)}{2}$.

Furthermore, since $R_{abcd} = R_{cdab}$, the number of independent components in the last two indices is also $N = \frac{n(n-1)}{2}$. Now, $R_{abcd} = R_{cdab}$ tells us that the first two indices and the last two indices are symmetric. We have N independent choices for both of them. Therefore, the number of independent components are

$$\frac{1}{2}N(N+1) = \frac{1}{2}\frac{n(n-1)}{2}\left(\frac{n(n-1)}{2}+1\right) = \frac{1}{8}n(n-1)\left(n^2-n+2\right). \tag{6.87}$$

However, we have not taken into account the first Bianchi identity (Proposition 6.5) $R_{a[bcd]} = 0$ if the connection is torsion-free, i.e.

$$R_{abcd} + R_{acdb} + R_{adbc} = 0. \tag{6.88}$$

If any of the indices b, c, d are equal to a, let's say $b = a$, then 6.88 becomes

$$R_{aacd} + R_{acdb} + R_{adbc} = 0 \iff R_{acda} + R_{adac} = 0 \iff R_{adac} = -R_{acda} = -R_{daac}, \tag{6.89}$$

which gives is the antisymmetry of the first two indices. So this does not give us any new information. If any two of b, c, d are equal, let's say $b = c$, then 6.88 becomes

$$R_{abbd} + R_{abdb} + R_{adbb} = 0 \iff R_{abbd} + R_{abdb} = 0 \iff R_{abbd} = -R_{abdb}, \tag{6.90}$$

which is nothing but the antisymmetry of the last two indices. Therefore, if any two indices a, b, c, d are equal, the Bianchi identity (6.88) does not give us any new symmetries. If a, b, c, d are all distinct, we get $\binom{n}{4}$ constraints. Therefore, the total number of independednt components of R_{abcd} are

$$\begin{aligned}\frac{1}{8}n(n-1)\left(n^2-n+2\right) - \binom{n}{4} &= \frac{1}{8}n(n-1)\left(n^2-n+2\right) - \frac{n(n-1)(n-2)(n-3)}{24} \\ &= \frac{n(n-1)}{8}\left(n^2-n+2-\frac{(n-2)(n-3)}{3}\right) \\ &= \frac{n(n-1)}{8}\frac{3n^2-3n-n^2+5n}{3} \\ &= \frac{1}{12}n^2\left(n^2-1\right).\end{aligned} \tag{6.91}$$

Therefore, the number of independent components of R_{abcd} is $\frac{1}{12}n^2\left(n^2-1\right)$.

Definition 6.4 (Ricci Curvature Tensor). The **Ricci curvature tensor** is the $(0,2)$ tensor defined by

$$R_{ab} = R^c{}_{acb}. \tag{6.92}$$

Proposition 6.9

The Ricci curvature tensor of Levi-Civita connection is a symmetric $(0,2)$ tensor, i.e. $R_{ab} = R_{ba}$.

Proof.

$$R_{ab} = R^c{}_{acb} = g^{cd} R_{dacb} = g^{dc} R_{cbda} = R^d{}_{bda} = R_{ba}. \tag{6.93}$$

■

Definition 6.5 (Ricci Scalar). The Ricci Scalar is defined by

$$R = g^{ab} R_{ab}. \tag{6.94}$$

Definition 6.6 (Einstein Tensor). The **Einstein tensor** is the symmetric $(0,2)$ tensor defined by

$$G_{ab} = R_{ab} - \frac{1}{2} R g_{ab}. \tag{6.95}$$

Proposition 6.10 (Contracted Bianchi Identity)

The Einstein tensor satisfies the contracted Bianchi identity:

$$\nabla^a G_{ab} = 0, \tag{6.96}$$

or equivalently

$$\nabla^a R_{ab} - \frac{1}{2} \nabla_b R = 0. \tag{6.97}$$

Proof. Recall the Bianchi Identity we proved earlier: $R^a{}_{b[cd;e]} = 0$. In fact, during the course of the proof, we saw that

$$R^a{}_{bcd;e} + R^a{}_{bec;d} + R^a{}_{bde;c} = 0. \tag{6.98}$$

Contracting with the metric tensor g_{fa}, we obtain

$$R_{fbcd;e} + R_{fbec;d} + R_{fbde;c} = 0. \tag{6.99}$$

Therefore

$$\begin{aligned} & g^{bn} g^{am} \left(R_{abmn;l} + R_{ablm;n} + R_{abnl;m}\right) = 0 \\ \Longrightarrow\ & g^{bn} \left(R^m{}_{bmn;l} + R^m{}_{blm;n} + R^m{}_{bnl;m}\right) = 0 \\ \Longrightarrow\ & g^{bn} \left(R^m{}_{bmn;l} - R^m{}_{bml;n} + R^m{}_{bnl;m}\right) = 0 \\ \Longrightarrow\ & g^{bn} \left(R_{bn;l} - R_{bl;n} + R^m{}_{bnl;m}\right) = 0 \\ \Longrightarrow\ & R_{;l} - R^n{}_{l;n} + g^{bn} R^m{}_{bnl;m} = 0. \end{aligned} \tag{6.100}$$

Now,

$$g^{bn} R^m{}_{bnl;m} = g^{bn} g^{am} R_{abnl;m} = -g^{bn} g^{am} R_{banl;m} = -g^{am} R^n{}_{anl;m} = -g^{am} R_{al;m} = -R^m{}_{l;m}. \tag{6.101}$$

Plugging 6.101 into 6.100, we obtain

$$R_{;l} - R^n{}_{l;n} - R^m{}_{l;m} = 0 \implies R_{;l} - 2R^n{}_{l;n} = 0$$
$$\implies \frac{1}{2}\nabla_l R = \nabla_m R^m{}_l = \nabla_m\left(g^{mn}R_{nl}\right) = g^{mn}\nabla_m R_{nl}$$
$$\implies \frac{1}{2}\nabla_l R = \nabla^n R_{nl}. \tag{6.102}$$

$\nabla_l R = g_{nl}\nabla^n R = \nabla^n\left(Rg_{nl}\right)$. Therefore, 6.102 becomes $\nabla^n R^{nl} - \frac{1}{2}\nabla^n\left(Rg_{nl}\right) = 0$, which is equivalent to $\nabla^n G_{nl} = 0$. ■

Definition 6.7 (Weyl Tensor). The **Weyl tensor** C_{abcd} is the trace-free part of Riemann curvature tensor defined by

$$C_{abcd} = R_{abcd} - \frac{2}{n-2}\left(g_{a[c}R_{d]b} - g_{b[c}R_{d]a}\right) + \frac{2}{(n-1)(n-2)}Rg_{a[c}g_{d]b}. \tag{6.103}$$

Proposition 6.11

$g^{ac}C_{abcd} = 0.$

Proof. Multiplying both sides of 6.103 by g^{ac}, we get

$$g^{ac}C_{abcd} = g^{ac}R_{abcd} - \frac{2}{n-2}\left(g^{ac}g_{a[c}R_{d]b} - g^{ac}g_{b[c}R_{d]a}\right) + \frac{2}{(n-1)(n-2)}g^{ac}Rg_{a[c}g_{d]b}$$
$$= R_{bd} - \frac{2}{n-2}\left(g^{ac}g_{a[c}R_{d]b} - g^{ac}g_{b[c}R_{d]a}\right) + \frac{2}{(n-1)(n-2)}g^{ac}Rg_{a[c}g_{d]b}. \tag{6.104}$$

Now,

$$2g^{ac}g_{a[c}R_{d]b} = g^{ac}g_{ac}R_{db} - g^{ac}g_{ad}R_{cb} = \delta^a_a R_{db} - \delta^c_d R_{cb} = nR_{db} - R_{db}, \tag{6.105}$$
$$2g^{ac}g_{b[c}R_{d]a} = g^{ac}g_{bc}R_{da} - g^{ac}g_{bd}R_{ca} = \delta^a_b R_{da} - g_{bd}g^{ca}R_{ca} = R_{db} - Rg_{bd}, \tag{6.106}$$
$$2g^{ac}Rg_{a[c}g_{d]b} = g^{ac}Rg_{ac}g_{db} - g^{ac}Rg_{ad}g_{cb} = R\delta^a_a g_{db} - R\delta^c_d g_{cb} = (n-1)\,Rg_{db}. \tag{6.107}$$

Substituting 6.105, 6.106, 6.107 into 6.104 and using the symmetry of $R_{\mu\nu}$ and $g_{\mu\nu}$, we get

$$g^{ac}C_{abcd} = R_{bd} - \frac{1}{n-2}\left(nR_{db} - R_{db} - R_{db} + Rg_{bd}\right) + \frac{1}{(n-1)\,(n-2)}\,(n-1)\,Rg_{db}$$
$$= R_{bd} - R_{db} - \frac{Rg_{bd}}{n-2} + \frac{Rg_{db}}{n-2}$$
$$= 0. \tag{6.108}$$

■

§6.7 Curvature of 2-Sphere

In this section, we shall compute the Riemann curvature tensor $R^a{}_{bcd}$, Ricci curvature tensor R_{ab} and the Ricci scalar R for the 2-sphere S^2. We have seen in Example 4.5 that the round unit metric on S^2 is given by

$$g = \mathrm{d}\theta\otimes\mathrm{d}\theta + \sin^2\theta\mathrm{d}\phi\otimes\mathrm{d}\phi. \tag{6.109}$$

Let $x^1 = \theta$ and $x^2 = \phi$. Then

$$g_{\mu\nu} = \begin{cases} 1 & \text{if } \mu=\nu=1 \\ \sin^2\theta & \text{if } \mu=\nu=2\,. \\ 0 & \text{otherwise} \end{cases} \tag{6.110}$$

Then $g^{\mu\nu}$ is

$$g^{\mu\nu} = \begin{cases} 1 & \text{if } \mu = \nu = 1 \\ \frac{1}{\sin^2\theta} & \text{if } \mu = \nu = 2\,. \\ 0 & \text{otherwise} \end{cases} \tag{6.111}$$

Now, $g_{22,1} = \sin(2\theta)$, and the rest of the $g_{\mu\nu,\rho}$ are 0. Therefore, the Christoffel symbols are

$$\Gamma^{\mu}_{\nu\rho} = \frac{1}{2}g^{\mu\sigma}\left(g_{\nu\sigma,\rho} + g_{\sigma\rho,\nu} - g_{\rho\nu,\sigma}\right) = \frac{1}{2}g^{\mu\mu}\left(g_{\nu\mu,\rho} + g_{\mu\rho,\nu} - g_{\rho\nu,\mu}\right) \text{ (no summation)}\,, \tag{6.112}$$

since g is diagonalized. If $\mu = 1$, the only nonzero contribution comes from $g_{22,1}$ when $\nu = \rho = 2$. Therefore,

$$\Gamma^{1}_{22} = -\frac{1}{2}g^{11}g_{22,1} = -\sin\theta\cos\theta, \tag{6.113}$$

and the other $\Gamma^{1}_{\nu\rho}$ are all 0. If $\mu = 2$, the only nonzero contribution comes from $g_{22,1}$ when $\{\nu, \rho\} = \{1, 2\}$. Therefore,

$$\Gamma^{2}_{21} = \Gamma^{2}_{12} = \frac{1}{2}g^{22}\left(g_{12,2} + g_{22,1} - g_{21,2}\right) = \frac{1}{2\sin^2\theta}\sin(2\theta) = \cot\theta, \tag{6.114}$$

and the other $\Gamma^{2}_{\nu\rho}$ are all 0. Now,

$$\frac{\partial\Gamma^{1}_{22}}{\partial x^1} = -\cos^2\theta + \sin^2\theta = -\cos(2\theta)\,, \frac{\partial\Gamma^{2}_{12}}{\partial x^1} = \frac{\partial\Gamma^{2}_{21}}{\partial x^1} = -\csc^2\theta. \tag{6.115}$$

Now, ${R^1}_{212}$ is (using 6.11),

$$\begin{aligned} {R^1}_{212} &= \frac{\partial\Gamma^{1}_{22}}{\partial x^1} - \frac{\partial\Gamma^{1}_{21}}{\partial x^2} + \Gamma^{\tau}_{22}\Gamma^{1}_{\tau1} - \Gamma^{\tau}_{21}\Gamma^{1}_{\tau2} = -\cos(2\theta) - \Gamma^{2}_{21}\Gamma^{1}_{22} \\ &= \sin^2\theta - \cos^2\theta + \cot\theta\sin\theta\cos\theta = \sin^2\theta. \end{aligned} \tag{6.116}$$

So $R_{1212} = g_{\mu1}{R^\mu}_{212} = g_{11}{R^1}_{212} = \sin^2\theta$. Using $R_{abcd} = -R_{bacd}$, and $R_{abcd} = R_{cdab}$, we get

$$R_{11\mu\nu} = R_{22\mu\nu} = R_{\mu\nu11} = R_{\mu\nu22} = 0. \tag{6.117}$$

The remaining ones are $R_{1221}, R_{2121}, R_{2112}$.

$$R_{1221} = R_{2112} = -R_{1212} = -\sin^2\theta\,, \quad \text{and } R_{2121} = -R_{1221} = \sin^2\theta. \tag{6.118}$$

So we have computed all the R_{abcd}. ${R^a}_{bcd} = g^a e R_{ebcd} = g^{aa}R_{abcd}$. Therefore, the remaining nonzero components of ${R^a}_{bcd}$ are

$${R^1}_{221} = g^{11}R_{1221} = -\sin^2\theta\,, \tag{6.119}$$

$${R^2}_{112} = g^{22}R_{2112} = \frac{1}{\sin^2\theta}\left(-\sin^2\theta\right) = -1\,, \tag{6.120}$$

$${R^2}_{121} = g^{22}R_{2121} = \frac{1}{\sin^2\theta}\sin^2\theta = 1. \tag{6.121}$$

Now, the Ricci curvature tensor R_{ab} is $R_{ab} = g_{cd}R_{cadb} = g^{cc}R_{cacb}$ (summation over c). Therefore,

$$R_{12} = R_{21} = g^{cc}R_{c1c2} = g^{11}R_{1112} + g^{22}R_{2122} = 0. \tag{6.122}$$

The nonzero components of R_{ab} are

$$R_{11} = g^{cc}R_{c1c1} = g^{11}R_{1111} + g^{22}R_{2121} = \frac{1}{\sin^2\theta}\sin^2\theta = 1\,, \tag{6.123}$$

$$R_{22} = g^{cc}R_{c2c2} = g^{11}R_{1212} + g^{22}R_{2222} = \sin^2\theta. \tag{6.124}$$

Therefore, the Ricci curvature is

$$R = g^{ab}R_{ab} = g^{11}R_{11} + g^{22}R_{22} = 2. \tag{6.125}$$

In a similar manner, one can show that the Ricci scalar for a sphere of radius r is $\frac{2}{r}$.

7 Einstein Equation and Linearized Theory

Einstein equation related spacetime curvature via Einstein tensor with matter distribution (via energy-momentum tensor). The Einstein equation is

$$G_{ab} = 8\pi G T_{ab}, \tag{7.1}$$

where G is Newton's constant, and T_{ab} is energy-momentum stress tensor which is a symmetric tensor. 7.1 can equivalently be written as

$$R_{ab} - \frac{1}{2} R g_{ab} = 8\pi G T_{ab}. \tag{7.2}$$

The Einstein equation is one of the postulates of general relativity. Note that, from Einsten equation and Contracted Bianchi Identity, it follows that the energy momentum tensor T_{ab} is conserved, i.e. $\nabla^a T_{ab} = 0$. Now, it's a natural question to ask how unique Einstein equation is. Is there any other tensor than G_{ab} that we could have put on the LHS of 7.1? Lovelock's theorem answers this question.

Theorem 7.1 (Lovelock, 1972)
Let H_{ab} be a symmetric tensor such that

(i) in any coordinate chart, at any point, $H_{\mu\nu}$ is a function of $g_{\mu\nu}$, $g_{\mu\nu,\rho}$ and $g_{\mu\nu,\rho\sigma}$ at that point;

(ii) $\nabla^a H_{ab} = 0$;

(iii) either spacetime is 4-dimensional or $H_{\mu\nu}$ depends linearly on $g_{\mu\nu,\rho\sigma}$.

Then there exist constants α and β such that

$$H_{ab} = \alpha G_{ab} + \beta g_{ab}. \tag{7.3}$$

Hence, Einstein realized there is a freedom to add a constant multiple of g_{ab} to the LHS of 7.1.

$$G_{ab} + \Lambda g_{ab} = 8\pi G T_{ab}. \tag{7.4}$$

This Λ is called the **cosmological constant**. We can rewrite 7.4 as

$$G_{ab} = 8\pi G \left(T_{ab} - \frac{\Lambda}{8\pi G} g_{ab} \right) = 8\pi G \left(T_{ab} - \rho_{\text{vac}} g_{ab} \right), \tag{7.5}$$

where $\rho_{\text{vac}} = \frac{\Lambda}{8\pi G}$ is the energy density of the vacuum. In vacuum, there is no matter, so T_{ab} is 0. Hence,

$$T^{(\text{vac})}_{ab} = -\rho_{\text{vac}} g_{ab}. \tag{7.6}$$

§7.1 Equivalence Principles

Incompatibility of Newtonian Gravity and Special Relativity

Newton's law of gravity is gven by the following Poisson's equation

$$\nabla^2 \phi = 4\pi G \rho, \tag{7.7}$$

where ϕ is the gravitational potential and ρ is the mass density.

Consider a sphere V of radius r. In Newtonian gravitational theory, $|\mathbf{g}| = \frac{GM}{r^2}$. Therefore,

$$\oiint_{\partial V} \mathbf{g} \cdot \mathrm{d}\mathbf{A} = \oiint_{\partial V} \frac{GM}{r^2} \left(-\mathbf{r}\right) \cdot \mathrm{d}A\mathbf{r} = -\oiint_{\partial V} \frac{GM}{r^2} \, \mathrm{d}A = -4\pi GM. \tag{7.8}$$

By divergence theorem,

$$\iiint_V \left(\boldsymbol{\nabla} \cdot \mathbf{g}\right) \, \mathrm{d}V = \oiint_{\partial V} \mathbf{g} \cdot \mathrm{d}\mathbf{A} = -4\pi GM = -4\pi G \iiint_V \rho \, \mathrm{d}V. \tag{7.9}$$

Therefore, $\boldsymbol{\nabla} \cdot \mathbf{g} = -4\pi G\rho$. Furthermore, $\mathbf{g} = -\boldsymbol{\nabla}\phi$. Hence, $\nabla^2\phi = 4\pi G\rho$.

Now, solution of the above Poisson equation (7.7) is given by

$$\phi\left(t, \mathbf{x}\right) = -G \int \mathrm{d}^3\mathbf{y} \, \frac{\rho\left(t, \mathbf{y}\right)}{|\mathbf{x} - \mathbf{y}|}. \tag{7.10}$$

Note the presence of the same t on both sides. The time at which one adds change to source distribution ρ is exactly the same time at which the gravitational potential responds to. These two events are simultaneous in one inertial frame, may not be simultaneous in other inertial reference frames. In fact, change in gravitational potential may precede the change in source distribution in some reference frames, i.e. the effect taking place before the cause violating principle of relativity. This is how Newtonian gravity is incompatible with special theory of relativity.

The incompatibility of Newtonian gravity with special theory of relativity is not a problem if the objects under consideration are moving non-relativistically (with speeds much less than the speed of light). For instance, in our solar system, Newtonian theory is very accurate.

Newtonian theory also breaks down when the gravitational field is very strong. Consider a partivle of mass m moving rouns a spherical body of mass M in a circular orbit of radius r. Then

$$\phi = -\frac{GM}{r}. \tag{7.11}$$

Newton's law gives

$$\frac{v^2}{r} = \frac{GM}{r^2} \implies \frac{v^2}{c^2} = \frac{|\phi|}{c^2}. \tag{7.12}$$

Newtonian theory requires non-relativistic motion, which is the case only when the gravitational field is weak: $\frac{|\phi|}{c^2} \ll 1$. In our solar syetem, $\frac{|\phi|}{c^2} < 10^{-5}$. GR is the theory that replaces both Newtonian gravity and special relativity.

The earliest form of all equivalence principles is called the **weak equivalence principle**, which dates from Galileo and Newton. It states that the inertial mass and gravitational mass of any object are equal.

Newton's second law relates the force exerted on an object to the acceleration it undergoes, setting them proportional to eacgh other with the proportionality constant being the inertial mass m_I:

$$\mathbf{F} = m_I\mathbf{a}. \tag{7.13}$$

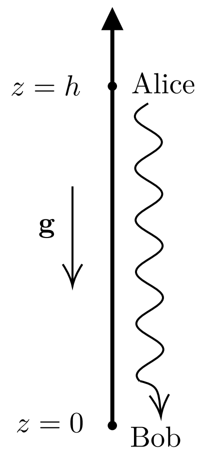
The inertial mass has a universal characteristic. It is the same constant no matter what kind of force is being exerted. The gravitational field $\mathbf{g}$, on the other hand the hand, is the negative gradient of gravitational potential ϕ, $\mathbf{g} = -\boldsymbol{\nabla}\phi$. Gravitational force F_g is proportional to this gravitational fiels with the proportionality constant being called the gravitational mass m_G.

$$\mathbf{F}_g = m_G\mathbf{g} = -m_G\boldsymbol{\nabla}\phi. \tag{7.14}$$

m_G has a very different character than m_I. It is a quantity specific to gravitational force. m_G is the "gravitational charge" of the body. However, the response of matter to gravitation is universal: objects of different composition falls from a given height at the same rate.

§7.2 Gravitational Red Shift

Consider the following scenario: Alice and Bob are at rest in a uniform gravitational field of strength g in the negative z direction. Alice is at $z = h$ and Bob is at $z = 0$. They have identical clocks. Alice sends light signals to Bob at constant proper time intervals which she measures to be $\Delta\tau_A$. What is the proper time interval between the signals received by Bob?



Alice and Bob have acceleration g with respect to a freely falling frame. We choose our freely falling frame so that Alice and Bob are at rest at $t = 0$. We shall neglect spacial relativistic contributions (i.e. $\frac{v}{c} \ll 1$). Then the position of Alice and Bob at time t are

$$z_A(t) = h + \frac{1}{2}gt^2, \quad \text{and } z_B(t) = \frac{1}{2}gt^2. \tag{7.15}$$

Alice and Bob have $v(t) = gt$, which we assume to be much smaller than c over the time it takes to perform the experiment. Hence, we shall neglect effects of order $\frac{g^2t^2}{c^2}$.

Suppose Alice emits the first signal at $t = t_1$. Then its trajectory is

$$z_A(t_1) - c(t - t_1) = h + \frac{1}{2}gt_1^2 - c(t - t_1). \tag{7.16}$$

It reaches Bob at $t = T_1$, when the above is equal to $z_B(T_1)$. So

$$h + \frac{1}{2}gt_1^2 - c(T_1 - t_1) = \frac{1}{2}gT_1^2. \tag{7.17}$$

The second light signal is emitted at $t = t_1 + \Delta\tau_A$, and it reaches Bob at $t = T_1 + \Delta\tau_B$. Therefore,

$$h + \frac{1}{2}g(t_1 + \Delta\tau_A)^2 - c(T_1 + \Delta\tau_B - t_1 - \Delta\tau_A) = \frac{1}{2}g(T_1 + \Delta\tau_B)^2. \tag{7.18}$$

7.18−7.17 gives us

$$c(\Delta\tau_A - \Delta\tau_B) + \frac{1}{2}g\Delta\tau_A(2t_1 + \Delta\tau_A) = \frac{1}{2}g\Delta\tau_B(2T_1 + \Delta\tau_B). \tag{7.19}$$

The terms quadratic in $\Delta\tau_A$ and $\Delta\tau_B$ are negligible as $g\Delta\tau_A \ll c$. If this does not hold, then Alice would reach relativistic speeds by the time she emitted the second signal. Similarly for $\Delta\tau_B$. Therefore,

$$\begin{aligned}
& c(\Delta\tau_A - \Delta\tau_B) + g\Delta\tau_A t_1 = g\Delta\tau_B T_1 \\
\implies & (gT_1 + c)\Delta\tau_B = (gt_1 + c)\Delta\tau_A \\
\implies & \Delta\tau_B = (gT_1 + c)^{-1}(gt_1 + c)\Delta\tau_A \\
\implies & \Delta\tau_B = \left(1 + \frac{gt_1}{c}\right)\left(1 + \frac{gT_1}{c}\right)^{-1}\Delta\tau_A \\
\implies & \Delta\tau_B \approx \left(1 + \frac{gt_1}{c}\right)\left(1 - \frac{gT_1}{c}\right)\Delta\tau_A \\
\implies & \Delta\tau_B \approx \left(1 + \frac{gt_1}{c} - \frac{gT_1}{c}\right)\Delta\tau_A.
\end{aligned} \tag{7.20}$$

$T_1 - t_1$ is the time light takes to travel the distance h between Alice and Bob. Therefore, $\frac{h}{c} = T_1 - t_1$. So

$$\Delta\tau_B \approx \left(1 - \frac{gh}{c^2}\right)\Delta\tau_A. \tag{7.21}$$

Therefore, the time interval between the light signals received by Bob is less than the time interval between the light signals emitted by Alice. In other words, Bob's clock appears to run slow compared to Alice's.

If $\Delta\tau_A$ is the period of light waves sent by Alice, then $\Delta\tau_A = \frac{\lambda_A}{c}$. Then 7.21 reduces to

$$\lambda_B \approx \left(1 - \frac{gh}{c^2}\right)\lambda_A. \tag{7.22}$$

Therefore, the light received by Bob is shorter in wavelength and hence is blue-shifted. In other words, light falling in a gravitational field is blue-shifted. An identical argument reveals that light falling out of a gravitational field is red-shifted.

§7.3 Linearized Theory

The Einstein equation $G_{ab} = 8\pi G T_{ab}$ is nonlinear. However, when gravity is weak, we consider spacetime as a perturbation of Minkowski spacetime. We assume our spacetime manifold is $M = \mathbb{R}^4$, and that there exists globally defined "almost inertial" coordinates x^μ for which the matric reads

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{7.23}$$

where $\eta_{\mu\nu} = \text{diag}\,(-1,1,1,1)$. The weakness of the gravitational field corresponds to components of $h_{\mu\nu}$ being small compared to 1. Notw that g_{ab} is the physical metric, i.e. free particles move along geodesics of g_{ab}. On the other hand, $h_{\mu\nu}$ are the components of a tensor field that transforms accordingly under Lorentz transformations of the coordinates x^μ.[1]

Let $\kappa^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}$, where $h^{\mu\nu} = \eta^{\mu\rho}\eta^{\nu\sigma}h_{\rho\sigma}$. Then

$$\begin{aligned}
\kappa^{\mu\nu} g_{\mu\tau} &= \left(\eta^{\mu\nu} - \eta^{\mu\rho}\eta^{\nu\sigma}h_{\rho\sigma}\right)\left(\eta_{\mu\tau} + h_{\mu\tau}\right) \\
&= \eta^{\mu\nu}\eta_{\mu\tau} + \eta^{\mu\nu}h_{\mu\tau} - \eta_{\mu\tau}\eta^{\mu\rho}\eta^{\nu\sigma}h_{\rho\sigma} - \eta^{\mu\rho}\eta^{\nu\sigma}\underbrace{h_{\rho\sigma}h_{\mu\tau}}_{\text{higher order}\to 0} \\
&= \delta^\nu{}_\tau + \eta^{\mu\nu}h_{\mu\tau} - \delta^\rho{}_\tau\eta^{\nu\sigma}h_{\rho\sigma} \\
&= \delta^\nu{}_\tau + \eta^{\mu\nu}h_{\mu\tau} - \eta^{\sigma\nu}h_{\sigma\tau} \\
&= \delta^\nu{}_\tau.
\end{aligned} \tag{7.24}$$

Therefore, $g^{\mu\nu} = \kappa^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}$ up to linear order in the perturbation $h_{\mu\nu}$. Now we shall determine the Einstein equation to first order in the perturbation $h_{\mu\nu}$.

To first order, the Christoffel symbols are

$$\Gamma^\nu_{\nu\rho} = \frac{1}{2}\eta^{\mu\sigma}\left(h_{\sigma\nu,\rho} + h_{\sigma\rho,\nu} - h_{\nu\rho,\sigma}\right). \tag{7.25}$$

The Riemann tensor is (neglecting $\Gamma\Gamma$ terms since they are second order in the perturbation)

$$R_{\mu\nu\rho\sigma} = \eta_{\mu\tau}\left(\partial_\rho\Gamma^\tau_{\nu\sigma} - \partial_\sigma\Gamma^\tau_{\nu\rho}\right). \tag{7.26}$$

Using 7.25, we get $\partial_\rho\Gamma^\tau_{\nu\sigma} = \frac{1}{2}\eta^{\tau\kappa}\left(h_{\kappa\nu,\sigma\rho} + h_{\kappa\sigma,\nu\rho} - h_{\nu\sigma,\kappa\rho}\right)$, and $\partial_\sigma\Gamma^\tau_{\nu\rho} = \frac{1}{2}\eta^{\tau\kappa}\left(h_{\kappa\nu,\rho\sigma} + h_{\kappa\rho,\nu\sigma} - h_{\nu\rho,\kappa\sigma}\right)$.

[1] **A convenient fiction:** We can think of a slightly curved spacetime as a flat spacetime with a tensor $h_{\mu\nu}$ defined on it. Then all physical physical fields, such as $R_{\mu\nu\rho\sigma}$ will be defined in terms of $h_{\mu\nu}$.

Hence,

$$\begin{aligned} R_{\mu\nu\rho\sigma} &= \frac{1}{2}\eta^{\tau\kappa}\eta_{\mu\tau}\left(h_{\kappa\sigma,\nu\rho} - h_{\nu\sigma,\kappa\rho} - h_{\kappa\rho,\nu\sigma} + h_{\nu\rho,\kappa\sigma}\right) \\ &= \frac{1}{2}\delta^{\kappa}{}_{\mu}\left(h_{\kappa\sigma,\nu\rho} - h_{\nu\sigma,\kappa\rho} - h_{\kappa\rho,\nu\sigma} + h_{\nu\rho,\kappa\sigma}\right) \\ &= \frac{1}{2}h_{\mu\sigma,\nu\rho} - h_{\nu\sigma,\mu\rho} - h_{\mu\rho,\nu\sigma} + h_{\nu\rho,\mu\sigma} \\ &= \frac{1}{2}\left(\partial_\rho\partial_\nu h_{\mu\sigma} + \partial_\sigma\partial_\mu h_{\nu\rho} - \partial_\rho\partial_\mu h_{\nu\sigma} - \partial_\sigma\partial_\nu h_{\mu\rho}\right). \end{aligned} \tag{7.27}$$

$$\begin{aligned} R^{\kappa}{}_{\nu\rho\sigma} &= \eta^{\kappa\mu}R_{\mu\nu\rho\sigma} = \frac{1}{2}\eta^{\kappa\mu}\left(\partial_\rho\partial_\nu h_{\mu\sigma} + \partial_\sigma\partial_\mu h_{\nu\rho} - \partial_\rho\partial_\mu h_{\nu\sigma} - \partial_\sigma\partial_\nu h_{\mu\rho}\right) \\ &= \frac{1}{2}\left(\partial_\rho\partial_\nu h^{\kappa}{}_{\sigma} + \partial_\sigma\partial^{\kappa} h_{\nu\rho} - \partial_\rho\partial^{\kappa} h_{\nu\sigma} - \partial_\sigma\partial_\nu h^{\kappa}{}_{\rho}\right). \end{aligned} \tag{7.28}$$

$$R_{\nu\sigma} = R^{\rho}{}_{\nu\rho\sigma} = \frac{1}{2}\left(\partial_\rho\partial_\nu h^{\rho}{}_{\sigma} + \partial_\sigma\partial^{\rho} h_{\nu\rho} - \partial_\rho\partial^{\rho} h_{\nu\sigma} - \partial_\sigma\partial_\nu h^{\rho}{}_{\rho}\right). \tag{7.29}$$

$\partial_\rho\partial_\nu h^{\rho}{}_{\sigma} = \partial_\rho\partial_\nu\eta^{\alpha\rho}h_{\alpha\sigma} = \eta^{\alpha\rho}\partial_\rho\partial_\nu h_{\alpha\sigma} = \partial^\rho\partial_\nu h_{\rho\sigma}$. Therefore,

$$\begin{aligned} R_{\nu\sigma} &= \frac{1}{2}\partial^\rho\partial_\nu h_{\sigma\rho} + \frac{1}{2}\partial^\rho\partial_\sigma h_{\nu\rho} - \frac{1}{2}\partial_\rho\partial^\rho h_{\nu\sigma} - \frac{1}{2}\partial_\sigma\partial_\nu h \\ &= \partial^\rho\partial_{(\nu}h_{\sigma)\rho} - \frac{1}{2}\partial^\rho\partial_\rho h_{\nu\sigma} - \frac{1}{2}\partial_\sigma\partial_\nu h. \end{aligned}$$

$$\therefore R_{\mu\nu} = \partial^\rho\partial_{(\mu}h_{\nu)\rho} - \frac{1}{2}\partial_\rho\partial^\rho h_{\mu\nu} - \frac{1}{2}\partial_\mu\partial_\nu h. \tag{7.30}$$

Neglecting the non-linear terms, the Ricci scalar (to the first order) is

$$\begin{aligned} R = R_{\mu\nu}\eta^{\mu\nu} &= \frac{1}{2}\eta^{\mu\nu}\partial^\rho\partial_\mu h_{\nu\rho} + \frac{1}{2}\eta^{\mu\nu}\partial^\rho\partial_\nu h_{\mu\rho} - \frac{1}{2}\eta^{\mu\nu}\partial_\rho\partial^\rho h_{\mu\nu} - \frac{1}{2}\eta^{\mu\nu}\partial_\mu\partial_\nu h \\ &= \frac{1}{2}\partial^\rho\partial^\nu h_{\nu\rho} + \frac{1}{2}\partial^\rho\partial^\mu h_{\mu\rho} - \frac{1}{2}\partial^\rho\partial_\rho h - \frac{1}{2}\partial_\nu\partial_\nu h \\ &= \partial^\rho\partial^\nu h_{\rho\nu} - \partial^\rho\partial_\rho h. \end{aligned} \tag{7.31}$$

The Einstein tensor is $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}R\eta_{\mu\nu}$ (again, in the first order). Einstein equation equates it to $8\pi T_{\mu\nu}$ (by choosing the unit $G = c = 1$). Let us define a new quantity

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}h\eta_{\mu\nu}. \tag{7.32}$$

Then we have

$$\bar{h} = \eta^{\mu\nu}\bar{h}_{\mu\nu} = \eta^{\mu\nu}h_{\mu\nu} - \frac{1}{2}h\eta^{\mu\nu}\eta_{\mu\nu} = h - \frac{1}{2}h\delta^{\mu}_{\mu} = -h. \tag{7.33}$$

This gives us

$$h_{\mu\nu} = \bar{h}_{\mu\nu} + \frac{1}{2}h\eta_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2}\bar{h}\eta_{\mu\nu}. \tag{7.34}$$

In this new variable, the expression for $G_{\mu\nu}$ is

$$\begin{aligned} G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}R\eta_{\mu\nu} &= \partial^\rho\partial_{(\mu}h_{\nu)\rho} - \frac{1}{2}\partial^\rho\partial_\rho h_{\mu\nu} - \frac{1}{2}\partial_\mu\partial_\nu h - \frac{1}{2}\eta_{\mu\nu}\left(\partial^\rho\partial^\sigma h_{\rho\sigma} - \partial^\rho\partial_\rho h\right) \\ &= \frac{1}{2}\partial^\rho\partial_\mu\left(\bar{h}_{\nu\rho} - \frac{1}{2}\bar{h}\eta_{\nu\rho}\right) + \frac{1}{2}\partial^\rho\partial_\nu\left(\bar{h}_{\mu\rho} - \frac{1}{2}\bar{h}\eta_{\mu\rho}\right) - \frac{1}{2}\partial^\rho\partial_\rho\left(\bar{h}_{\mu\nu} - \frac{1}{2}\bar{h}\eta_{\mu\nu}\right) \\ &\quad + \frac{1}{2}\partial_\mu\partial_\nu\bar{h} - \frac{1}{2}\eta_{\mu\nu}\partial^\rho\partial^\sigma\left(\bar{h}_{\rho\sigma} - \frac{1}{2}\bar{h}\eta_{\rho\sigma}\right) - \frac{1}{2}\eta_{\mu\nu}\partial^\rho\partial_\rho\bar{h} \\ &= \frac{1}{2}\partial^\rho\partial_\mu\bar{h}_{\nu\rho} - \frac{1}{4}\partial_\nu\partial_\mu\bar{h} + \frac{1}{2}\partial^\rho\partial_\nu\bar{h}_{\mu\rho} - \frac{1}{4}\partial_\mu\partial_\nu\bar{h} - \frac{1}{2}\partial^\rho\partial_\rho\bar{h}_{\mu\nu} + \frac{1}{4}\eta_{\mu\nu}\partial^\rho\partial_\rho\bar{h} \\ &\quad + \frac{1}{2}\partial_\mu\partial_\nu\bar{h} - \frac{1}{2}\eta_{\mu\nu}\partial^\rho\partial^\sigma\bar{h}_{\rho\sigma} + \frac{1}{4}\eta_{\mu\nu}\eta_{\rho\sigma}\partial^\rho\partial_\rho\bar{h} - \frac{1}{2}\eta_{\mu\nu}\partial^\rho\partial_\rho\bar{h} \\ &= \frac{1}{2}\partial^\rho\partial_\mu\bar{h}_{\nu\rho} + \frac{1}{2}\partial^\rho\partial_\nu\bar{h}_{\mu\rho} - \frac{1}{2}\partial^\rho\partial_\rho\bar{h}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\partial^\rho\partial^\sigma\bar{h}_{\rho\sigma} \\ &= -\frac{1}{2}\partial^\rho\partial_\rho\bar{h}_{\mu\nu} + \partial^\rho\partial_{(\mu}\bar{h}_{\nu)\rho} - \frac{1}{2}\eta_{\mu\nu}\partial^\rho\partial^\sigma\bar{h}_{\rho\sigma}. \end{aligned} \tag{7.35}$$

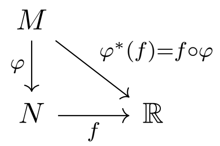
Therefore, the linearized Einstein equation is

$$-\frac{1}{2}\partial^\rho\partial_\rho\bar{h}_{\mu\nu} + \partial^\rho\partial_{(\mu}\bar{h}_{\nu)\rho} - \frac{1}{2}\eta_{\mu\nu}\partial^\rho\partial^\sigma\bar{h}_{\rho\sigma} = 8\pi T_{\mu\nu}. \tag{7.36}$$

8 Diffeomorphisms and Lie Derivative

§8.1 Pull Back and Push Forward

Definition 8.1. Let $\varphi : M \to N$ and $f : N \to \mathbb{R}$ be smooth maps. The **pull back** of f by φ is the map $\varphi^* (f) : M \to \mathbb{R}$ defined by $\varphi^* (f) = f \circ \varphi$. In other words, $\varphi^* : C^\infty (N) \to C^\infty (M)$ is the map given by $f \mapsto f \circ \varphi$.



Definition 8.2 (Push Forward of Tangent Vector)**.** Let $\varphi : M \to N$ be a smooth map. Let $p \in M$ and $X_p \in T_pM$. The **push forward** of X_p with respect to φ is the vector $\phi_* (X_p) \in T_{\varphi(p)}N$ given by

$$(\phi_* (X_p)) (f) = X_p (\varphi^* f) , \tag{8.1}$$

where $f \in C^\infty (N)$.

Definition 8.3 (Pull Back of Covector)**.** Let $\varphi : M \to N$ be a smooth map. Let $p \in M$ and $\eta_{\varphi(p)} \in T^*_{\varphi(p)}N$. The **pull back** of $\eta_{\varphi(p)}$ by φ is the covector $\varphi^* \left(\eta_{\varphi(p)}\right) \in T^*_p M$ given by

$$\varphi^* \left(\eta_{\varphi(p)}\right) (X_p) = \eta_{\varphi(p)} (\phi_* (X_p)) , \tag{8.2}$$

where $X_p \in T_pM$.

Definition 8.4 (Diffeomorphism)**.** A map $\varphi : M \to N$ is a **diffeomorphism** iff it is bijective, smooth and has a smooth inverse.

If $\varphi : M \to N$ is a diffeomorphism, then M and N have the same dimension. In fact, they have identical manifold structure. If (U, ψ) is a coordinate chart in M, then $\left(\varphi (U) , \psi \circ \varphi^{-1}\right)$ is a coordinate chart in N. With diffeomorphism, we can extend our definition of push forward to any type of tensor.

Definition 8.5 (Push Forward of a Tensor)**.** Let $\varphi : M \to N$ be a diffeomorphism and $T\big|_p$ a tensor of type (r, s) at $p \in M$. The **push forward** of $T\big|_p$ is a tensor $\varphi_* \left(T\big|_p\right)$ of type (r, s) at $\varphi (p) \in N$ defined by

$$\begin{aligned} &\varphi_* \left(T\big|_p\right) \left(\eta_1\big|_{\varphi(p)}, \ldots, \eta_r\big|_{\varphi(p)}, X_1\big|_{\varphi(p)}, \ldots, X_s\big|_{\varphi(p)}\right) \\ &= T\big|_p \left(\varphi^* \left(\eta_1\big|_{\varphi(p)}\right), \ldots, \varphi^* \left(\eta_r\big|_{\varphi(p)}\right), \left(\varphi^{-1}\right)_* X_1\big|_{\varphi(p)}, \ldots, \left(\varphi^{-1}\right)_* X_s\big|_{\varphi(p)}\right) . \end{aligned} \tag{8.3}$$

Remark 8.1. In GR, we descrive physics with a manifold M on which certain tensor fields, such as the metric g, the Maxwell field F etc. are defined. If $\varphi : M \to N$ is a diffeomorphism, then there is no way of distinguishing $(M, g, F, \ldots)$ from $(N, \varphi_* (g) , \varphi_* (F) , \ldots)$. They give equivalent description of physics. If we set $N = M$, this reveals that the collection of tensor fields $(\varphi_* (g) , \varphi_* (F) , \ldots)$ is physically indistinguishable from $(g, F, \ldots)$. It follows that diffeomorphism are **gauge symmetry** in GR.

Definition 8.6 (Symmetry Transformation). A diffeomorphism $\varphi : M \to M$ is a **symmetry transformation** of a tensor field T iff

$$\varphi_* \left(T\big|_p \right) = T\big|_{\varphi(p)} \qquad \forall\, p \in M. \tag{8.4}$$

A symmetry transformation of the metric tensor is called an **isometry**.

We want to compute $\varphi_* \left(T\big|_p \right)$ in a coordinate basis. Let $\psi_M = \left(x^1, \ldots, x^n\right)$ be coordinates around $p \in M$, and $\psi_N = \left(y^1, \ldots, y^n\right)$ coordinates around $\varphi(p) \in N$. Now,

$$\left(\varphi_* T\big|_p\right)^{\mu_1 \cdots \mu_r}{}_{\nu_1 \cdots \nu_s} = \left(\varphi_* T\big|_p\right) \left(\mathrm{d}y^{\mu_1}\big|_{\varphi(p)}, \ldots, \mathrm{d}y^{\mu_r}\big|_{\varphi(p)}, \frac{\partial}{\partial y^{\nu_1}}\bigg|_{\varphi(p)}, \ldots, \frac{\partial}{\partial y^{\nu_s}}\bigg|_{\varphi(p)} \right). \tag{8.5}$$

So we need to compute how the basis covectors and basis tangent vectors transform with pull back by φ and push forward by φ^{-1}, respectively.

$$\left(\varphi^{-1}\right)_* \left(\frac{\partial}{\partial y^{\nu_i}}\bigg|_{\varphi(p)} \right)(f) = \frac{\partial}{\partial y^{\nu_i}}\bigg|_{\varphi(p)} \left(\left(\varphi^{-1}\right)^* f \right) = \frac{\partial}{\partial y^{\nu_i}}\bigg|_{\varphi(p)} \left(f \circ \varphi^{-1} \right), \tag{8.6}$$

for $f \in C^\infty(M)$. Using chain rule,

$$\begin{aligned}
\frac{\partial}{\partial y^{\nu_i}}\bigg|_{\varphi(p)} \left(f \circ \varphi^{-1} \right) &= \frac{\partial \left(f \circ \psi_M^{-1} \circ \psi_M \circ \varphi^{-1} \circ \psi_N^{-1} \right)}{\partial r^{\nu_i}}\bigg|_{\psi_N(\varphi(p))} \\
&= \frac{\partial \left(\psi_M \circ \varphi^{-1} \circ \psi_N^{-1} \right)^i}{\partial r^{\nu_i}}\bigg|_{\psi_N(\varphi(p))} \frac{\partial \left(f \circ \psi_M^{-1} \right)}{\partial r^i}\bigg|_{\psi_M(p)} \\
&= \frac{\partial \left(x^i \circ \varphi^{-1} \right)}{\partial y^{\nu_i}} \frac{\partial f}{\partial x^i}.
\end{aligned} \tag{8.7}$$

We shall write $\frac{\partial \left(x^i \circ \varphi^{-1} \right)}{\partial y^{\nu_i}}$ as $\frac{\partial x^i}{\partial y^{\nu_i}}$. So

$$\begin{aligned}
\frac{\partial}{\partial y^{\nu_i}}\bigg|_{\varphi(p)} \left(f \circ \varphi^{-1} \right) &= \frac{\partial x^i}{\partial y^{\nu_i}}\bigg|_{\varphi(p)} \frac{\partial f}{\partial x^i}\bigg|_p \\
&= \frac{\partial x^{\sigma_i}}{\partial y^{\nu_i}} \left(\varphi(p) \right) \frac{\partial}{\partial x^{\sigma_i}}\bigg|_p f \\
\therefore \left(\varphi^{-1}\right)_* \left(\frac{\partial}{\partial y^{\nu_i}}\bigg|_{\varphi(p)} \right) &= \frac{\partial x^{\sigma_i}}{\partial y^{\nu_i}} \left(\varphi(p) \right) \frac{\partial}{\partial x^{\sigma_i}}\bigg|_p .
\end{aligned} \tag{8.8}$$

Now we shall compute $\varphi^* \left(\mathrm{d}y^{\mu_i}\big|_{\varphi(p)} \right)$. For any $X_p \in T_pM$,

$$\varphi^* \left(\mathrm{d}y^{\mu_i}\big|_{\varphi(p)} \right) X_p = \mathrm{d}y^{\mu_i}\big|_{\varphi(p)} \left(\varphi_* X_p \right). \tag{8.9}$$

Using an analogous manner as 8.8, one can show that

$$\varphi_* \left(\frac{\partial}{\partial x^i}\bigg|_p \right) = \frac{\partial y^j}{\partial x^i}(p) \frac{\partial}{\partial y^j}\bigg|_{\varphi(p)}, \tag{8.10}$$

where $\frac{\partial y^j}{\partial x^i}$ is a shorthand for $\frac{\partial \left(y^j \circ \varphi \right)}{\partial x^i}$. Therefore,

$$\varphi_* X_p = \varphi_* \left(X^i \frac{\partial}{\partial x^i}\bigg|_p \right) = X^i \frac{\partial y^j}{\partial x^i}(p) \frac{\partial}{\partial y^j}\bigg|_{\varphi(p)}. \tag{8.11}$$

Applying $\mathrm{d}y^{\mu_i}\big|_{\varphi(p)}$ on 8.11, one obtains

$$\begin{aligned}
\mathrm{d}y^{\mu_i}\big|_{\varphi(p)}(\varphi_* X_p) &= \mathrm{d}y^{\mu_i}\big|_{\varphi(p)}\left(X^i \frac{\partial y^j}{\partial x^i}(p)\,\frac{\partial}{\partial y^j}\bigg|_{\varphi(p)}\right) = X^i \frac{\partial y^j}{\partial x^i}(p)\,\delta^{\mu_i}_j \\
&= \frac{\partial y^{\mu_i}}{\partial x^i}(p)\,X^{\mu}\delta^i_{\mu} = \frac{\partial y^{\mu_i}}{\partial x^i}(p)\,\mathrm{d}x^i\big|_p\left(X^{\mu}\,\frac{\partial}{\partial x^{\mu}}\bigg|_p\right) \\
&= \frac{\partial y^{\mu_i}}{\partial x^{\rho_i}}(p)\,\mathrm{d}x^{\rho_i}\big|_p(X_p) \\
\therefore \varphi^*\left(\mathrm{d}y^{\mu_i}\big|_{\varphi(p)}\right) &= \frac{\partial y^{\mu_i}}{\partial x^{\rho_i}}(p)\,\mathrm{d}x^{\rho_i}\big|_p. \qquad (8.12)
\end{aligned}$$

Combining 8.3, 8.5, 8.8, 8.12, one obtains

$$\begin{aligned}
\left(\varphi_* T\big|_p\right)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} &= T\big|_p\left(\frac{\partial y^{\mu_1}}{\partial x^{\rho_1}}(p)\,\mathrm{d}x^{\rho_1}\big|_p, \ldots, \frac{\partial y^{\mu_r}}{\partial x^{\rho_r}}(p)\,\mathrm{d}x^{\rho_r}\big|_p,\right. \\
&\qquad\left.\frac{\partial x^{\sigma_1}}{\partial y^{\nu_1}}(\varphi(p))\,\frac{\partial}{\partial x^{\sigma_1}}\bigg|_p, \ldots, \frac{\partial x^{\sigma_s}}{\partial y^{\nu_s}}(\varphi(p))\,\frac{\partial}{\partial x^{\sigma_s}}\bigg|_p\right) \\
&= \prod_{i=1}^{r}\frac{\partial y^{\mu_i}}{\partial x^{\rho_i}}(p)\,\prod_{j=1}^{s}\frac{\partial x^{\sigma_j}}{\partial y^{\nu_j}}(\varphi(p))\;T\big|_p\left(\mathrm{d}x^{\rho_1}\big|_p, \ldots, \mathrm{d}x^{\rho_r}\big|_p, \frac{\partial}{\partial x^{\sigma_1}}\bigg|_p, \ldots, \frac{\partial}{\partial x^{\sigma_s}}\bigg|_p\right) \\
&= \prod_{i=1}^{r}\frac{\partial y^{\mu_i}}{\partial x^{\rho_i}}(p)\,\prod_{j=1}^{s}\frac{\partial x^{\sigma_j}}{\partial y^{\nu_j}}(\varphi(p))\left(T\big|_p\right)^{\rho_1\cdots\rho_r}{}_{\sigma_1\cdots\sigma_s}. \qquad (8.13)
\end{aligned}$$

§8.2 Lie Derivative

Let $X \in \mathfrak{X}(M)$, and $p \in M$. Suppose γ is the integral curve of X going through p. WLOG, $\gamma(0) = p$. Then let φ_t be the map that sends p to the point parameter distance t along γ, i.e. $\varphi_t(p) = \gamma(t)$. This might be defined for only small t.

It can be shown that φ_t is a diffeomorphism. Note that φ_0 is the identity map, and $\varphi_s \circ \varphi_t = \varphi_{s+t}$. Hence, $\varphi_t^{-1} = \varphi_{-t}$. Therefore, if φ_t is defined for all $t \in \mathbb{R}$, the diffeomorphisms φ_t form a 1-parameter abelian group, with the group operation being composition.

Definition 8.7 (Lie Derivative)**.** The **Lie derivative** of a tensor field T with respect to a vector field X at $p \in M$ is

$$(\mathcal{L}_X T)_p = \lim_{t\to 0}\frac{(\varphi_{-t})_* T\big|_{\varphi_t(p)} - T\big|_p}{t}. \qquad (8.14)$$

One can easily verify that Lie derivative is, indeed, a derivation. First, let's verify that $\mathcal{L}_X$ is linear. If S and T are (r, s) tensor fields,

$$(\mathcal{L}_X(\alpha S + \beta T))_p = \lim_{t\to 0}\frac{(\phi_{-t})_*\left(\alpha S\big|_{\phi_t(p)} + \beta T\big|_{\phi_t(p)}\right) - \left(\alpha S\big|_p + \beta T\big|_p\right)}{t}. \qquad (8.15)$$

$(\phi_{-t})_*$ is linear, so

$$\begin{aligned}
(\mathcal{L}_X(\alpha S + \beta T))_p &= \lim_{t\to 0}\frac{\alpha(\phi_{-t})_* S\big|_{\phi_t(p)} + \beta(\phi_{-t})_* T\big|_{\phi_t(p)} - \alpha S\big|_p - \beta T\big|_p}{t} \\
&= \alpha\lim_{t\to 0}\frac{(\phi_{-t})_* S\big|_{\phi_t(p)} - S\big|_p}{t} + \beta\lim_{t\to 0}\frac{(\phi_{-t})_* T\big|_{\phi_t(p)} - T\big|_p}{t} \\
&= \alpha\,(\mathcal{L}_X S)_p + \beta\,(\mathcal{L}_X T)_p. \qquad (8.16)
\end{aligned}$$

So, linearity is verified. Now, in order to show that $\mathcal{L}_X$ is indeed a derivation, we need to verify the Leibniz rule, i.e.

$$\mathcal{L}_X (S \otimes T)_p = \mathcal{L}_X (S)_p \otimes T\big|_p + S\big|_p \otimes \mathcal{L}_X (T)_p . \tag{8.17}$$

Let S and T be (r,s) and (k,l) tensor fields, respectively. Then

$$\mathcal{L}_X (S \otimes T)_p = \lim_{t\to 0} \frac{(\phi_{-t})_* (S\otimes T)\big|_{\phi_t(p)} - (S\otimes T)\big|_p}{t} \tag{8.18}$$

Take $\omega_1,\dots,\omega_r,\eta_1,\dots,\eta_k \in T_p^*M$ and $X_1,\dots,X_s,Y_1,\dots,Y_l \in T_pM$. Then

$$\begin{aligned}
&(\phi_{-t})_* (S\otimes T)\big|_{\phi_t(p)} (\omega_1,\dots,\omega_r,\eta_1,\dots,\eta_k,X_1,\dots,X_s,Y_1,\dots,Y_l)\\
&= (S\otimes T)\big|_p \left(\phi^*_{-t}\omega_1,\dots,\phi^*_{-t}\omega_r,\phi^*_{-t}\eta_1,\dots,\phi^*_{-t}\eta_k,(\phi_t)_* X_1,\dots,(\phi_t)_* X_s,(\phi_t)_* Y_1,\dots,(\phi_t)_* Y_l\right)\\
&= S\big|_p \left(\phi^*_{-t}\omega_1,\dots,\phi^*_{-t}\omega_r,(\phi_t)_* X_1,\dots,(\phi_t)_* X_s\right) T\big|_p \left(\phi^*_{-t}\eta_1,\dots,\phi^*_{-t}\eta_k,(\phi_t)_* Y_1,\dots,(\phi_t)_* Y_l\right).
\end{aligned}$$

Let $A = (\omega_1,\dots,\omega_r,\eta_1,\dots,\eta_k,X_1,\dots,X_s,Y_1,\dots,Y_l)$, $B = \left(\phi^*_{-t}\omega_1,\dots,\phi^*_{-t}\omega_r,(\phi_t)_* X_1,\dots,(\phi_t)_* X_s\right)$, $C = \left(\phi^*_{-t}\eta_1,\dots,\phi^*_{-t}\eta_k,(\phi_t)_* Y_1,\dots,(\phi_t)_* Y_l\right)$, $D = (\omega_1,\dots,\omega_r,X_1,\dots,X_s)$, $E = (\eta_1,\dots,\eta_k,Y_1,\dots,Y_l)$. Now,

$$\begin{aligned}
&\left[(\phi_{-t})_* (S\otimes T)\big|_{\phi_t(p)} - (S\otimes T)\big|_p\right](A)\\
&= S\big|_p (B)\, T\big|_p (C) - S\big|_p (D)\, T\big|_p (E)\\
&= S\big|_p (B)\, T\big|_p (C) - S\big|_p (D)\, T\big|_p (C) + S\big|_p (D)\, T\big|_p (C) - S\big|_p (D)\, T\big|_p (E)\\
&= \left(S\big|_p (B) - S\big|_p (D)\right) T\big|_p (C) + S\big|_p (D) \left(T\big|_p (C) - T\big|_p (E)\right).
\end{aligned} \tag{8.19}$$

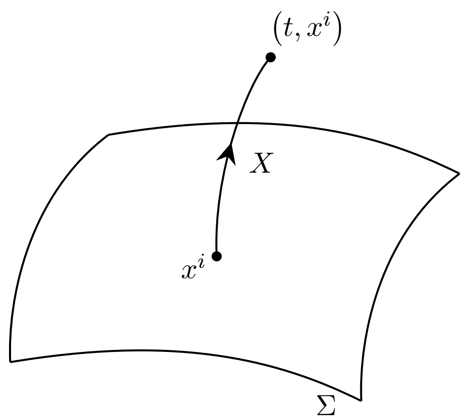
Now we shall divide 8.19 by t and take the limit $t \to 0$. Thus we obtain

$$\begin{aligned}
\mathcal{L}_X (S\otimes T)_p (A) &= \lim_{t\to 0} \frac{S\big|_p (B) - S\big|_p (D)}{t} \lim_{t\to 0} T\big|_p (C) + S\big|_p (D) \lim_{t\to 0} \frac{T\big|_p (C) - T\big|_p (E)}{t}\\
&= \mathcal{L}_X (S)_p (D)\;\, T\big|_p (E) + S\big|_p (D)\;\, \mathcal{L}_X (T)_p (E)\\
&= \left[\mathcal{L}_X (S)_p \otimes T\big|_p\right](A) + \left[S\big|_p \otimes \mathcal{L}_X (T)_p\right](A).
\end{aligned} \tag{8.20}$$

Therefore, 8.17 holds, and hence Lie derivative is a derivation.

Convenient Coordinates

Let $X \in \mathfrak{X}(M)$, and Σ a hypersurface (surface with dimension $\dim M - 1$) in M such that X is nowhere tangent to Σ. Let x^i, $i = 1,\dots,n-1$ be coordinates on Σ. Now, assign coordinates $\left(t, x^i\right)$ to the point parameter distance t along the integral curve of X that starts at the points with coordinates x^i on Σ.



This defines a coordinate chart $\left(t, x^i\right)$ at least for small t, i.e. in a neighborhood of Σ. Furthermore, the integral curves of X are the curves $\left(t, x^i\right)$ with fixed x^i and parameter t. The tangent to these curves is $\frac{\partial}{\partial t}$. So we have constructed coordinates such that $X = \frac{\partial}{\partial t}$.

φ_t sends a point p with coordinates $x^\mu = \left(t_p, x^i_p\right)$ to the point with coordinates $y^\mu = \left(t_p + t, x^i_p\right)$. Then we have

$$\frac{\partial y^\mu}{\partial x^\nu} = \delta^\mu{}_\nu. \tag{8.21}$$

Now, writing the formula for push forward of an (r, s) tensor in the component form (8.13),

$$\begin{aligned}\left(\left(\varphi_t\right)_* T\big|_p\right)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} &= \prod_{i=1}^{r} \frac{\partial y^{\mu_i}}{\partial x^{\rho_i}}\left(p\right) \prod_{j=1}^{s} \frac{\partial x^{\sigma_j}}{\partial y^{\nu_j}}\left(\varphi\left(p\right)\right) \left(T\big|_p\right)^{\rho_1\cdots\rho_r}{}_{\sigma_1\cdots\sigma_s} \\ &= \left(T\big|_p\right)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s},\end{aligned} \tag{8.22}$$

so

$$\left(\left(\varphi_{-t}\right)_* T\big|_{\varphi_t(p)}\right)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} = \left(T\big|_{\varphi_t(p)}\right)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s}. \tag{8.23}$$

In this coordinate chart, $p = \left(t_p, x^i_p\right)$, and $\varphi_t\left(p\right) = \left(t_p + t, x^i_p\right)$. Therefore,

$$\begin{aligned}&\left(\mathcal{L}_X T\right)_p^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} \\ &= \lim_{t\to 0} \frac{\left(\left(\varphi_{-t}\right)_* T\big|_{\varphi_t(p)}\right)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} - \left(T\big|_p\right)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s}}{t} \\ &= \lim_{t\to 0} \frac{\left(T\big|_{\varphi_t(p)}\right)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} - \left(T\big|_p\right)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s}}{t} \\ &= \lim_{t\to 0} \frac{1}{t}\left[T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s}\left(t_p + t, x^i_p\right) - T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s}\left(t_p, x^i_p\right)\right] \\ &= \frac{\partial}{\partial t} T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s}\left(t_p, x^i_p\right).\end{aligned} \tag{8.24}$$

In this chart, we have $X\left(f\right) = \frac{\partial f}{\partial t}$. Therefore,

$$\mathcal{L}_X f = X\left(f\right). \tag{8.25}$$

Both sides of this equation are scalars, and it must therefore be basis independent. Next, consider a vector field Y. In our coordinate above,

$$\left(\mathcal{L}_X Y\right)^\mu = \frac{\partial Y^\mu}{\partial t}. \tag{8.26}$$

In our coordinate X^μ is either 1 or 0, so $\frac{\partial X^\mu}{\partial x^\nu} = 0$.

$$\left[X, Y\right]^\mu = X^\nu \frac{\partial Y^\mu}{\partial x^\nu} - Y^\nu \frac{\partial X^\mu}{\partial x^\nu} = \frac{\partial Y^\mu}{\partial x^0} = \left(\mathcal{L}_X Y\right)^\mu. \tag{8.27}$$

Therefore, $\left(\mathcal{L}_X Y\right)^\mu = \left[X, Y\right]^\mu$ in our coordinate. If two vectors have the same components in one basis, they are equal in any other basis. Therefore,

$$\mathcal{L}_X Y = \left[X, Y\right]. \tag{8.28}$$

Now we shall compute the Lie derivative of a 1-form. Let ω be a 1-form. Consider a vector field U. Then $\omega_\mu U^\mu$ is a scalar, i.e. a smooth function on the manifold. Therefore,

$$\mathcal{L}_X\left(\omega_\mu U^\mu\right) = X\left(\omega_\mu U^\mu\right) = X^\nu \frac{\partial \omega_\mu}{\partial x^\nu} U^\mu + X^\nu \omega_\mu \frac{\partial U^\mu}{\partial x^\nu}. \tag{8.29}$$

Since Lie derivative is a derivation,

$$\begin{aligned}\mathcal{L}_X\left(\omega_\mu U^\mu\right) &= \left(\mathcal{L}_X\omega\right)_\mu U^\mu + \omega_\mu\left(\mathcal{L}_X U\right)^\mu \\ &= \left(\mathcal{L}_X\omega\right)_\mu U^\mu + \omega_\mu\left[X,U\right]^\mu \\ &= \left(\mathcal{L}_X\omega\right)_\mu U^\mu + \omega_\mu X^\nu\frac{\partial U^\mu}{\partial x^\nu} - \omega_\mu U^\nu\frac{\partial X^\mu}{\partial x^\nu}.\end{aligned} \tag{8.30}$$

Combining 8.29 and 8.30, we get

$$\begin{aligned}\left(\mathcal{L}_X\omega\right)_\mu U^\mu &= X^\nu\frac{\partial\omega_\mu}{\partial x^\nu}U^\mu + \omega_\mu U^\nu\frac{\partial X^\mu}{\partial x^\nu} \\ &= X^\nu\frac{\partial\omega_\mu}{\partial x^\nu}U^\mu + \omega_\nu U^\mu\frac{\partial X^\nu}{\partial x^\mu}.\end{aligned}$$

Since $U \in \mathfrak{X}(M)$ is arbitrary,

$$\left(\mathcal{L}_X\omega\right)_\mu = X^\nu\frac{\partial\omega_\mu}{\partial x^\nu} + \omega_\nu\frac{\partial X^\nu}{\partial x^\mu}. \tag{8.31}$$

In exactly similar fashion, one can define the Lie derivative of an (r,s) tensor field as

$$\begin{aligned}\left(\mathcal{L}_X T\right)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} = X^\sigma\partial_\sigma T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} &- \sum_{i=1}^{r}\left(\partial_\lambda X^\mu\right)T^{\mu_1\cdots\mu_{i-1}\lambda\mu_{i+1}\cdots\mu_r}{}_{\nu_1\cdots\nu_s} \\ &+ \sum_{j=1}^{s}\left(\partial_{\mu_j}X^\lambda\right)T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{j-1}\lambda\nu_{j+1}\cdots\nu_s}.\end{aligned} \tag{8.32}$$

Manifestly Covariant Form of Lie Derivative

Now, we shall show that if T is a (r,s) tensor field, and ∇ is the Levi-Civita connection,

$$\begin{aligned}\left(\mathcal{L}_X T\right)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} = X^\sigma\nabla_\sigma T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} &- \sum_{i=1}^{r}\left(\nabla_\lambda X^{\mu_i}\right)T^{\mu_1\cdots\mu_{i-1}\lambda\mu_{i+1}\cdots\mu_r}{}_{\nu_1\cdots\nu_s} \\ &+ \sum_{j=1}^{s}\left(\nabla_{\nu_j}X^\lambda\right)T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{j-1}\lambda\nu_{j+1}\cdots\nu_s}.\end{aligned} \tag{8.33}$$

We know that

$$T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s;\sigma} = T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s,\sigma} + \sum_{i=1}^{r}\Gamma^{\mu_i}_{\lambda\sigma}T^{\mu_1\cdots\mu_{i-1}\lambda\mu_{i+1}\cdots\mu_r}{}_{\nu_1\cdots\nu_s} - \sum_{j=1}^{s}\Gamma^{\lambda}_{\nu_j\sigma}T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{j-1}\lambda\nu_{j+1}\cdots\nu_s}. \tag{8.34}$$

Therefore, $X^\sigma\nabla_\sigma T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s}$ is equal to

$$X^\sigma\partial_\sigma T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} + \sum_{i=1}^{r}X^\sigma\Gamma^{\mu_i}_{\lambda\sigma}T^{\mu_1\cdots\mu_{i-1}\lambda\mu_{i+1}\cdots\mu_r}{}_{\nu_1\cdots\nu_s} - \sum_{j=1}^{s}X^\sigma\Gamma^{\lambda}_{\nu_j\sigma}T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{j-1}\lambda\nu_{j+1}\cdots\nu_s}. \tag{8.35}$$

$\nabla_\lambda X^{\mu_i} = X^{\mu_i}_{;\lambda} = \partial_\lambda X^\mu + \Gamma^\mu_{\sigma\lambda}X^\sigma$. Therefore,

$$\begin{aligned}-\sum_{i=1}^{r}\left(\nabla_\lambda X^{\mu_i}\right)T^{\mu_1\cdots\mu_{i-1}\lambda\mu_{i+1}\cdots\mu_r}{}_{\nu_1\cdots\nu_s} &= -\sum_{i=1}^{r}\left(\partial_\lambda X^\mu + \Gamma^\mu_{\sigma\lambda}X^\sigma\right)T^{\mu_1\cdots\mu_{i-1}\lambda\mu_{i+1}\cdots\mu_r}{}_{\nu_1\cdots\nu_s} \\ &= -\sum_{i=1}^{r}\left(\partial_\lambda X^\mu\right)T^{\mu_1\cdots\mu_{i-1}\lambda\mu_{i+1}\cdots\mu_r}{}_{\nu_1\cdots\nu_s} - \sum_{i=1}^{r}\Gamma^\mu_{\lambda\sigma}X^\sigma T^{\mu_1\cdots\mu_{i-1}\lambda\mu_{i+1}\cdots\mu_r}{}_{\nu_1\cdots\nu_s}.\end{aligned} \tag{8.36}$$

Furthermore, $\nabla_{\nu_j} X^\lambda = X^\lambda{}_{;\nu_j} = \partial_{\mu_j} X^\lambda + \Gamma^\lambda_{\sigma\nu_j} X^\sigma$. Therefore,

$$\begin{aligned}
&\sum_{j=1}^s \left(\nabla_{\nu_j} X^\lambda\right) T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{j-1}\lambda\nu_{j+1}\cdots\nu_s} \\
&= \sum_{j=1}^s \left(\partial_{\mu_j} X^\lambda + \Gamma^\lambda_{\sigma\nu_j} X^\sigma\right) T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{j-1}\lambda\nu_{j+1}\cdots\nu_s} \\
&= \sum_{j=1}^s \left(\partial_{\mu_j} X^\lambda\right) T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{j-1}\lambda\nu_{j+1}\cdots\nu_s} + \sum_{j=1}^s \Gamma^\lambda_{\nu_j\sigma} X^\sigma T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{j-1}\lambda\nu_{j+1}\cdots\nu_s}.
\end{aligned} \tag{8.37}$$

Adding 8.35, 8.36, 8.37, we obtain that the RHS of 8.33 is equal to

$$X^\sigma \partial_\sigma T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} - \sum_{i=1}^r \left(\partial_\lambda X^\mu\right) T^{\mu_1\cdots\mu_{i-1}\lambda\mu_{i+1}\cdots\mu_r}{}_{\nu_1\cdots\nu_s} + \sum_{j=1}^s \left(\partial_{\mu_j} X^\lambda\right) T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{j-1}\lambda\nu_{j+1}\cdots\nu_s}, \tag{8.38}$$

which is exactly the formula for $(\mathcal{L}_X T)^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s}$. Therefore, 8.33 holds.

Now, using 8.33,

$$\begin{aligned}
(\mathcal{L}_X g)_{\mu\nu} &= X^\sigma \nabla_\sigma g_{\mu\nu} + \left(\nabla_\mu X^\lambda\right) g_{\lambda\nu} + \left(\nabla_\nu X^\lambda\right) g_{\mu\lambda} \\
&= \nabla_\mu \left(X^\lambda g_{\lambda\nu}\right) + \nabla_\nu \left(X^\lambda g_{\mu\lambda}\right) \\
&= \nabla_\mu X_\nu + \nabla_\nu X_\mu \\
&= 2\nabla_{(\mu} X_{\nu)}.
\end{aligned} \tag{8.39}$$

If φ_t is a symmetry transformation of T for all t, then $(\phi_t)_* T\big|_p = T\big|_{\varphi_t(p)}$ and $(\varphi_{-t})_* T\big|_{\varphi_t(p)} = T\big|_p$ for all t. Therefore,

$$(\mathcal{L}_X T)_p = \lim_{t\to 0} \frac{1}{t}\left[(\varphi_{-t})_* T\big|_{\varphi_t(p)} - T\big|_p\right] = 0. \tag{8.40}$$

So $\mathcal{L}_X T = 0$ if φ_t is a symmetry transformation. If φ_t is a one-parameter group of isometries, then $\mathcal{L}_X g = 0$. In other words,

$$\nabla_a X_b + \nabla_b X_a = 0. \tag{8.41}$$

This is the **Killing equation**, and the solutions X to this equation are called **Killing vector fields**.

9 Linearized Theory Revisited

§9.1 Gauge Symmetry in Linearized Theory

We have seen earlier that diffeomorphisms are gauge symmetries in GR. If φ is a diffeomorphism, $(M, g, F, \ldots)$ and $(\varphi_*(g), \varphi_*(F), \ldots)$ are physically equivalent. Let us now go back to the metric

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{9.1}$$

where $\eta_{\mu\nu} = \mathrm{diag}(-1,1,1,1)$. We focus on diffeomorphisms that preserve 9.1. A general diffeomorphism φ would lead to $(\varphi_*\eta)_{\mu\nu}$ very different from $\mathrm{diag}(-1,1,1,1)$. However, if we consider 1-parameter family of diffeomorphisms φ_t, then φ_0 is the identity map. So, if t is small, then φ_t is close to identity. Therefore, small t guarantees that $((\varphi_t)_*\eta)_{\mu\nu}$ will be close to $\mathrm{diag}(-1,1,1,1)$ and the form 9.1 will be preserved. For small t,

$$(\varphi_{-t})_*\left(T\big|_{\varphi_t(p)}\right) = T\big|_p + t\,(\mathcal{L}_X T)_p + \mathcal{O}\left(t^2\right) = T\big|_p + (\mathcal{L}_\xi T)_p + \mathcal{O}\left(t^2\right), \tag{9.2}$$

where X is the vector field that generates φ_t (φ_t is defined of integral curves of X) and $\xi^a = tX^a$. t is small, so we can treat ξ^a as first order quantity. If we apply 9.2 to the energy momentum tensor, which itself is in the 1st order, because $T_{\mu\nu} \ll 1$ for spacetime being nearly flat. Then $\mathcal{L}_\xi T$ is higher order and hence negligible. Therefore, energy momentum tensor is gauge invariant in the first order. The same is true for any tensor that vanishes in the unperturbed spacetime, e.g. the Riemann curvature tensor.

Now, what abouot the metric tensor $g_{\mu\nu}$?

$$(\varphi_{-t})_*\left(g\big|_{\varphi_t(p)}\right) = g\big|_p + (\mathcal{L}_\xi g)_p + \cdots = \eta\big|_p + h\big|_p + (\mathcal{L}_\xi \eta)_p + \cdots, \tag{9.3}$$

where we have neglected $\mathcal{L}_\xi h$ which is very small as ξ and h are both of the first order. But we want $(\varphi_{-t})_*\left(g\big|_{\varphi_t(p)}\right) = g\big|_p$ for all $p \in M$. Comparing 9.1 and 9.3, we can deduce that $h\big|_p$ and $h\big|_p + (\mathcal{L}_\xi\eta)_p$ are **equivalent metric perturbations**. Therefore, the Linearized theort possesses the symmetry $h \to h + \mathcal{L}_\xi \eta$ for small ξ^μ. In our chart $\{x^\mu\}$, $(\mathcal{L}_\xi\eta)_{\mu\nu} = \partial_\mu\xi_\nu + \partial_\nu\xi_\mu$ (covariant derivatives are replaced by partial derivatives to the first order). Therefore, the gauge symmetry is

$$\boxed{h_{\mu\nu} \to h_{\mu\nu} + \partial_\mu\xi_\nu + \partial_\nu\xi_\mu.} \tag{9.4}$$

Now, consider the linearized Einstein equation (7.36):

$$-\frac{1}{2}\partial^\rho\partial_\rho\bar{h}_{\mu\nu} + \partial^\rho\partial_{(\mu}\bar{h}_{\nu)\rho} - \frac{1}{2}\eta_{\mu\nu}\partial^\rho\partial^\sigma\bar{h}_{\rho\sigma} = 8\pi T_{\mu\nu}. \tag{9.5}$$

We have seen earlier that under gauge transformation $h'_{\mu\nu} = h_{\mu\nu} + \partial_\mu\xi_\nu + \partial_\nu\xi_\mu$. Now we want to know how $\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}h\eta_{\mu\nu}$ transforms under such gauge transformation.

$$\begin{aligned}
\bar{h}'_{\mu\nu} = h'_{\mu\nu} - \frac{1}{2}h'\eta_{\mu\nu} &= h_{\mu\nu} + \partial_\mu\xi_\nu + \partial_\nu\xi_\mu - \frac{1}{2}\eta_{\mu\nu}\left(\eta^{\rho\sigma}h'_{\rho\sigma}\right)\\
&= h_{\mu\nu} + \partial_\mu\xi_\nu + \partial_\nu\xi_\mu - \frac{1}{2}\eta_{\mu\nu}\eta^{\rho\sigma}\left(h_{\rho\sigma} + \partial_\rho\xi_\sigma + \partial_\sigma\xi_\rho\right)\\
&= h_{\mu\nu} + \partial_\mu\xi_\nu + \partial_\nu\xi_\mu - \frac{1}{2}\eta_{\mu\nu}h - \eta_{\mu\nu}\partial^\rho\xi_\rho\\
&= \bar{h}_{\mu\nu} + \partial_\mu\xi_\nu + \partial_\nu\xi_\mu - \eta_{\mu\nu}\partial^\rho\xi_\rho.
\end{aligned}$$

Therefore, under gauge transformation,

$$\boxed{\bar{h}_{\mu\nu} \to \bar{h}_{\mu\nu} + \left(\partial_\mu\xi_\nu + \partial_\nu\xi_\mu - \eta_{\mu\nu}\partial^\rho\xi_\rho\right).} \tag{9.6}$$

Under this transformation,

$$\begin{aligned}\partial^\nu \bar{h}'_{\mu\nu} &= \partial^\nu \bar{h}_{\mu\nu} + \partial^\nu \partial_\mu \xi_\nu + \partial^\nu \partial_\nu \xi_\mu - \eta_{\mu\nu} \partial^\nu \partial^\rho \xi_\rho \\ &= \partial^\nu \bar{h}_{\mu\nu} + \partial_\mu \partial^\nu \xi_\nu + \partial^\nu \partial_\nu \xi_\mu - \partial_\mu \partial^\rho \xi_\rho \\ &= \partial^\nu \bar{h}_{\mu\nu} + \partial^\nu \partial_\nu \xi_\mu .\end{aligned}$$

Therefore, under gauge transformation,

$$\boxed{\partial^\nu \bar{h}_{\mu\nu} \to \partial^\nu \bar{h}_{\mu\nu} + \partial^\nu \partial_\nu \xi_\mu .} \tag{9.7}$$

If we choose ξ_μ to satisfy the wave equation $\partial^\nu \partial_\nu \xi_\mu = -\partial^\nu \bar{h}_{\mu\nu}$ (which we can solve using a Green's function), then the transformed equation will read

$$\partial^\nu \bar{h}'_{\mu\nu} = 0. \tag{9.8}$$

This is called the **harmonic gauge**. In this gauge, the linearized Einstein equation (9.5) reduces to

$$-\frac{1}{2}\partial^\rho \partial_\rho \bar{h}'_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{9.9}$$

leading to

$$\Box \bar{h}'_{\mu\nu} = -16\pi T_{\mu\nu}. \tag{9.10}$$

§9.2 The Newtonian Limit

We will now see how GR reduces to Newtonian theory in the limit of non-relativistic motion and a weak gravitational field. We expect Newtonian theory to be valid as $c \to \infty$. We stick to $c = 1$, but introduce a small parameter $0 < \varepsilon \ll 1$ and write ε everywhere whenever $\frac{1}{c}$ would appear.

We take an almost inertial coordinate system (freely falling system) $x^\mu = \left(t, x^i\right)$. The 3-velocity of a particle is $v^i = \frac{\mathrm{d}x}{\mathrm{d}\tau} = \mathcal{O}\left(\varepsilon\right)$. In Newtonian theory, $\Phi = -\frac{GM}{r}$ and $\frac{v^2}{r} = \frac{GM}{r^2}$. So $|\Phi| = v^2 = \mathcal{O}\left(\varepsilon^2\right)$. Therefore, we expect tha gravitational field to be $\mathcal{O}\left(\varepsilon^2\right)$. We further assume

$$h_{00} = \mathcal{O}\left(\varepsilon^2\right), \quad h_{0i} = \mathcal{O}\left(\varepsilon^3\right), \quad h_{iij} = \mathcal{O}\left(\varepsilon^2\right). \tag{9.11}$$

Since the matter which generates the gravitational field moves non-relativistically ($v = \mathcal{O}\left(\varepsilon\right)$), time derivative of the gravitational field will be small compared to spatial derivatives: the gravitational field at a point $\mathbf{x}$ due to a body of mass m located at $\mathbf{x}\left(t\right)$ is given by

$$\Phi = -\frac{m}{|\mathbf{x} - \mathbf{x}(t)|}. \tag{9.12}$$

Let $L = |\mathbf{x} - \mathbf{x}(t)|$. Then

$$\partial_i \Phi = \boldsymbol{\nabla}\Phi = \frac{m\left(\mathbf{x} - \mathbf{x}(t)\right)}{|\mathbf{x} - \mathbf{x}(t)|^3}. \tag{9.13}$$

Hence,

$$|\partial_i \Phi| = \frac{m}{|\mathbf{x} - \mathbf{x}(t)|^2} = \frac{|\Phi|}{L}. \tag{9.14}$$

On the other hand, using $\dot{\mathbf{x}} = \mathcal{O}\left(\varepsilon\right)$, we get

$$\partial_0 \Phi = \frac{\partial \Phi}{\partial t} = \frac{\partial \Phi}{\partial x^i}\frac{\mathrm{d}x^i}{\mathrm{d}t} = \partial_i \Phi \dot{c}^i = \left(\boldsymbol{\nabla}\Phi\right) \cdot \dot{\mathbf{x}}. \tag{9.15}$$

Hence,

$$\partial_0 \Phi \le |\boldsymbol{\nabla}\Phi| \cdot |\dot{\mathbf{x}}| = \mathcal{O}\left(\frac{|\Phi|}{L}\right)\mathcal{O}\left(\varepsilon\right) \ll \mathcal{O}\left(\frac{|\Phi|}{L}\right). \tag{9.16}$$

Therefore, $\partial_0 \Phi \ll \partial_i \Phi$.

Consider the Lagrangian $\hat{L}$ of a time-like geodesic.

$$\begin{aligned}\hat{L}^2 &= -g_{\mu\nu}(x(\tau))\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} = -(\eta_{\mu\nu}+h_{\mu\nu})\dot{x}^\mu\dot{x}^\nu \\ &= -(-1+h_{00})\dot{x}^0\dot{x}^0 - h_{0i}\dot{x}^0\dot{x}^i - h_{i0}\dot{x}^i\dot{x}^0 - (\delta_{ij}+h_{ij})\dot{x}^i\dot{x}^j \\ &= (1-h_{00})\dot{t}^2 - 2h_{0i}\dot{t}\dot{x}^i - (\delta_{ij}+h_{ij})\dot{x}^i\dot{x}^j. \end{aligned} \tag{9.17}$$

Now, recall the definition od proper time:

$$\mathrm{d}\tau^2 = -g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu \implies 1 = -g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu. \tag{9.18}$$

Therefore, $\hat{L}^2 = 1$ on the time like geodesic. In other words,

$$(1-h_{00})\dot{t}^2 - 2h_{0i}\dot{t}\dot{x}^i - (\delta_{ij}+h_{ij})\dot{x}^i\dot{x}^j = 1. \tag{9.19}$$

Observe that $\frac{\partial}{\partial t} = \frac{\mathrm{d}x^i}{\mathrm{d}t}\frac{\partial}{\partial x^i} = v^i\frac{\partial}{\partial x^i}$. Since $v^i = \mathcal{O}(\varepsilon)$, $\frac{\partial}{\partial t} \ll \frac{\partial}{\partial x^i}$, i.e. time derivatives are much smaller than spatial derivatives. Also,

$$\dot{x}^i = \frac{\mathrm{d}x^i}{\mathrm{d}\tau} = \frac{\mathrm{d}x^i}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\tau} = v^i\dot{t} = \mathcal{O}(\varepsilon)\,\dot{t} \implies \dot{t} = \mathcal{O}(1). \tag{9.20}$$

Now, rewriting 9.19, we get

$$\begin{aligned} (1-h_{00})\dot{t}^2 - \delta_{ij}\dot{x}^i\dot{x}^j &= 1 + \underbrace{2h_{0i}\dot{t}\dot{x}^i + h_{ij}\dot{x}^i\dot{x}^j}_{\mathcal{O}(\varepsilon^4)} \\ \implies (1-h_{00})\dot{t}^2 &= 1 + \delta_{ij}\dot{x}^i\dot{x}^j + \mathcal{O}(\varepsilon^4) \\ \implies (1-h_{00})^{\frac{1}{2}}\dot{t} &= 1 + \frac{1}{2}\delta_{ij}\dot{x}^i\dot{x}^j + \mathcal{O}(\varepsilon^4) \\ \implies \dot{t} &= \left(1 + \frac{1}{2}\delta_{ij}\dot{x}^i\dot{x}^j + \mathcal{O}(\varepsilon^4)\right)(1-h_{00})^{-\frac{1}{2}} \\ \implies \dot{t} &= \left(1 + \frac{1}{2}\delta_{ij}\dot{x}^i\dot{x}^j + \mathcal{O}(\varepsilon^4)\right)\left(1 + \frac{1}{2}h_{00} + \mathcal{O}(\varepsilon^4)\right) \\ \implies \dot{t} &= 1 + \frac{1}{2}h_{00} + \frac{1}{2}\delta_{ij}\dot{x}^i\dot{x}^j + \mathcal{O}(\varepsilon^4). \end{aligned} \tag{9.21}$$

Now, recall 9.17. The Euler-Lagrange equation for x^i is

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\partial\hat{L}}{\partial\dot{x}^i}\right) = \frac{\partial\hat{L}}{\partial x^i}. \tag{9.22}$$

From 9.17, one has

$$2\hat{L}\frac{\partial\hat{L}}{\partial x^i} = -h_{00,i}\dot{t}^2 - 2h_{0j,i}\dot{t}\dot{x}^j - h_{mj,i}\dot{x}^m\dot{x}^j. \tag{9.23}$$

Furthermore,

$$2\hat{L}\frac{\partial\hat{L}}{\partial\dot{x}^i} = -2h_{0i}\dot{t} - 2(\delta_{ij}+h_{ij})\dot{x}^j. \tag{9.24}$$

Then the Euler-Lagrange equation for x^i reduces to

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left[2h_{0i}\dot{t} - 2(\delta_{ij}+h_{ij})\dot{x}^j\right] = -h_{00,i}\dot{t}^2 - 2h_{0j,i}\dot{t}\dot{x}^j - h_{mj,i}\dot{x}^m\dot{x}^j. \tag{9.25}$$

$h_{0j,i} = \mathcal{O}\left(\frac{\varepsilon^3}{L}\right)$, so $h_{0j,i}\dot{t}\dot{x}^j = \mathcal{O}\left(\frac{\varepsilon^4}{L}\right)$. Similarly, $h_{mj,i}\dot{x}^m\dot{x}^j = \mathcal{O}\left(\frac{\varepsilon^4}{L}\right)$. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left[2h_{0i}\dot{t} - 2(\delta_{ij}+h_{ij})\dot{x}^j\right] = -h_{00,i} + \mathcal{O}\left(\frac{\varepsilon^4}{L}\right). \tag{9.26}$$

Note that

$$\frac{\mathrm{d}}{\mathrm{d}\tau} = \frac{\mathrm{d}x^k}{\mathrm{d}\tau}\frac{\partial}{\partial x^k} = \dot{x}^k\frac{\partial}{\partial x^k}, \tag{9.27}$$

which is of order $\frac{\varepsilon}{L}$. $h_{0i}\dot{t}$ is $\mathcal{O}\left(\varepsilon^3\right)$, so $\frac{\mathrm{d}\left(h_{0i}\dot{t}\right)}{\mathrm{d}\tau}$ is $\mathcal{O}\left(\frac{\varepsilon^4}{L}\right)$. Similarly, $h_{ij}\dot{x}^j$ is $\mathcal{O}\left(\varepsilon^3\right)$, so $\frac{\mathrm{d}\left(h_{ij}\dot{x}^j\right)}{\mathrm{d}\tau}$ is also $\mathcal{O}\left(\frac{\varepsilon^4}{L}\right)$. Therefore, 9.26 gives us

$$-2\ddot{x}^i + \mathcal{O}\left(\frac{\varepsilon^4}{L}\right) = -h_{00,i} + \mathcal{O}\left(\frac{\varepsilon^4}{L}\right). \tag{9.28}$$

Retaining only the leading order terms gives us

$$\ddot{x}^i = \frac{1}{2}h_{00,i}. \tag{9.29}$$

Using chain rule,

$$\begin{aligned}\ddot{x}^i &= \frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\mathrm{d}x^i}{\mathrm{d}\tau}\right) = \frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\mathrm{d}x^i}{\mathrm{d}t}\dot{t}\right)\\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}x^i}{\mathrm{d}t}\dot{t}\right)\dot{t}\\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}x^i}{\mathrm{d}t}\right)\left(\dot{t}\right)^2 + \text{ subleading terms}\\ &= \frac{\mathrm{d}^2x^i}{\mathrm{d}t^2} + \text{ subleading terms }.\end{aligned} \tag{9.30}$$

For $\Phi = -\frac{1}{2}h_{00}$, 9.29 reduces to

$$\frac{\mathrm{d}^2x^i}{\mathrm{d}t^2} = -\partial_i\Phi, \tag{9.31}$$

which is exactly Newton's equation for a body moving in a gravitational field Φ.

§9.3 Gravitational Waves

In vacuum, the linearized Einstein equation reduces to the source-free wave equation (9.10):

$$\partial^\rho\partial_\rho\bar{h}'_{\mu\nu} = 0. \tag{9.32}$$

Let's look for plane wave solutions to this equation:

$$\bar{h}'_{\mu\nu} = \mathrm{Re}\left(H'_{\mu\nu}e^{ik_\sigma x^\sigma}\right). \tag{9.33}$$

where $H_{\mu\nu}$ is a constant symmetric complex matrix. (We shall suppress Re in all subsequent equations.) Then we have

$$\begin{aligned}&\partial^\rho\partial_\rho\left(H'_{\mu\nu}e^{ik_\sigma x^\sigma}\right) = 0\\ \implies\ &g^{\rho\tau}\partial_\tau\left(H'_{\mu\nu}\left(ik_\rho\right)e^{ik_\sigma x^\sigma}\right) = 0\\ \implies\ &g^{\rho\tau}\left(H'_{\mu\nu}\left(ik_\rho\right)\left(ik_\tau\right)e^{ik_\sigma x^\sigma}\right) = 0\\ \implies\ &-H'_{\mu\nu}k_\rho k^\rho e^{ik_\sigma x^\sigma} = 0.\end{aligned} \tag{9.34}$$

Therefore, $k_\rho k^\rho = 0$, i.e. k^μ is a null vector so it propagates with the speed of light.

Now consider the harmonic gauge condition (9.8).

$$\begin{aligned}\partial^\nu\bar{h}'_{\mu\nu} = 0 \implies\ &g^{\nu\sigma}\partial_\sigma\left(H'_{\mu\nu}e^{ik_\rho x^\rho}\right) = 0\\ \implies\ &g^{\nu\sigma}H'_{\mu\nu}\left(ik_\sigma\right)e^{ik_\rho x^\rho} = 0\\ \implies\ &k^\nu H'_{\mu\nu} = 0.\end{aligned} \tag{9.35}$$

This means the waves are **transverse**, i.e. oscillations are orthogonal to the wave vector. Since H is symmetric, there are 10 independent components on the matrix. But now the transverse wave condition gives us four constraints: $k^\nu H'_{\mu\nu}$ one condition for each $\nu \in \{0,1,2,3\}$. Hence, there now will be $10-4=6$ independent components of H.

However, $\partial^\nu \bar{h}'_{\mu\nu} = 0$ does not eliminate all gauge freedom. Recall the harmonic gauge condition (9.8).

$$\partial^\nu \bar{h}_{\mu\nu} = -\partial^\nu \partial_\nu \xi_\mu = -\Box \xi_\mu. \tag{9.36}$$

If we choose $\xi_\mu \to \xi_\mu + \zeta_\mu$ with $\Box \zeta_\mu = 0$, then 9.36 is still satisfied. This $\Box \xi_\mu = 0$ condition is called **residual gauge symmetry**. We look for the solution to the wave equation $\partial^\nu \partial_\nu \xi_\mu = 0$. If we take

$$\xi_\mu = X_\mu e^{ik_\rho x^\rho}, \tag{9.37}$$

this obviously satisfies $\partial^\nu \partial_\nu \xi_\mu = 0$ since k^μ is a null vector. Using 9.6, we get

$$\begin{aligned}
\bar{h}_{\mu\nu} e^{-ik_\sigma x^\sigma} &\to \bar{h}_{\mu\nu} e^{-ik_\sigma x^\sigma} + \left(\partial_\mu \xi_\nu\right) e^{-ik_\sigma x^\sigma} + \left(\partial_\nu \xi_\mu\right) e^{-ik_\sigma x^\sigma} - \eta_{\mu\nu} \left(\partial^\rho \xi_\rho\right) e^{-ik_\sigma x^\sigma} \\
&= \bar{h}_{\mu\nu} e^{-ik_\sigma x^\sigma} + \partial_\mu \left(X_\nu e^{ik_\rho x^\rho}\right) e^{-ik_\sigma x^\sigma} + \partial_\nu \left(X_\mu e^{ik_\rho x^\rho}\right) e^{-ik_\sigma x^\sigma} - \eta_{\mu\nu} \eta^{\rho\beta} \partial_\beta \left(X_\rho e^{ik_\alpha x^\alpha}\right) e^{-ik_\sigma x^\sigma} \\
&= \bar{h}_{\mu\nu} e^{-ik_\sigma x^\sigma} + X_\nu \left(ik_\mu\right) + X_\mu \left(ik_\nu\right) - \eta_{\mu\nu} \eta^{\rho\beta} X_\rho \left(ik_\beta\right)
\end{aligned}$$

$$\therefore \boxed{H_{\mu\nu} \to H_{\mu\nu} + i\left(k_\mu X_\nu + k_\nu X_\mu - \eta_{\mu\nu} k^\beta X_\beta\right).} \tag{9.38}$$

We want the gravitational wave $\bar{h}'_{\mu\nu}$ to be purely spatial. Hence, $\bar{h}'_{0\nu} = 0$. Therefore, we have

$$H'_{0\nu} = 0, \tag{9.39}$$

which is known as the **longitudinal gauge condition**. This still does not determine X_μ uniquely. So we add the **tracelessness condition**:

$$H'^\mu{}_\mu = 0. \tag{9.40}$$

Using 9.38 and 9.40, we get

$$\begin{aligned}
0 = H'^\mu{}_\mu = \eta^{\mu\nu} H'_{\mu\nu} &= \eta^{\mu\nu} H_{\mu\nu} + i\eta^{\mu\nu} \left(k_\mu X_\nu + k_\nu X_\mu - \eta_{\mu\nu} k^\beta X_\beta\right) &(9.41)\\
&= H^\mu{}_\mu + i\left(k^\nu X_\nu + k^\mu X_\mu - 4k^\beta X_\beta\right) &(9.42)\\
&= H^\mu{}_\mu - 2ik^\nu X_\nu. &(9.43)
\end{aligned}$$

$$\therefore k^\nu X_\nu = -\frac{i}{2} H^\mu{}_\mu. \tag{9.44}$$

Now, we impose 9.39. For $\nu = 0$, we have

$$0 = H'_{00} = H_{00} + i\left(k_0 X_0 + k_0 X_0 + k^\beta X_\beta\right) = H_{00} + 2ik_0 X_0 + \frac{1}{2} H^\mu{}_\mu. \tag{9.45}$$

$$\therefore X_0 = -\frac{1}{2ik_0}\left(H_{00} + +\frac{1}{2} H^\mu{}_\mu\right) = \frac{i}{2k_0}\left(H_{00} + +\frac{1}{2} H^\mu{}_\mu\right). \tag{9.46}$$

For $\nu \neq 0$, we have

$$\begin{aligned}
& H_{0\nu} + i\left(k_0 X_\nu + k_\nu X_0\right) = H'_{0\nu} = 0 \\
\implies\ & ik_0 X_\nu = -H_{0\nu} - ik_\nu \frac{i}{2k_0}\left(H_{00} + +\frac{1}{2} H^\mu{}_\mu\right) \\
\implies\ & ik_0 X_\nu = \frac{1}{2k_0}\left[-2k_0 H_{0\nu} + k_\nu \left(H_{00} + +\frac{1}{2} H^\mu{}_\mu\right)\right] \\
\therefore\ & X_\nu = -\frac{i}{2k_0^2}\left[-2k_0 H_{0\nu} + k_\nu \left(H_{00} + +\frac{1}{2} H^\mu{}_\mu\right)\right].
\end{aligned} \tag{9.47}$$

Now the X_ν are determined uniquely. So we have eliminated all the gauge freedom.

So, we see that, for these choices of X_ν (9.46 and 9.47), we will have $H'^{\mu}{}_{\mu} = 0$ and $H'_{0\nu} = 0$. Thus, we began with 10 independent components in the symmetric matrix $H_{\mu\nu}$ –choosing the harmonic gauge implied the transverse wave condition, which led to 4 constraints– bringing the number of independent components to 6. We used our remaining residual freedom to choose X_μ such that $H'^{\mu}{}_{\mu} = 0$ (1 condition) and $H_{0\nu} = 0$ (4 conditions). But when $\nu = 0$, $H_{0\mu} = 0$ gives us

$$k_\mu H^{\mu\nu} = k_\mu H^{\mu 0} = 0. \tag{9.48}$$

So there are only three new conditions in $H0\nu = 0$, not four.

Thus the number of independent components of $H_{\mu\nu}$ is then $6 - 3 - 1 = 2$, and using all our freedom, we can have the only non-zero components of $H'_{\mu\nu}$ to be: $H'_{11}, H'_{12}, H'_{21}$, and H'_{22}. The trace-free condition gives: $H'_{11} + H'_{22} = 0$ so $H'_{11} = -H'_{22} = H_+$, and symmetry gives: $H'_{12} = H'_{21} = H_\times$.

$$H'_{\mu\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & H_+ & H_\times & 0 \\ 0 & H_\times & -H_+ & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{9.49}$$

So, the gravitational wave is transverse and has two possible polarizations. This is another way of interpreting that the gravitational field has two degrees of freedom per spacetime point (event).

Using all of our gauge freedom, we have gone to a sub-gauge of the harmonic gauge, known as the "transverse trace-less gauge", or the "radiation gauge". We shall denote $\bar{h}'_{\mu\nu}$ by $\bar{h}'^{\mathsf{TT}}_{\mu\nu}$. The superscript TT for denoting that it is in the Transverse Traceless gauge. Then we have

$$\bar{h}'^{\mathsf{TT}}_{\mu\nu} = h'^{\mathsf{TT}}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h'^{\mathsf{TT}}. \tag{9.50}$$

But

$$h'^{\mathsf{TT}} = \eta^{\mu\nu} h'^{\mathsf{TT}}_{\mu\nu} = \eta^{\mu\nu} H'_{\mu\nu} e^{ik_\rho X^\rho} = 0 \tag{9.51}$$

since $\eta^{\mu\nu} h'^{\mathsf{TT}}_{\mu\nu} = 0$. Therefore, $h'^{\mathsf{TT}}_{\mu\nu} = \bar{h}'^{\mathsf{TT}}_{\mu\nu}$ in the traceless transverse gauge. In this gauge, the linearized Einstien equation then can be written as

$$\partial^\rho \partial_\rho \bar{h}'^{\mathsf{TT}}_{\mu\nu} = \partial^\rho \partial_\rho h'^{\mathsf{TT}}_{\mu\nu} = 0. \tag{9.52}$$

§9.4 Physical Effects due to Gravitational Waves

Let us get a feeling for the physical effects due to gravitational waves. How would one detect a gravitational wave? To detect such waves, an observer can set up a family of test particles locally. Since the displacement vector S^a from the observer to any particle is governed by the geodesic deviation equation, we can use it to predict what the observer will see. But, we need to do that analysis carefully, as it would be to write out the geodesic deviation equation using the 'almost inertial' coordinates, and thereby, determine S^μ. The point is, S^μ are the components of S in a certain basis, so it would be tough to determine which variation in S^μ is due to the (variation of the) basis and which variation arises from S (due to the gravity-waves). One can make this approach work, but let us take a different approach.

Let us consider a freely falling observer, i.e. one that is following a geodesic in a general space-time. Our observer, in their frame, will have a set of measuring clocks with which they can measure distances. At some point p on their worldline, we could introduce a local inertial frame (Riemann normal coordinates) X, Y, Z in which the observer is at rest. Now, in the case that the observer sets up measuring rods of unit length pointing in the X, Y, Z directions at p, T_pM has an orthonormal basis e_α where $e_0^a = u^a$ is the 4-velocity of the observer, and e_i^a are the space-like vectors satisfying $u_a e_i^a = 0$ (meaning e_0 and e_j's are orthogonal) and $g_{ab} e_i^a e_j^b$ (meaning e_i and e_j are orthogonal spacelike).

In Minkowski spacetime, this basis can be extended to the observer's entire worldline by taking the basis vectors to have constant components (in an inertial frame), i.e. they do not depend on proper time τ. This implies, in particular, that the orthonormal basis is non-rotating. (No surprises, since in the Riemann normal coordinates, the connection components of the Levi-Civita connection, i.e. the Christoffel symbols all vanish.) Since the basis vectors have constant components, they are parallelly transported along the worldline.

In a curved spacetime, the analogue of this is to take the basis vectors e_i to be parallelly transported along the worldline, i.e. to let $(\nabla_U e_i)^a = 0$ which implies $U^b(\nabla_{e_b} e_i)^a = 0$. This equation determines the e_i^a's uniquely along the whole worldline of the observer if they were specified at any point p. They are just being parallelly transported along the geodesic. [Fact: parallel transport preserves inner product.] Since parallel transport preserves inner product, and since the e_i^a's were orthogonal at p, they will remain orthogonal.

The basis just constructed, is called a "parallelly transported frame". It is the kind of basis that would be constructed by an observer freely falling with 3 gyroscopes whose spin axes define the spatial basis vectors. Using such a basis, we can be sure that an increase in a component of S is really an increase in the distance to the particle in a particular direction, rather than a basis-independent effect.

Now, let us imagine that the observer sets up a family of test particles near his worldline. The deviation vector to any infinitesimally nearby particles satisfies the geodesic deviation equation:

$$\begin{aligned} &(\nabla_U \nabla_U S)^a = R(U,S,U)^a \\ \implies &\left(U^b \nabla_{e_b}\left(U^c \nabla_{e_c} S\right)\right)^a = R^a{}_{bcd} U^b U^c S^d \\ \implies &\left(U^b \nabla_{e_b}\left(U^c \nabla_{e_c} S\right)\right)_a = R_{abcd} U^b U^c S^d. \end{aligned} \tag{9.53}$$

Let us contract it with e_α^a and use the fact that $(\nabla_U e_i)^a = 0$ to essentially push it inside the covariant derivative:

$$R_{abcd} e_\alpha^a U^b U^c S^d = U^b \nabla_b \left(U^c \nabla_c \left(e_\alpha^a S_a\right)\right). \tag{9.54}$$

Here $e_\alpha^a S_a$ is a scalar, and hence we see that, $\nabla_c(e_\alpha^a S_a) = \partial_c S_\alpha$. So,

$$U^b \nabla_b \left(U^c \nabla_c \left(e_\alpha^a S_a\right)\right) = U^b \nabla_b \left(\frac{\mathrm{d}x^c}{\mathrm{d}\tau}\frac{\partial S_\alpha}{\partial x^c}\right) = \frac{\mathrm{d}x^b}{\mathrm{d}\tau}\frac{\partial}{\partial x^b}\left(\frac{\mathrm{d}S_\alpha}{\mathrm{d}\tau}\right) = \frac{\mathrm{d}^2 S_\alpha}{\mathrm{d}\tau^2}. \tag{9.55}$$

Therefore,

$$\frac{\mathrm{d}^2 S_\alpha}{\mathrm{d}\tau^2} = R_{abcd} e_\alpha^a U^b U^c S^d = R_{abcd} e_\alpha^a U^b U^c e_\beta^d S^\beta, \tag{9.56}$$

where τ is the proper time and $S_\alpha = e_\alpha^a S_a$ is one of the components of S_a in the parallely transported frame. On the RHS, we have used $e_\beta^d S^\beta = S^d$.

In the linearized theory, R_{abcd} is a quantity of first order, so we only need to evaluate the other quantities to leading order, i.e. we can evaluate them as if the spacetime were flat. The extra contributions would be negligible since they would end up being $\mathcal{O}(h^2)$ or $\mathcal{O}(|\partial_i h|^2)$.

Let us assume that the observer is at rest in the "almost-inertial" coordinates, i.e. to leading order, $U^\mu = (1,0,0,1)$. Hence, from 9.56, we see that the terms on RHS become 0 when $b = c \neq 0$. Hence, when $b = c = 0$, we have

$$\frac{\mathrm{d}^2 S_\alpha}{\mathrm{d}\tau^2} = R_{abcd} e_\alpha^a U^b U^c e_\beta^d S^\beta \approx R_{a00d} e_\alpha^a e_\beta^d S^\beta. \tag{9.57}$$

Now,

$$R_{abcd} = \frac{1}{2}\left(h'_{ad,bc} + h'_{bc,ad} - h'_{bd,ac} - h'_{ac,bd}\right) \implies R_{\mu 00\sigma} = \frac{1}{2} h'_{\mu\sigma,00}. \tag{9.58}$$

Therefore,

$$\begin{aligned} \frac{\mathrm{d}^2 S_\alpha}{\mathrm{d}\tau^2} &\approx R_{\mu 00\nu} e_\alpha^\mu e_\beta^\sigma S^\beta \approx \frac{1}{2} h'_{\mu\sigma,00} e_\beta^\sigma S^\beta. \\ \therefore \frac{d^2 S_\alpha}{d\tau^2} &\approx \frac{1}{2} h'_{\mu\nu,00} e_\beta^\nu S^\beta. \end{aligned} \tag{9.59}$$

In Minkowski spacetime, we could take e_i^a to be aligned with the X, Y, Z axes respectively, i.e. $e_1^\mu = (0,1,0,0), e_2^\mu = (0,0,1,0)$ and $e_3^\mu = (0,0,0,1)$ (in a basis, locally). We can use the same result here as we only need to evaluate e_α^μ upto leading order. Using $h'_{0\mu} = h'_{3\mu} = 0$, we see that,

$$\frac{\mathrm{d}^2 S_0}{\mathrm{d}\tau^2} = \frac{1}{2}\frac{\partial^2 h'_{\mu\nu}}{\partial t^2}\delta_0^\mu \delta_\beta^\nu S^\beta = \frac{1}{2}\frac{\partial^2 h'_{0\nu}}{\partial t^2}\delta_\beta^\nu S^\beta = -\frac{1}{2}k_0^2 h'_{0\nu} S^\nu = 0, \tag{9.60}$$

since $h'_{0\mu} = 0$. Similarly,

$$\frac{\mathrm{d}^2 S_3}{\mathrm{d}\tau^2} = \frac{1}{2}\frac{\partial^2 h'_{\mu\nu}}{\partial t^2}\delta_3^\mu \delta_\beta^\nu S^\beta = \frac{1}{2}\frac{\partial^2 h'_{3\nu}}{\partial t^2}\delta_\beta^\nu S^\beta = -\frac{1}{2}k_3^2 h'_{3\nu} S^\nu = 0 \tag{9.61}$$

to this order of approximation. Hence, the observer sees no relative acceleration in the t and z axes (the z axis is thought to be the axis along which the gravitational wave is propagating). We can choose our "almost-inertial" coordinates so that the observer has coordinates $x^\mu = (\tau, 0, 0, 0)$ (i.e. $t = \tau$ up to leading order along the observer's worldline).

For a $+$-polarized wave, we can let $H_\times = 0$, and then,

$$\begin{aligned}\frac{\mathrm{d}^2 S_1}{\mathrm{d}\tau^2} &= \frac{1}{2}\frac{\partial^2 h'_{\mu\nu}}{\partial t^2}\delta_1^\mu \delta_\beta^\nu S^\beta = \frac{1}{2}\frac{\partial^2 h'_{1\nu}}{\partial t^2}\delta_\beta^\nu S^\beta \\ &= \frac{1}{2}\frac{\partial^2}{\partial t^2}(h'_{11})S^1 = \frac{1}{2}\frac{\partial^2}{\partial t^2}(H'_{11}e^{ik_\rho x^\rho})S^1 \\ &= -\frac{1}{2}k_0^2 H_+ \eta^{1\sigma} S_\sigma = -\frac{1}{2}k_0^2 H_+ \eta^{11} S_1 \\ &= -\frac{1}{2}k_0^2 H_+ S_1. \end{aligned} \tag{9.62}$$

In essence, we should have $\frac{\mathrm{d}^2 S_1}{\mathrm{d}\tau^2} = \mathrm{Re}\left(-\frac{1}{2}k_0^2 H_+ S_1\right)$. So

$$\frac{\mathrm{d}^2 S_1}{\mathrm{d}\tau^2} = -\frac{1}{2}\omega^2 \,\mathrm{Re}\left(|H_+|\, e^{i\,\mathrm{Arg}(H_+)} e^{i(\omega\tau)} S_1\right) = -\frac{1}{2}\omega^2 \,|H_+| \cos(\omega\tau - \alpha) S_1. \tag{9.63}$$

Similarly,

$$\begin{aligned}\frac{\mathrm{d}^2 S_2}{\mathrm{d}\tau^2} &= \frac{1}{2}\frac{\partial^2 h'_{\mu\nu}}{\partial t^2}\delta_2^\mu \delta_\beta^\nu S^\beta = \frac{1}{2}\frac{\partial^2 h'_{2\nu}}{\partial t^2}\delta_\beta^\nu S^\beta \\ &= \frac{1}{2}\frac{\partial^2}{\partial t^2}(h'_{22})S^1 = \frac{1}{2}\frac{\partial^2}{\partial t^2}\left(H'_{22}e^{ik_\rho x^\rho}\right)S^1 \\ &= -\frac{1}{2}k_0^2 (-H_+) \eta^{2\sigma} S_\sigma = \frac{1}{2}k_0^2 H_+ \eta^{22} S_2 \\ &= \frac{1}{2}k_0^2 H_+ S_2. \end{aligned} \tag{9.64}$$

In essence, similarly as before, we should have,

$$\frac{\mathrm{d}^2 S_2}{\mathrm{d}\tau^2} = \mathrm{Re}\left(\frac{1}{2}k_0^2 H_+ S_2\right) = \frac{1}{2}\omega^2 \,\mathrm{Re}\left(|H_+|\, e^{i\,\mathrm{Arg}(H_+)} e^{i\omega\tau} S_2\right) = \frac{1}{2}\omega^2 \,|H_+| \cos(\omega\tau - \alpha) S_2. \tag{9.65}$$

So, we have got two differential equations: $\frac{\mathrm{d}^2 S_1}{\mathrm{d}\tau^2} = -\frac{1}{2}\omega^2 \,|H_+| \cos(\omega\tau - \alpha) S_1$, and $\frac{\mathrm{d}^2 S_2}{\mathrm{d}\tau^2} = \frac{1}{2}\omega^2 \,|H_+| \cos(\omega\tau - \alpha) S_2$, where $|H_+|$ is small. The exact solutions of these differential equations involve Mathieu functions. Some plots of the solution of the ODE (with some initial value) are shared below (Figure 9.1 to Figure 9.6). They were plotted using Maple 2020.1.

Since $|H_+|$ is small, we see that upto $\mathcal{O}\left(|H_+|^2\right)$,

$$S_1(\tau) \approx \bar{S}_1 \left(1 + \frac{1}{2}|H_+| \cos(\omega\tau - \alpha)\right) \tag{9.66}$$

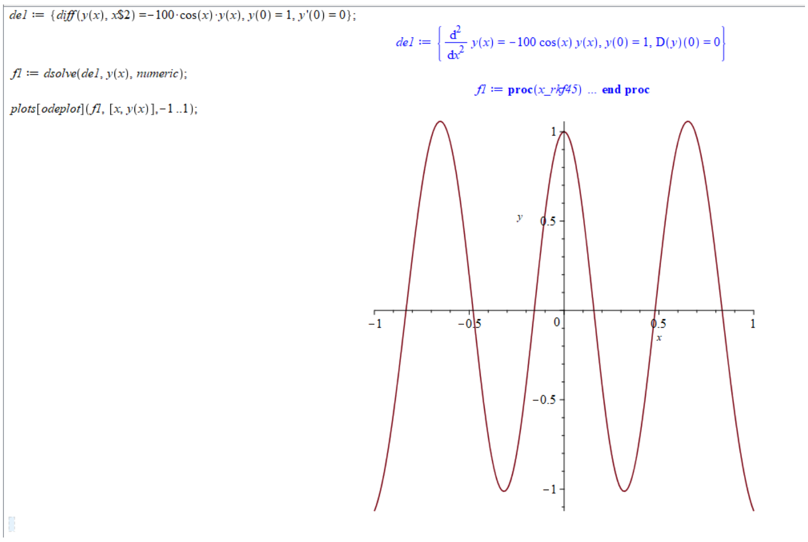


Figure 9.1: $S_1(\tau)$ when $\omega = 1$, $H_+ = 100$, $S_1(0) = 1$, and $\frac{\mathrm{d}S_1}{\mathrm{d}\tau}(0) = 0$.

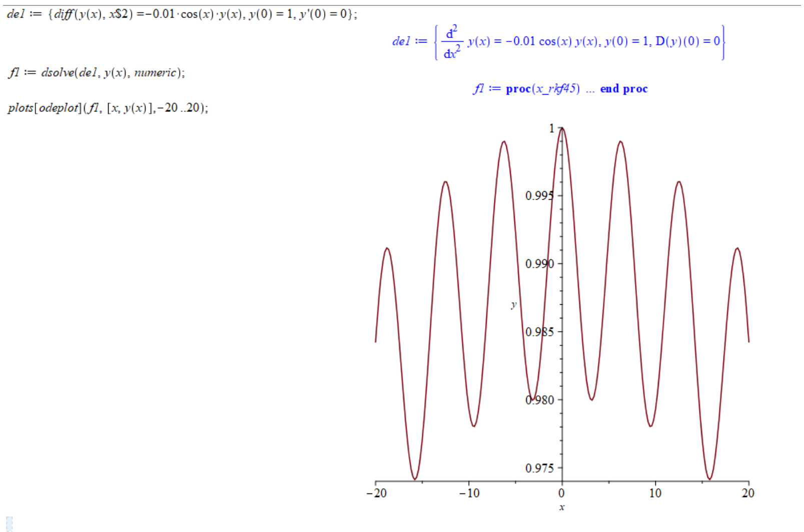


Figure 9.2: $S_1(\tau)$ when $\omega = 1$, $H_+ = 0.01$, $S_1(0) = 1$, and $\frac{\mathrm{d}S_1}{\mathrm{d}\tau}(0) = 0$.

is a solution where $\bar{S}_1$ is a constant. We verify this by plugging it into the differential equation. Indeed,

$$\begin{aligned}\frac{\mathrm{d}^2 S_1}{\mathrm{d}\tau^2} &\approx \frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\left(\bar{S}_1\left(1+\frac{1}{2}|H_+|\cos(\omega\tau-\alpha)\right)\right)\\ &= \frac{1}{2}\bar{S}_1|H_+|\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}(\cos(\omega\tau-\alpha))\\ &= -\frac{1}{2}\omega^2|H_+|\cos(\omega\tau-\alpha)\bar{S}_1. \end{aligned}\tag{9.67}$$

Also,

$$\begin{aligned}-\frac{1}{2}\omega^2|H_+|\cos(\omega\tau-\alpha)S_1 &\approx -\frac{1}{2}\omega^2|H_+|\cos(\omega\tau-\alpha)\left(\bar{S}_1+\frac{\bar{S}_1}{2}|H_+|\cos(\omega\tau-\alpha)\right)\\ &= -\frac{1}{2}\omega^2|H_+|\cos(\omega\tau-\alpha)\bar{S}_1+\mathcal{O}\left(|H_+|^2\right). \end{aligned}\tag{9.68}$$

So we see that, $S_1(\tau) \approx \bar{S}_1\left(1+\frac{1}{2}|H_+|\cos(\omega\tau-\alpha)\right)$ works upto leading order term. Similarly, we would have $S_2(\tau) \approx \bar{S}_2\left(1-\frac{1}{2}|H_+|\cos(\omega\tau-\alpha)\right)$ to work upto leading order term for the differential

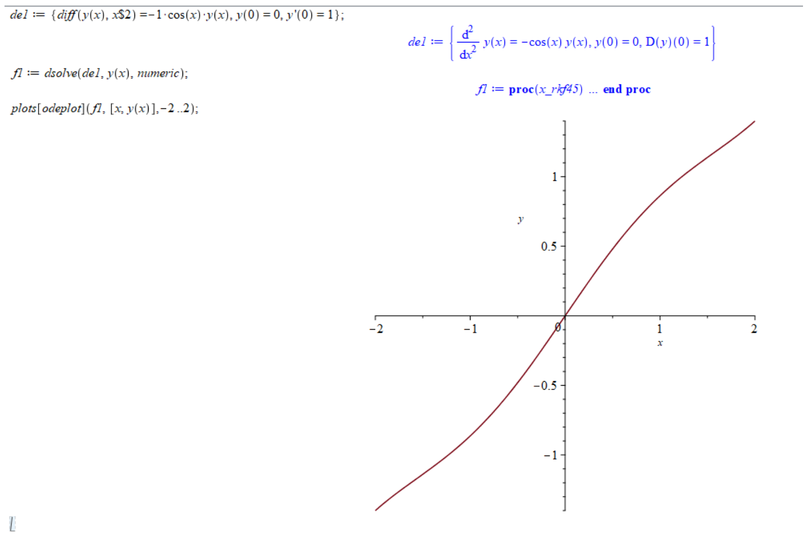


Figure 9.3: $S_1(\tau)$ when $\omega = 1$, $H_+ = 1$, $S_1(0) = 0$, and $\frac{\mathrm{d}S_1}{\mathrm{d}\tau}(0) = 1$.

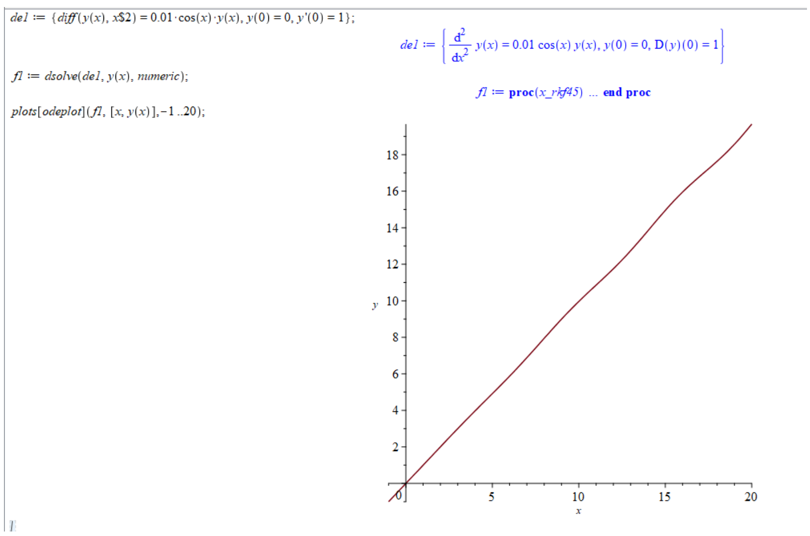


Figure 9.4: $S_2(\tau)$ when $\omega = 1$, $H_+ = 0.01$, $S_2(0) = 0$, and $\frac{\mathrm{d}S_2}{\mathrm{d}\tau}(0) = 1$. Notice that this is not entirely a straight line. There are some "wiggles".

equation we got for S_2. We see that,

$$\begin{aligned}\left(\frac{S_1(\tau)}{\bar{S}_1}\right)^2 + \left(\frac{S_2(\tau)}{\bar{S}_2}\right)^2 &= \left(1 + \frac{1}{2}|H_+|\cos(\omega\tau - \alpha)\right)^2 + \left(1 - \frac{1}{2}|H_+|\cos(\omega\tau - \alpha)\right)^2 \\ &= 1 + |H_+|\cos(\omega\tau - \alpha) + 1 - |H_+|\cos(\omega\tau - \alpha) + \mathcal{O}\left(|H_+|^2\right) \\ &\approx 2,\end{aligned} \tag{9.69}$$

ignoring terms $\mathcal{O}\left(|H_+|^2\right)$. Therefore,

$$\left(\frac{S_1(\tau)}{\sqrt{2}\bar{S}_1}\right)^2 + \left(\frac{S_2(\tau)}{\sqrt{2}\bar{S}_2}\right)^2 \approx 1. \tag{9.70}$$

This is (almost) an ellipse. Thus, particles initially separated in the x-direction will oscillate back and forth in the x-direction, and likewise for those with an initial y-separation. Thus, if we start a ring of stationary particles in the xy-plane, as the wave passes, they will bounce back and forth in the shape of a "+". See Figure 9.7.

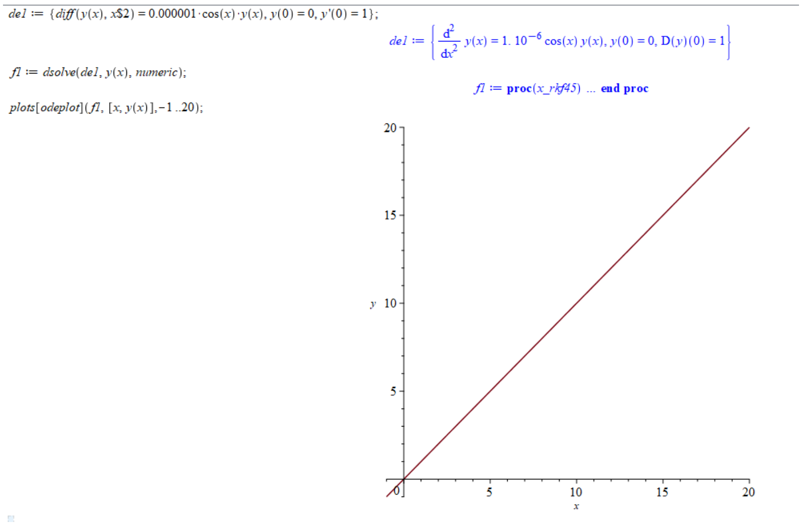


Figure 9.5: $S_2(\tau)$ when $\omega = 1$, $H_+ = 0.000001$, $S_2(0) = 0$, and $\frac{\mathrm{d}S_2}{\mathrm{d}\tau}(0) = 1$. No wonder this is a linearized theory.

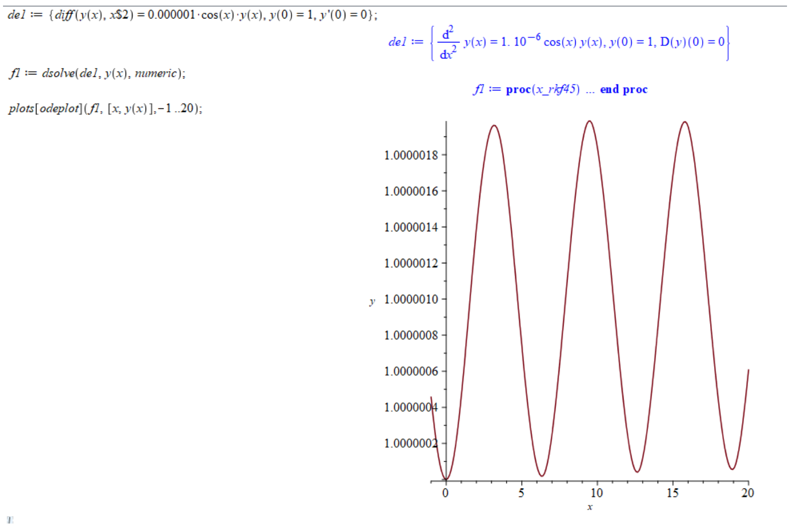


Figure 9.6: $S_2(\tau)$ when $\omega = 1$, $H_+ = 0.000001$, $S_2(0) = 1$, and $\frac{\mathrm{d}S_2}{\mathrm{d}\tau}(0) = 0$.

For a $\times$-polarised wave, $H_+ = 0$, and then we would have:

$$\begin{aligned}
\frac{\mathrm{d}^2 S_1}{\mathrm{d}\tau^2} &= \frac{1}{2}\frac{\partial^2 h'_{\mu\nu}}{\partial\tau^2}\delta^\mu_1\delta^\nu_\beta S^\beta = \frac{1}{2}\frac{\partial^2 h'_{1\nu}}{\partial\tau^2} S^\nu \\
&= \frac{1}{2}\frac{\partial^2}{\partial t^2}\left(h'_{12}\right) S_2 = \frac{1}{2}\frac{\partial^2}{\partial t^2}\left(H'_{12} e^{ik_\rho x^\rho}\right) S_2 \\
&= \frac{1}{2}k_0^2 H_\times \eta^{2\sigma} S_\sigma = \frac{1}{2}k_0^2 H_\times \eta^{22} S_2 \\
&= \frac{1}{2}k_0^2 H_\times S_2.
\end{aligned} \tag{9.71}$$

Again, similarly as before, we should have

$$\begin{aligned}
\frac{\mathrm{d}^2 S_1}{\mathrm{d}\tau^2} &= \mathrm{Re}\left(\frac{1}{2}k_0^2 H_\times S_2\right) = \frac{1}{2}\omega^2\,\mathrm{Re}\left(|H_\times| e^{i\,\mathrm{Arg}(H_\times)} e^{i\omega\tau} S_2\right) \\
&= \frac{1}{2}\omega^2\,|H_\times|\cos(\omega\tau - \alpha) S_2.
\end{aligned} \tag{9.72}$$

Similarly, for S_2, we will have

$$\frac{\mathrm{d}^2 S_2}{\mathrm{d}\tau^2} = \frac{1}{2}\omega^2\,|H_\times|\cos(\omega\tau - \alpha) S_1. \tag{9.73}$$

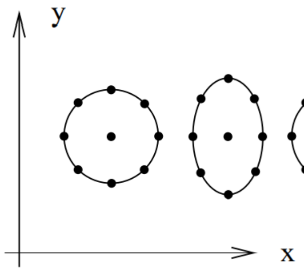


Figure 9.7: Effect of a +-polarised Gravitational Wave as it passes through a ring of stationary particles in the xy-plane

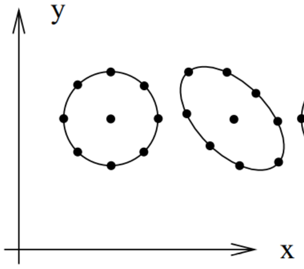


Figure 9.8: Effect of a ×-polarised Gravitational Wave as it passes through a ring of stationary particles in the xy-plane

We have got two coupled differential equations. We see that,

$$\frac{\mathrm{d}^2(S_1+S_2)}{\mathrm{d}\tau^2} = \frac{1}{2}\omega^2 \left|H_\times\right| \cos(\omega\tau-\alpha)\left(S_1+S_2\right) \tag{9.74}$$

Therefore,

$$S_1+S_2 \approx \bar{S}_{12}\left(1+\frac{1}{2}\left|H_\times\right|\cos(\omega\tau-\alpha)\right), \tag{9.75}$$

ignoring terms $\mathcal{O}\left(\left|H_\times\right|^2\right)$. Also,

$$\frac{\mathrm{d}^2\left(S_1-S_2\right)}{\mathrm{d}\tau^2} = -\frac{1}{2}\omega^2\left|H_\times\right|\cos(\omega\tau-\alpha)\left(S_1-S_2\right). \tag{9.76}$$

Therefore, again ignoring terms $\mathcal{O}\left(\left|H_\times\right|^2\right)$,

$$S_1-S_2 \approx \bar{S}_{21}\left(1-\frac{1}{2}\left|H_\times\right|\cos(\omega\tau-\alpha)\right). \tag{9.77}$$

After solving for $S_1(\tau)$ and $S_2(\tau)$, we get

$$S_1(\tau) = \frac{\bar{S}_{12}+\bar{S}_{21}}{2}+\frac{\bar{S}_{12}-\bar{S}_{21}}{4}\left|H_\times\right|\cos(\omega\tau-\alpha), \tag{9.78}$$

$$S_2(\tau) = \frac{\bar{S}_{12}-\bar{S}_{21}}{2}+\frac{\bar{S}_{12}+\bar{S}_{21}}{4}\left|H_\times\right|\cos(\omega\tau-\alpha). \tag{9.79}$$

We see that the solutions, upto first order in $|H_\times|$, are:

$$S_1(\tau) = C+\frac{D}{2}\left|H_\times\right|\cos(\omega\tau-\alpha) \text{ and } S_2(\tau) = D+\frac{C}{2}\left|H_\times\right|\cos(\omega\tau-\alpha). \tag{9.80}$$

It will also fall into an equation of an ellipse whose major and minor axes are a bit 'crossed'/rotated in the coordinate we are using. In this case, the circle/ring of particles would bounce back and forth in the shape of a "×". See Figure 9.8. [Hence, the notation "H_+" and "$H_\times$" should make sense now.]

We could also consider right and left-handed circularly polarized modes by defining:

$$H_R = \frac{1}{\sqrt{2}}\left(H_+ + iH_\times\right) \text{ and } H_L = \frac{1}{\sqrt{2}}\left(H_+ - iH_\times\right) \tag{9.81}$$

The effect of a pure H_R wave would be to rotate the particles in a right-handed sense (see Figure 9.9), and similarly for the left-handed mode H_L. [Note that the individual particles do not travel around the ring; they just move in epicycles.]

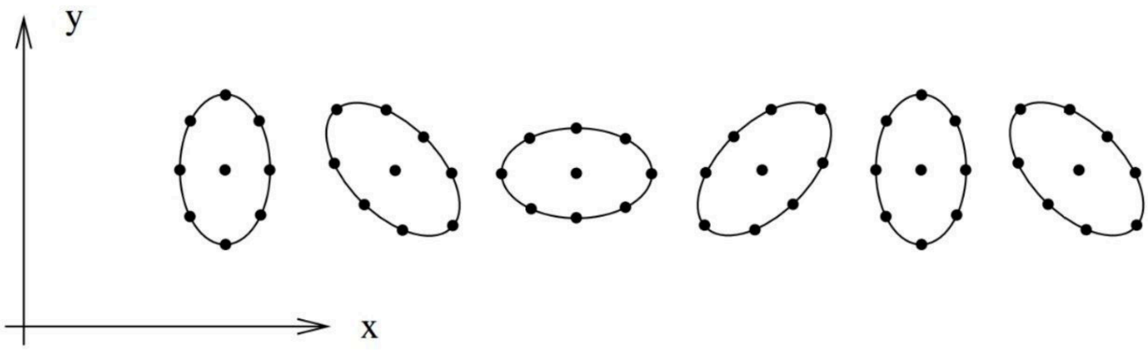


Figure 9.9: Effect of an H_R gravitational wave to an bunch of particles put on an ellipse. The effect is as though the ellipse gets being rotated in a right hand sense.

Now, we want to see that the two polarizations are 45° apart. Let us rotate our local "almost inertal" coordinates clockwise by 45° about the z-axis. Then,

$$(t, x, y, z) \to (t', x', y', z') = \left(t, \frac{x}{\sqrt{2}} - \frac{y}{\sqrt{2}}, \frac{x}{\sqrt{2}} + \frac{y}{\sqrt{2}}, z\right). \tag{9.82}$$

In this coordinate,

$$H'_{\mu\nu} = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu} H_{\rho\sigma}. \tag{9.83}$$

From 9.82, we get

$$x = \frac{x'}{\sqrt{2}} + \frac{y'}{\sqrt{2}}, \text{ and } y = -\frac{x'}{\sqrt{2}} + \frac{y'}{\sqrt{2}}. \tag{9.84}$$

Therefore,

$$\begin{aligned} H'_{11} &= \frac{\partial x}{\partial x'}\frac{\partial x}{\partial x'}H_{11} + \frac{\partial x}{\partial x'}\frac{\partial y}{\partial x'}H_{12} + \frac{\partial y}{\partial x'}\frac{\partial x}{\partial x'}H_{21} + \frac{\partial y}{\partial y'}\frac{\partial x}{\partial x'}H_{22} \\ &= \frac{1}{2}H_+ - \frac{1}{2}H_\times - \frac{1}{2}H_\times + \frac{1}{2}\left(-H_+\right) = -H_\times. \end{aligned} \tag{9.85}$$

$$\begin{aligned} H'_{12} &= \frac{\partial x}{\partial x'}\frac{\partial x}{\partial y'}H_{11} + \frac{\partial x}{\partial x'}\frac{\partial y}{\partial y'}H_{12} + \frac{\partial y}{\partial x'}\frac{\partial x}{\partial y'}H_{21} + \frac{\partial y}{\partial y'}\frac{\partial x}{\partial y'}H_{22} \\ &= \frac{1}{2}H_+ + \frac{1}{2}H_\times - \frac{1}{2}H_\times - \frac{1}{2}\left(-H_+\right) = H_+. \end{aligned} \tag{9.86}$$

So, in $H \to H'$, we see that the (independent) components have switched places so that the polarizations have switched places. What we called $\times$-polarization initially, after rotating clockwise by 45° around the z-axis, will now seem to be a $+$-polarization and vice versa for $+$ to $\times$.