2. Geometry

2.0. The Special and General Theories of Relativity

2.0.1. The Special Theory

2.0.1.a. Newton's Laws (1686)

1. Every body continues in its state of rest, or of uniform motion in a right (straight) line, unless it is compelled to change that state by forces impressed on it.
2. The change of motion is proportional to the motive force impressed; and is made in the direction of the right line in which that force is impressed.

2.0.1.b. Galilean Relativity

An **inertial frame** is a reference frame in which Newton's 1st law holds. Any 2 inertial frames must either be at rest or moving with uniform velocity with respect to each other. Consider 2 inertial frames S and S' that coincide at $t = t' = 0$. Let S' be moving with respect to S with velocity $\mathbf{v} = (v, 0, 0)$ (see Fig.2.1). According to Galileo, the coordinates for a given event with respect to these frames are related by

$$x' = x - vt \qquad y' = y \qquad z' = z \qquad t' = t \tag{2.1}$$

or

$$\mathbf{x}' = \mathbf{x} - \mathbf{v}t \quad = (x - vt, y, z) \qquad \text{and} \qquad t' = t$$

If the velocity of the particle with respect to S is $\mathbf{u} = \frac{d\mathbf{x}}{dt}$, its velocity with respect to S' is

$$\mathbf{u}' = \frac{d\mathbf{x}'}{dt'} \quad = \frac{d}{dt}(\mathbf{x} - \mathbf{v}t) = \mathbf{u} - \mathbf{v} \quad = (u_x - v, u_y, u_z)$$

Thus, its acceleration as seen in these frames are equal:

$$\mathbf{a}' = \frac{d\mathbf{u}'}{dt'} \quad = \frac{d}{dt}(\mathbf{u} - \mathbf{v}) \quad = \frac{d\mathbf{u}}{dt} \quad = \mathbf{a}$$

2.0.1.c. Special Relativity

The Galilean relativity failed to provide a consistent account of various experimental results concerning the properties of light. This problem was resolved by Einstein's theory of special relativity, which proposed

1. The speed c of light in vaccum is the same in any inertial frame.
2. The laws of physics are the same in all inertial frames.

Application of this to the Maxwell's equations led to the replacing of the Galilean relativity eq(2.1) by [see Exercise 2.1]

$$x' = \gamma\left(x - vt\right) \qquad y' = y \qquad z' = z \qquad ct' = \gamma\left(ct - \beta x\right) \tag{2.2}$$

where

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{1}{\sqrt{1-\beta^2}} \qquad \text{and} \qquad \beta = \frac{v}{c}$$

Eq(2.2) is known as the **Lorentz transformation**. That (2.2) does indeed satisfies proposal 1 can be seen by applying it to

$$x^2 + y^2 + z^2 = ct^2$$

so that

$$x'^2 + y'^2 + z'^2 - c^2 t'^2 = \gamma^2\left(x - vt\right)^2 + y^2 + z^2 - \gamma^2\left(ct - \beta x\right)^2$$

$$= \gamma^2\left(x - vt + ct - \beta x\right)\left(x - vt - ct + \beta x\right) + y^2 + z^2$$

$$= \gamma^2\left[\left(1-\beta\right)x + ct\left(1-\beta\right)\right]\left[\left(1+\beta\right)x - ct\left(1+\beta\right)\right] + y^2 + z^2$$

$$= \left(x + ct\right)\left(x - ct\right) + y^2 + z^2$$

$$= x^2 + y^2 + z^2 - ct^2 \tag{2.2a}$$

This means that light emitted from a point source at the origin $\mathbf{x} = \mathbf{x}' = \mathbf{0}$ at time $t = t' = 0$ has wavefronts that are spherical and propagating with speed c in both, and hence all inertial, frames.

2.0.1.d. Spacetimes

The importance of eqs(2.1, 2) is that they contain information about the structure of space and time that is independent of any frame of reference. For example, consider 2 events with spacetime coordinates $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$ relative to S and $(\mathbf{x}'_1, t'_1)$ and $(\mathbf{x}'_2, t'_2)$ relative to S'.

According to the Galilean transformation,

$$t'_2 - t'_1 = t_2 - t_1 \quad \equiv \Delta t \qquad (2.3a)$$

so that events that are simultaneous in S will remain so in S'. Furthermore, the spatial separations of simultaneous events are also the same in both frames, i.e., if $t_2 = t_1$,

$$\left|\mathbf{x}'_2 - \mathbf{x}'_1\right| = \left|\mathbf{x}_2 - \mathbf{v}t_2 - (\mathbf{x}_1 - \mathbf{v}t_1)\right| \quad = \left|\mathbf{x}_2 - \mathbf{x}_1\right| \quad \equiv \Delta r \qquad (2.3b)$$

Thus, measurements done by any inertial observer give the same time interval between any 2 given events, as well as the same spatial separation between 2 given simultaneous events. Therefore, Δt and Δr possess real physical meanings that are independent of coordinates.

According to the Lorentz transformation, neither Δt nor Δr is the same for all inertial observers. In their stead, there is a frame- independent quantity called the **proper time interval** $\Delta\tau$ and defined by

$$c^2 \Delta\tau^2 = c^2 \Delta t^2 - \Delta r^2 \qquad (2.3)$$

where $\Delta\tau^2$ is a shorthand for $(\Delta\tau)^2$. [See (2.2a) for proof of frame- independence]

These frame- (coordinate-) independent properties are fundamental characteristics of spacetime (or, more accurately, the models we use to describe it). Spacetimes that satisfy (2.3a and b) are called **Galilean spacetimes**. Those that satisfy (2.3) are **Minkowski spacetimes**. Used this way, c in (2.3) is no longer the velocity of something: it is a conversion factor between the spatial and temporal dimensions of spacetime. As a theory of spactime, special relativity also stands above the domains of mechanics and electromagnetics.

According to Einstein, the geometric structure of spacetime restricts the possible forms of physical laws. This is stated as proposal 2 in §2.0.1.b. Mathematically, this means any set of equations that expresses a physical law should be **covariant**, i.e., has the same form in every inertial system.

2.0.1.e. Newtonian Gravitation

According to the Newtonian theory of gravity, the motion of 2 objects with masses m_1 and m_2, and positions $\mathbf{x}_1$ and $\mathbf{x}_2$, are governed by

$$m_1 \frac{d^2\mathbf{x}_1}{dt^2} = Gm_1m_2 \frac{\mathbf{x}_2 - \mathbf{x}_1}{\left|\mathbf{x}_2 - \mathbf{x}_1\right|^3} \qquad (2.4)$$

$$m_2 \frac{d^2\mathbf{x}_2}{dt^2} = Gm_2m_1 \frac{\mathbf{x}_1 - \mathbf{x}_2}{\left|\mathbf{x}_1 - \mathbf{x}_2\right|^3}$$

where $G \simeq 6.67 \times 10^{-11}\ Nm^2kg^{-2}$ is Newton's **gravitational constant**. It is straightforward to show that (2.4) is covariant under a Galilean transformation, but not so under a Lorentz transformation. Thus, (2.4) cannot be a valid theory in Minkowski spacetime.

2.0.2. The General Theory

The following is a summary of chapter 9 of "Introducing Einstein's Relativity" by R. D'Inverno.

There are perhaps 5 principles that lead to the general theory:

1. **Mach's principle**:

   Matter distribution of the universe determines inertial forces (geometry).

2. **Principle of equivalence**:

   a) Strong principle: the motion of a gravitational test particle is independent of its mass and composition (gravitation mass equal to inertial mass).

   b) Weak principle: the gravitaional field couples to everything (since mass = energy).

   c) A free-falling frame is localy indistinguishable from an inertial frame.

   d) A frame under constant acceleration is locally indistinguishable from one at rest in a gravitational field.

3. **Principle of general relativity**:

   All observers are equivalent.

   Principle of general covariance:

   Equations of physics must be in tensorial form.

4. **Principle of minimal gravitational coupling**:

   No terms containing the cuvature tensor explicitly should be added in making the transition from the special to the general theory.

5. **Correspondence principle**:

   The general theory should reduce to the special and the Newtonian theories under appropriate conditions.

The exact phrasings of, as well as the significance attributed to, these principles vary with authors. However, it is generally accepted that the principle of equivalence is the key assumption.

A simple starting point of the general theory is as follows. In the special theory, the proper time interval $d\tau$ is defined in terms of Cartesian coordinates as

$$c^2 (d\tau)^2 = c^2 (dt)^2 - (d\mathbf{r})^2 \qquad (2.6)$$

In terms of some general coordinates $\{x^\mu \, ; \, \mu = 0,1,2,3\}$, this becomes

$$c^2 \left(d\tau\right)^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu}\left(x\right) dx^\mu dx^\nu \qquad (2.7)$$

where the **metric tensor** $g_{\mu\nu}$ is related to Minkowski metric tensor

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \qquad (2.8)$$

by a similarity transformation corresponding to the coordinate transformation $\left(t,\mathbf{r}\right) \rightarrow \left(x^\mu\right)$. In the general theory, the principle of equivalence asserts that gravitational forces can be treated locally as an inertial forces, i.e., $g_{\mu\nu}\left(x\right)$ takes the form (2.8) locally in a free-falling frame. Next, in accordance with Mach's principle, $g_{\mu\nu}\left(x\right)$ is determined by the distribution of gravitational sources. The last 3 principles then serve as the guidance for writing the **field equations** that relates g to the sources.

2.1. Spacetime as a Differentiable Manifold

2.1.1. Topology of the Real Line *R* and R^d

Topology is the mathematical tool for studying the basic properties of a **deformable continuum** that are independent of the concept of "lengths". Of particular interest to physicists are differentiable manifolds, which are smooth (differentiable) topological spaces.

1. A system $\mathcal{U}$ of subsets of a set X defines a **topology** on X if $\mathcal{U}$ contains
   a) $\varnothing$ and X.
   b) the union of every one of its subsystems.
   c) the intersection of every one of its *finite* subsystems.

   The sets (subsystems) in $\mathcal{U}$ are called **open sets** of the **topological space** $(X,\mathcal{U})$, often abbreviated to X.
2. The intersection of an infinite sequence of subsystems is a point. Thus, only a finite number of intersections is allowed in criterion c so that we have the option to define a point as not-open (closed). This is important since, later on, we will define the concept of continuity in terms of open sets so that discontinuities are related to closed sets.
3. Let X be any non-empty set.
   a) The topology with open sets $\varnothing$ and X is called **trivial**.
   b) The topology with all subsets of X, $\varnothing$ and X included, is called **discrete**.
4. The topology on $\mathbb{R}$ using unions of open intervals $a < x < b$ as open sets is called the **usual topology** on $\mathbb{R}$.
5. The usual topology on $\mathbb{R}^2$ using unions of open rectangular regions (x, y) with $a < x < b$ and $c < y < d$ as open sets. Generalization to the case $\mathbb{R}^d$ is obvious.
6. A **neighborhood** of a point $x \in X$ is a set $N(x)$ containing an open set which contains x. The neighborhood $N(A)$ of a set $A \subset X$ is similarly defined.
7. A function $f : X \to Y$ is **continuous** if for any open set U in Y, the inverse image $V = f^{-1}(U) \subset X$ is open in X. [see Fig.2.4. for example of

discontinuity in the function $f:\mathbb{R}\to\mathbb{R}$].

8. Some observations on continuous mappings:
   a) The concept of continuity as defined in (7) is for the *entire* mapping, not for any specific point. However, the equivalence of this definition to the more familiar δ–ε definition used in elementary calculus can be easily demonstrated using the concept of neighborhoods.
   b) The seemingly more straightforward definition

      A function $f:X\to Y$ is **continuous** if for any open set U in X, the image $V=f(U)\subset Y$ is open in Y.

      is not acceptable because it treats a constant function as *not* continuous. The definition (7) does not have this kind of problem because a mapping that assigns many values in Y to a point in X, e.g. a vertical line in the graph $f(x)$ vs x, is by definition not a function.
   c) For the case where many points in X have the same value U in Y, e.g., periodic functions, the inverse image $f^{-1}(U)$ will be a union of disjoint sets. The definition (7) can remain valid only if we require the unions of open sets to be an open set. This is part of the reason for including the latter in the definition of topology.
   d) We require the empty set Φ to be open since for a function that is not onto, we have $f^{-1}(U)=\Phi$ for some $U\subset Y$.

9. Note that every function on a discrete topological space is continuous.
10. A topology imposes 2 kinds of structures on the space:
   a) **Local topology** determines how open sets are fit together inside any globally small region. Notions like continuity are defined here.
   b) **Global topology** determines how the open sets cover the whole space. For example, the plane $\mathbb{R}^2$, the sphere S^2 and the torus T^2 have the same local topology but different global ones.

2.1.2. Differentiable Spacetime Manifold

Let $\mathbb{R}^n$ be the set of all *n*-tuples of real numbers $\left(x^1,\cdots,x^n\right)$. A set *M* of points is a (**topological**) **manifold** if each point *P* in it has an open neighborhood *U* **homeomorphic** to some open set *V* in $\mathbb{R}^n$. In other words, there is a **bi-continuous** (map and inverse map both continuous) **bijection** (1-1 onto map)

$$\phi : U \to V \qquad \text{by} \qquad P \mapsto \phi(P) = \left(x^1(P),\cdots,x^n(P)\right)$$

for all *P* in *M*. The *n* numbers $x^j(P)$ are called the **coordinates** of *P* and *n* is the **dimension** of *M*. Thus, the topology of *M* is the same as $\mathbb{R}^n$ locally.

The pair (U,ϕ) is called a **chart**, or a **local coordinate system**. An **atlas** on *M* is a set $\{(U_\alpha,\phi_\alpha)\}$ of charts so that the domains $\{U_\alpha\}$ covers *M*, i.e., every *P* is in some U_α. For reasons of compatibility, an atlas of class C^k requires the maps

$$\phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha\left(U_\alpha \cap U_\beta\right) \to \phi_\beta\left(U_\alpha \cap U_\beta\right) \qquad \text{(a)}$$

to be maps of class C^k. Note that $\phi_\beta \circ \phi_\alpha^{-1}$ is a map between open sets of $\mathbb{R}^n$. In fact, it represents a **coordinate transformation** for points in the overlap region $U_\alpha \cap U_\beta$ of two coordinate systems given by ϕ_α and ϕ_β. A manifold with an atlas of class C^k is said to be a C^k manifold. Those with $k > 1$ are called **differentiable manifolds**. For convenience, we shall deal only with C^∞ or C^ω manifolds.

2.1.3. Summary and Examples

In essence, a C^∞ manifold is a "smooth" continuum that is locally the same as $\mathbb{R}^n$.

Example 1

Consider the manifold *M* indicated by the flat 2-D rectangle in Fig.2.7(a). Since a manifold must be an open set, points on the boundaries of the rectangle are excluded from *M*. Consider the circular open region *U* inside *M*. An orthogonal coordinate system $\left(x^1, x^2\right)$ with grids as shown in Fig.2.7(a) maps it to the shaded open area *V* in $\mathbb{R}^2$ shown in Fig.2.7(b). The differences in shapes and sizes should be noted between *U* and *V*, which are homeomorphic to each other. This means, among other things, that the distances between points, which should be independent of the particular coordinate system used, can only be defined by imposing an additional structure, called a **metric**, on the manifold.

Example 2

Consider polar coordinates $\left(x^1, x^2\right) = \left(r, \phi\right)$ for the circular open region *U* in example 1. The coordinate grids are shown in Fig.2.8(a). Since every point in *U* and its boundary can be described by $0 \le r \le 4$ and $0 \le \phi \le 2\pi$, their images in $\mathbb{R}^2$ are in the shaded rectangle in Fig.2.8(b). On the other hand, a coordinate patch must map open regions 1-1 onto open regions in $\mathbb{R}^n$. Therefore, the boundary lines $\left(0, x^2\right)$, $\left(4, x^2\right)$, $\left(x^1, 0\right)$ and $\left(x^1, 2\pi\right)$ must be excluded from the image *V′* of points in the region *U′* covered by the coordinate patch. This in turn implies the inverse images of these lines must be excluded from *U′*. Thus, points on the line $\phi = 0$ must be excluded from the coordinate patch *U′*. In other words, *U* cannot be covered by the coordinate patch. With respect to such coordinates, a perfectly smooth function on *U* may become "artificially" non-analytic for $\phi = 0$. Since such behavior is entirely the fault of the coordinate system used, it is also "removable" by suitable means.

2.2. Tensors

2.2.1. Methods

Ways to express physical laws without making special assumptions about the coordinate system:

1. **Tensor anaylsis**: equations are written in a form that applies to any coordinate system. Here, one deals always with the components of tensors.
2. **Differential geometry**: equations are written without reference to any coordinate system. Here, one deals with each tensor as a single quantity.

2.2.2. Classification of Tensors

Tensors are classified by their ranks $\begin{pmatrix} a \\ b \end{pmatrix}$.

2.2.2.1. Scalars

Rank $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ tensors are called **scalars**.

A **scalar field** is a real-valued function

$$f : M \to R \quad \text{with} \quad P \mapsto f(P) \quad \text{or} \quad x^{\mu} \mapsto f(x^{\mu})$$

such that, under a coordinate transformation $x^{\mu} \to x^{\mu'}$ so that $f \to f'$, we have

$$f'(x^{\mu'}) = f(x^{\mu}) = f(P) \qquad (2.9)$$

2.2.2.2. Vectors

Rank $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ tensors are called **vectors** in differential geometry and **contravariant vectors** in tensor analysis. The prototypical vector is the **tangent** vector to a curve. Now, a **curve** (**path**) is a mapping from the real line to a manifold

$$C : R \to M \quad \text{with} \quad \lambda \mapsto C(\lambda) = \{x^\mu(\lambda)\}$$

where λ is called the **parameter** of the curve. The tangent vector of C is defined as

$$\mathbf{V}(\lambda) = \frac{d}{d\lambda} = \left\{\frac{dx^\mu}{d\lambda}\right\} = \sum_\mu \frac{dx^\mu}{d\lambda}\frac{\partial}{\partial x^\mu} = \sum_\mu V^\mu \partial_\mu$$

$$= V^\mu \partial_\mu = V^\mu X_\mu \qquad (2.11)$$

where $V^\mu = \frac{dx^\mu}{d\lambda}$, $\partial_\mu = \frac{\partial}{\partial x^\mu} = X_\mu$ and the last equality made use of the **Einstein summation notation**. Note that X_μ are the basis vectors for contravariant vectors.

For a differentiable scalar field $f(P)$, its derivative along the curve C is

$$\frac{df}{d\lambda} = \sum_\mu \frac{dx^\mu}{d\lambda}\frac{\partial f}{\partial x^\mu} = V^\mu \partial_\mu f \qquad (2.10)$$

Now, a curve can also be defined as the envelope of all of its tangent vectors $\frac{d}{d\lambda}$. For the convenience in algebraic manipulation, we shall treat each parametrization as a distinct curve. Thus, the envelopes of $\frac{d}{d\sigma}$ and $\frac{d}{d\lambda}$ always denote 2 different curves even if the trajectories parametrized by σ and λ are identical in *M*. One example is the **affinely related** parameters $\sigma = a\lambda + b$, where *a* and *b* are constants. Under a change of coordinates $x^\mu \to x^{\mu'}$, the chain rule of differentiation gives

$$X_{\mu'} = \frac{\partial}{\partial x^{\mu'}} = \frac{\partial x^\mu}{\partial x^{\mu'}}\frac{\partial}{\partial x^\mu} = \Lambda^\mu_{\mu'}\partial_\mu$$

where $\Lambda^\mu_{\mu'} = \frac{\partial x^\mu}{\partial x^{\mu'}}$. Hence,

$$\mathbf{V} = V^\mu \partial_\mu = V^\mu \frac{\partial x^{\mu'}}{\partial x^\mu}\frac{\partial}{\partial x^{\mu'}} = V^\mu \Lambda^{\mu'}_\mu \partial_{\mu'} = V^{\mu'}\partial_{\mu'}$$

where

$$\Lambda^{\mu'}_{\mu} = \frac{\partial x^{\mu'}}{\partial x^{\mu}} \qquad (2.13)$$

and $$V^{\mu'} = V^{\mu}\Lambda^{\mu'}_{\mu} \qquad (2.12)$$

Note that

$$\Lambda^{\sigma}_{\mu'}\Lambda^{\mu'}_{\mu} = \frac{\partial x^{\sigma}}{\partial x^{\mu'}}\frac{\partial x^{\mu'}}{\partial x^{\mu}} = \frac{\partial x^{\sigma}}{\partial x^{\mu}} = \delta^{\sigma}_{\mu} \qquad (2.14)$$

If we adopt the convention to represent a $\begin{pmatrix}1\\1\end{pmatrix}$ tensor *A* by a matrix **A** with components $a_{ij} = A^{i}_{j} = A^{i}{}_{j}$, eq(2.14) can be interpreted as saying the matrix *M* of components $m_{\sigma\mu'} = \Lambda^{\sigma}_{\mu'}$ is the inverse of the matrix *N* of components $n_{\mu'\mu} = \Lambda^{\mu'}_{\mu}$.

In particular, (2.14) means $MN = I$. Similarly,

$$\Lambda^{\mu}_{\sigma'}\Lambda^{\mu'}_{\mu} = \frac{\partial x^{\mu}}{\partial x^{\sigma'}}\frac{\partial x^{\mu'}}{\partial x^{\mu}} = \frac{\partial x^{\mu'}}{\partial x^{\sigma'}} = \delta^{\mu'}_{\sigma'}$$

implies $NM = I$. Hence, $N = M^{-1}$.

Note that equivalent conclusions can be drawn if we choose to represent a $\begin{pmatrix}1\\1\end{pmatrix}$ tensor *A* by a matrix **B** with components $b_{ij} = A^{j}_{i} = A_{i}{}^{j}$.

2.2.2.3. One-Forms

Rank $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ tensors are called **one-forms** in differential geometry and **covariant vectors** in tensor analysis. More precisely, a one-form ω is a linear mapping

$$\omega : M \to R \qquad \text{with} \qquad \mathbf{V} \mapsto \omega(\mathbf{V})$$

and $$\omega(a\mathbf{V} + b\mathbf{U}) = a\omega(\mathbf{V}) + b\omega(\mathbf{U})$$

where a and b are any 2 constants (scalar fields if ω is a 1-form field). Given a coordinate system, we can write

$$\omega(\mathbf{V}) = \omega_\mu V^\mu \tag{2.15}$$

where ω_μ is the components of ω. Note that (2.15) is the generalization of the dot product in Euclidean space in which each vector **V** can be assigned an 1-form with components $V_\mu = V^\mu$.

The prototype of an 1-form is the gradient of a scalar field f. We denote this 1-form as ω_f with components $\omega_{f\mu} = \partial_\mu f$. If $\mathbf{V} = \frac{d}{d\lambda}$ is the tangent vector to a curve $x^\mu(\lambda)$, the contraction $\omega_f(\mathbf{V})$ is a scalar field given by

$$\omega_f(\mathbf{V}) = \frac{\partial f}{\partial x^\mu} V^\mu \quad = \frac{\partial f}{\partial x^\mu} \frac{dx^\mu}{d\lambda} \quad = \frac{df}{d\lambda} \tag{2.16}$$

where is just the rate of change of f along the curve.

By definition, $\omega(\mathbf{V})$ is a scalar, i.e.,

$$\omega(\mathbf{V}) = \omega_\mu V^\mu = \omega_{\mu'} V^{\mu'} \tag{2.18}$$

Since $V^{\mu'} = \Lambda^{\mu'}_{\mu} V^\mu$, we have

$$\omega_\mu = \omega_{\mu'} \Lambda^{\mu'}_{\mu} \qquad \Rightarrow \qquad \omega_{\mu'} = \omega_\mu \Lambda^{\mu}_{\mu'} \tag{2.17}$$

In tensor analysis, (2.17) is the definition of a covariant vector.

2.2.2.4. Tensors

In tensor analysis, a tensor of contravariant rank *a* and covariant rank *b*, or simply of rank $\begin{pmatrix} a \\ b \end{pmatrix}$, has d^{a+b} components in a *d*-dimensional manifold. Given the specification of its components $T^{\alpha\beta\cdots}_{\mu\nu\cdots}$ in some coordinate system, its components in another coordinate system are given by

$$T^{\alpha'\beta'\cdots}_{\mu'\nu'\cdots} = \Lambda^{\alpha'}_{\alpha}\Lambda^{\beta'}_{\beta}\cdots\Lambda^{\mu}_{\mu'}\Lambda^{\nu}_{\nu'}\cdots T^{\alpha\beta\cdots}_{\mu\nu\cdots} \qquad (2.19)$$

In fact, (2.19) is the definition of a rank $\begin{pmatrix} a \\ b \end{pmatrix}$ tensor in tensor analysis. If a physical law can be written in tensorial form in a particular coordinate systems, it is guaranteed to be valid for all coordinate systems.

A **contraction** is the summation over a pair of contravariant – covariant indices. After the contraction, a rank $\begin{pmatrix} a \\ b \end{pmatrix}$ tensor is reduced to rank $\begin{pmatrix} a-1 \\ b-1 \end{pmatrix}$. Obviously, the contraction can be continued until one of the ranks is reduced to 0.

2.3. Extra Geometrical Structures

2.3.0. Introduction

There are 2 basic geometrical structures of our spacetime continuum that are not yet defined in a manifold, i.e.,

1. **Parallelism** (to be represented by an **affine connection**).
2. **Lengths and angles** (to be represented by a **metric**).

In Euclid's geometry, parallelism is defined through the famous 5th axiom in terms of lengths of angles. However, if our manifold is, say, a function space, the concept of parallelism may be valid but "distances" are often meaningless. Thus, in general, we must treat parallelism and lengths-angles as independent geometrical structures. Indeed, many physical theories (e.g., gauge theories) made use of manifolds endowed with an affine connection but no metric.

Obviously, the Euclidean space is described by a special kind of affine connection that can be deduced from a metric. Such a connection is called the **metric**, or **Levi-Civita**, **connection**. In relativity theories, the spacetime continuum is described by a metric connection.

2.3.1. The Affine Connection

Tools provided by the affine connection:

1. Parallelism.
2. Curvature.
3. Covariant derivative.
4. Geodesic.

Applications:

(a) General relativity: path of "free" particle is a geodesic (path with only parallel tangents).

(b) Covariant Derivatives

   i) Transformation of $\partial_\mu V^\nu$

   ii) Geometric reason why $\partial_\mu V^\nu$ is not a tensor

   iii) Parallel transport

   iv) Covariant derivatives on vectors

   v) Covariant derivatives on tensors

2.3.1.b.i. Transformation of $\partial_\mu V^\nu$

From

$$\partial_{\mu'} V^{\nu'} = \Lambda^{\mu}_{\mu'} \partial_\mu \left(\Lambda^{\nu'}_{\nu} V^\nu \right)$$

$$= \Lambda^{\mu}_{\mu'} \Lambda^{\nu'}_{\nu} \partial_\mu V^\nu + \Lambda^{\mu}_{\mu'} \left(\partial_\mu \Lambda^{\nu'}_{\nu} \right) V^\nu \qquad (2.20)$$

we see that $\partial_\mu V^\nu$ is not a tensor. Later on, we shall define a **covariant derivative** ∇_μ so that for a $\begin{pmatrix} p \\ q \end{pmatrix}$ tensor T, ∇T is a $\begin{pmatrix} p \\ q+1 \end{pmatrix}$ tensor. In particular,

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + connection\ terms$$

is a $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ tensor tensor.

2.3.1.b.ii. Geometric Reason Why $\partial_\mu V^\nu$ Is Not a Tensor

Consider the derivatives of V^μ along a curve $C(\lambda)$.

Let P and Q be 2 points on the curve with parameters λ and $\lambda+\delta\lambda$, respectively. Treating V^μ as a scalar function, its derivative at P is

$$\begin{aligned}\frac{dV^\mu}{d\lambda} &= \lim_{\delta\lambda\to 0}\frac{V^\mu(Q)-V^\mu(P)}{\delta\lambda}\\ &= \frac{dx^\nu}{d\lambda}\partial_\nu V^\mu \qquad (2.21)\end{aligned}$$

where the last equality is just the chain rule. In vector form, we have

$$\frac{d\mathbf{V}}{d\lambda} = \lim_{\delta\lambda\to 0}\frac{\mathbf{V}(Q)-\mathbf{V}(P)}{\delta\lambda}$$

Now $\mathbf{V}(Q) - \mathbf{V}(P)$ is the difference of 2 vectors defined at different points in space. Since only vectors defined at the same space point form a vector space, $\mathbf{V}(Q) - \mathbf{V}(P)$ is not a member of any vector space, i.e., it is not a vector. Hence, $d\mathbf{V}/d\lambda$ is not a vector and $\partial_\nu V^\mu$ is in general not a tensor.

2.3.1.b.iii. Parallel transport

One way to define a derivative of **V** that remains a vector is

$$\left.\frac{D\mathbf{V}}{d\lambda}\right|_P = \lim_{\delta\lambda\to 0}\frac{\mathbf{V}(Q)-\mathbf{V}(P\to Q)}{\delta\lambda} \qquad (2.22)$$

where $\mathbf{V}(P\to Q)$ is a vector at Q that represents **V**(P). Given an affine connection, we can use the accompanying parallelism to define $\mathbf{V}(P\to Q)$ as the **parallel transport** of **V**(P) along the curve C parametrized by λ. If the connection is metric, the parallel transport can also be chosen to preserve lengths. Given a coordinate system so that $P(\lambda)=x^\mu(\lambda)$, we define an infinitesimal parallel transport along C as

$$V^\mu(P\to Q)=V^\mu(P)-\delta\lambda\,\Gamma^\mu_{\nu\sigma}(P)V^\nu(P)\frac{dx^\sigma}{d\lambda} \qquad (2.23)$$

where the **affine connection coefficients** $\Gamma^\mu_{\nu\sigma}(P)$ are characteristics of the affine connection. Obviously, transport over a finite distance can be obtained by integration. Note that $\mathbf{V}(P\to Q)$ depends, through $\frac{dx^\mu}{d\lambda}$, on the curve C of transport. The effect is most striking when the transport is over finite distances in a curved manifold (see Fig.2.11).

2.3.1.b.iv. Covariant Derivatives On Vectors

The covariant derivative $\nabla_\sigma V^\mu$ is defined by

$$\frac{DV^\mu}{d\lambda} = \frac{dx^\sigma}{d\lambda}\nabla_\sigma V^\mu \qquad (2.24a)$$

Putting (2.23) into (2.22), we have

$$\frac{DV^\mu}{d\lambda} = \lim_{\delta\lambda\to 0}\left[\frac{V^\mu(Q)-V^\mu(P)}{\delta\lambda} + \Gamma^\mu_{\nu\sigma}(P)V^\nu(P)\frac{dx^\sigma}{d\lambda}\right]$$

$$= \frac{dV^\mu}{d\lambda} + \Gamma^\mu_{\nu\sigma}V^\nu\frac{dx^\sigma}{d\lambda}$$

$$= \frac{dx^\sigma}{d\lambda}\partial_\sigma V^\mu + \Gamma^\mu_{\nu\sigma}V^\nu\frac{dx^\sigma}{d\lambda}$$

so that by comparison with (2.24a), we have

$$\nabla_\sigma V^\mu = \partial_\sigma V^\mu + \Gamma^\mu_{\nu\sigma}V^\nu \qquad (2.24)$$

Note that most relativistists prefer to write (2.24) as

$$V^\mu{}_{;\sigma} = V^\mu{}_{,\sigma} + \Gamma^\mu_{\nu\sigma}V^\nu$$

Since $\nabla_\sigma V^\mu$ is a $\begin{pmatrix}1\\1\end{pmatrix}$ tensor, a coordinate transformation will turn (2.24) into

$$\nabla_{\sigma'}V^{\mu'} = \Lambda^\sigma_{\sigma'}\Lambda^{\mu'}_\mu\nabla_\sigma V^\mu \quad = \Lambda^\sigma_{\sigma'}\Lambda^{\mu'}_\mu\left(\partial_\sigma V^\mu + \Gamma^\mu_{\nu\sigma}V^\nu\right)$$

$$= \partial_{\sigma'}V^{\mu'} + \Gamma^{\mu'}_{\nu'\sigma'}V^{\nu'} \quad = \Lambda^\sigma_{\sigma'}\partial_\sigma\left(\Lambda^{\mu'}_\mu V^\mu\right) + \Gamma^{\mu'}_{\nu'\sigma'}\Lambda^{\nu'}_\nu V^\nu$$

$$= \Lambda^\sigma_{\sigma'}\Lambda^{\mu'}_\mu\partial_\sigma V^\mu + \Lambda^\sigma_{\sigma'}\left(\partial_\sigma\Lambda^{\mu'}_\mu\right)V^\mu + \Gamma^{\mu'}_{\nu'\sigma'}\Lambda^{\nu'}_\nu V^\nu$$

Thus, we must have

$$\Lambda^\sigma_{\sigma'}\Lambda^{\mu'}_\mu\Gamma^\mu_{\nu\sigma}V^\nu = \Lambda^\sigma_{\sigma'}\left(\partial_\sigma\Lambda^{\mu'}_\mu\right)V^\mu + \Gamma^{\mu'}_{\nu'\sigma'}\Lambda^{\nu'}_\nu V^\nu$$

$$\Rightarrow \qquad \Lambda^\sigma_{\sigma'}\Lambda^{\mu'}_\mu\Gamma^\mu_{\nu\sigma} = \Lambda^\sigma_{\sigma'}\left(\partial_\sigma\Lambda^{\mu'}_\nu\right) + \Gamma^{\mu'}_{\nu'\sigma'}\Lambda^{\nu'}_\nu$$

$$\therefore \qquad \Gamma^{\mu'}_{\nu'\sigma'} = \Lambda^\nu_{\nu'}\left[\Lambda^\sigma_{\sigma'}\Lambda^{\mu'}_\mu\Gamma^\mu_{\nu\sigma} - \Lambda^\sigma_{\sigma'}\left(\partial_\sigma\Lambda^{\mu'}_\nu\right)\right]$$

$$= \Lambda^{\mu'}_\mu\Lambda^\nu_{\nu'}\Lambda^\sigma_{\sigma'}\Gamma^\mu_{\nu\sigma} - \Lambda^\nu_{\nu'}\left(\partial_{\sigma'}\Lambda^{\mu'}_\nu\right) \qquad (2.26a)$$

Now, from $\Lambda^\nu_{\nu'}\Lambda^{\mu'}_\nu = \delta^{\mu'}_{\nu'}$, we have

$$\left(\partial_{\sigma'}\Lambda^{\nu}_{\nu'}\right)\Lambda^{\mu'}_{\nu} + \Lambda^{\nu}_{\nu'}\left(\partial_{\sigma'}\Lambda^{\mu'}_{\nu}\right) = 0$$

Hence (2.26a) becomes

$$\Gamma^{\mu'}_{\nu'\sigma'} = \Lambda^{\mu'}_{\mu}\,\Lambda^{\nu}_{\nu'}\,\Lambda^{\sigma}_{\sigma'}\,\Gamma^{\mu}_{\nu\sigma} + \Lambda^{\mu'}_{\nu}\left(\partial_{\sigma'}\Lambda^{\nu}_{\nu'}\right) \qquad (2.26)$$

Therefore, the connection coefficients $\Gamma^{\mu}_{\nu\sigma}$ do not in general represent a tensor.

2.3.1.b.v. Covariant Derivatives on Tensors

Expressions for covariant derivatives of tensors of other ranks can be derived using the criteria that

1. ∇T is a $\begin{pmatrix} p \\ q+1 \end{pmatrix}$ tensor if T is a $\begin{pmatrix} p \\ q \end{pmatrix}$ tensor.
2. The Leibnitz rule for derivation:

$$\nabla(abc\cdots) = (\nabla a)bc\cdots + a(\nabla b)c\cdots + ab(\nabla c)\cdots + \cdots$$

3. The contraction of a pair of indices between a $\begin{pmatrix} p \\ q \end{pmatrix}$ and a $\begin{pmatrix} r \\ s \end{pmatrix}$ tensors is a $\begin{pmatrix} p+r-1 \\ q+s-1 \end{pmatrix}$ tensor.

Scalar field *f*

By definition,

$$\nabla_\mu f = \partial_\mu f \qquad (2.27a)$$

1-form ω

Since $\omega_\mu V^\mu$ is a scalar, we have

$$\nabla_\sigma\left(\omega_\mu V^\mu\right) = \left(\nabla_\sigma \omega_\mu\right) V^\mu + \omega_\mu\left(\nabla_\sigma V^\mu\right) \qquad \text{[Leibnitz rule]}$$

$$= \left(\nabla_\sigma \omega_\mu\right) V^\mu + \omega_\mu\left(\partial_\sigma V^\mu + \Gamma^\mu_{\nu\sigma} V^\nu\right) \qquad \text{[(2.24) used]} \qquad \text{(a)}$$

$$= \partial_\sigma\left(\omega_\mu V^\mu\right) = \left(\partial_\sigma \omega_\mu\right) V^\mu + \omega_\mu\left(\partial_\sigma V^\mu\right) \qquad \text{[(2.27a) used]} \qquad \text{(b)}$$

Comparing (a) and (b) gives

$$\left(\nabla_\sigma \omega_\mu\right) V^\mu + \omega_\mu \Gamma^\mu_{\nu\sigma} V^\nu = \left(\partial_\sigma \omega_\mu\right) V^\mu$$

$$\Rightarrow \qquad \nabla_\sigma \omega_\mu = \partial_\sigma \omega_\mu - \Gamma^\nu_{\mu\sigma}\, \omega_\nu \qquad (2.27)$$

General

By the Leibnitz rule, $\nabla_\sigma T^{\alpha\beta\cdots}_{\mu\nu\cdots}$ is a sum of terms each of which has ∇ operating on a single index of *T*. Using (2.24) for an upper index and (2.27) for an lower index, we have

$$\nabla_\sigma T^{\alpha\beta\cdots}_{\mu\nu\cdots} = \partial_\sigma T^{\alpha\beta\cdots}_{\mu\nu\cdots} + \Gamma^\alpha_{a\sigma} T^{a\beta\cdots}_{\mu\nu\cdots} + \Gamma^\beta_{b\sigma} T^{\alpha b\cdots}_{\mu\nu\cdots} + \cdots - \Gamma^m_{\mu\sigma} T^{\alpha\beta\cdots}_{m\nu\cdots} - \Gamma^n_{\nu\sigma} T^{\alpha\beta\cdots}_{\mu n\cdots} - \cdots$$

2.3.2. Geodesics

A geodesic is a path on which the tangents at every pair of points are parallel to each other. Using (2.23), we see that the parallel transport of the tangent vector $\frac{dx^\mu}{d\lambda}$ at P to Q along a curve $C(\lambda)$ is

$$\frac{dx^\mu(P\to Q)}{d\lambda}=\frac{dx^\mu(P)}{d\lambda}-\delta\lambda\,\Gamma^\mu_{\nu\sigma}(P)\frac{dx^\nu(P)}{d\lambda}\frac{dx^\sigma(P)}{d\lambda} \qquad \text{(a)}$$

If $C(\lambda)$ is a geodesic, we have

$$\frac{dx^\mu(P\to Q)}{d\lambda}=c(\delta\lambda)\frac{dx^\mu(Q)}{d\lambda} \qquad \text{(b)}$$

where $c(\delta\lambda)$ is some arbitrary function. Obviously, $c(0)=1$ so that for infinitesimal $\delta\lambda$, we can write $c(\delta\lambda)\simeq 1-f(\lambda)\,\delta\lambda$, where $f(\lambda)$ is some arbitrary function. Thus, for small $\delta\lambda$, eq(b) becomes

$$\begin{aligned}\frac{dx^\mu(P\to Q)}{d\lambda}&\simeq\left[1-f(\lambda)\delta\lambda\right]\frac{dx^\mu(Q)}{d\lambda} \qquad (2.30)\\ &\simeq\left[1-f(\lambda)\delta\lambda\right]\left[\frac{dx^\mu(P)}{d\lambda}+\frac{d^2x^\mu(P)}{d\lambda^2}\delta\lambda+\cdots\right]\\ &\simeq\frac{dx^\mu(P)}{d\lambda}+\left[\frac{d^2x^\mu(P)}{d\lambda^2}-f(\lambda)\frac{dx^\mu(P)}{d\lambda}\right]\delta\lambda+O(\delta\lambda)^2\end{aligned}$$

Putting this into (a) gives

$$\left[\frac{d^2x^\mu(P)}{d\lambda^2}-f(\lambda)\frac{dx^\mu(P)}{d\lambda}+\Gamma^\mu_{\nu\sigma}(P)\frac{dx^\nu(P)}{d\lambda}\frac{dx^\sigma(P)}{d\lambda}\right]\delta\lambda+O(\delta\lambda)^2=0$$

$\Rightarrow$

$$\frac{d^2x^\mu}{d\lambda^2}+\Gamma^\mu_{\nu\sigma}\frac{dx^\nu}{d\lambda}\frac{dx^\sigma}{d\lambda}=f(\lambda)\frac{dx^\mu}{d\lambda} \qquad (2.31)$$

which is called the **geodesic equation**. If we switch to a new parametrization τ, we have

$$\frac{d}{d\lambda}=\frac{d\tau}{d\lambda}\frac{d}{d\tau}$$

$$\frac{d^2}{d\lambda^2} = \frac{d^2\tau}{d\lambda^2}\frac{d}{d\tau} + \frac{d\tau}{d\lambda}\frac{d^2}{d\lambda d\tau} = \frac{d^2\tau}{d\lambda^2}\frac{d}{d\tau} + \left(\frac{d\tau}{d\lambda}\right)^2 \frac{d^2}{d\tau^2}$$

so that (2.31) becomes

$$\frac{d^2\tau}{d\lambda^2}\frac{dx^\mu}{d\tau} + \left(\frac{d\tau}{d\lambda}\right)^2 \frac{d^2x^\mu}{d\tau^2} + \left(\frac{d\tau}{d\lambda}\right)^2 \Gamma^\mu_{\nu\sigma}\frac{dx^\nu}{d\tau}\frac{dx^\sigma}{d\tau} = f(\lambda)\left(\frac{d\tau}{d\lambda}\right)\frac{dx^\mu}{d\tau}$$

$$\Rightarrow \qquad \frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\nu\sigma}\frac{dx^\nu}{d\tau}\frac{dx^\sigma}{d\tau} = \left(\frac{d\tau}{d\lambda}\right)^{-2}\left[f(\lambda)\left(\frac{d\tau}{d\lambda}\right) - \frac{d^2\tau}{d\lambda^2}\right]\frac{dx^\mu}{d\tau} \qquad (2.32)$$

$$\equiv f(\tau)\frac{dx^\mu}{d\tau}$$

Thus, under a change of parametrization, the geodesic equation retains the same form except for a change of the function *f*. In particular, the parameter that has $f = 0$ is called the **affine parameter**. Note that if λ is an affine parameter, then all of its affinely related parameters $\tau = a\lambda + b$ are also affine parameters.

2.3.3. The Riemann Curvature Tensor

2.3.3.a. Parallel Transport Over Different Paths

Consider 2 points P and Q with coordinates x^μ and $x^\mu + \delta x^\mu$, respectively. Without loss of generality, we can assume δx^μ to be nonzero only for $\mu = a$ or b, i.e., $\delta x^\mu = \delta^{\mu a}\delta x^a + \delta^{\mu b}\delta x^b$. Let R and S be two "mid-points" with coordinates $x^\mu + \delta^{\mu a}\delta x^a$ and $x^\mu + \delta^{\mu b}\delta x^b$, respectively. [see Fig.2.12] Thus,

$$V^\mu(P\to R) = V^\mu(P) - \Gamma^\mu_{\nu a}(P)V^\nu(P)\delta x^a$$

$$V^\mu(P\to R\to Q) = V^\mu(P\to R) - \Gamma^\mu_{\nu a}(R)V^\nu(P\to R)\delta x^b$$

$$= V^\mu(P) - \Gamma^\mu_{\nu a}(P)V^\nu(P)\delta x^a$$

$$- \Gamma^\mu_{\nu b}(R)\left[V^\nu(P) - \Gamma^\nu_{\sigma a}(P)V^\sigma(P)\delta x^a\right]\delta x^b$$

and similarly,

$$V^\mu(P\to S) = V^\mu(P) - \Gamma^\mu_{\nu b}(P)V^\nu(P)\delta x^b$$

$$V^\mu(P\to S\to Q) = V^\mu(P) - \Gamma^\mu_{\nu b}(P)V^\nu(P)\delta x^b$$

$$- \Gamma^\mu_{\nu a}(S)\left[V^\nu(P) - \Gamma^\nu_{\sigma b}(P)V^\sigma(P)\delta x^b\right]\delta x^a$$

Using

$$\Gamma^\mu_{\nu b}(R) \simeq \Gamma^\mu_{\nu b}(P) + \Gamma^\mu_{\nu b,\,a}(P)\,\delta x^a$$

$$\Gamma^\mu_{\nu a}(S) \simeq \Gamma^\mu_{\nu a}(P) + \Gamma^\mu_{\nu a,\,b}(P)\,\delta x^b$$

we have

$$V^\mu(P\to S\to Q) - V^\mu(P\to R\to Q)$$

$$= \delta x^a \delta x^b \left\{ -\Gamma^\mu_{\nu a,b}(P)V^\nu(P) + \Gamma^\mu_{\nu a}(P)\Gamma^\nu_{\sigma b}(P)V^\sigma(P)\right.$$

$$\left. + \Gamma^\mu_{\nu b,a}(P)V^\nu(P) - \Gamma^\mu_{\nu b}(P)\Gamma^\nu_{\sigma a}(P)V^\sigma(P) \right\}$$

$$= \delta x^a \delta x^b \left(\Gamma^\mu_{\sigma b,a} - \Gamma^\mu_{\sigma a,b} + \Gamma^\mu_{\nu a}\Gamma^\nu_{\sigma b} - \Gamma^\mu_{\nu b}\Gamma^\nu_{\sigma a}\right)V^\sigma$$

$$= \delta x^a \delta x^b \; R^\mu_{\sigma ab}\, V^\sigma$$

where

$$R^{\mu}_{\sigma ab} = \Gamma^{\mu}_{\sigma b,a} - \Gamma^{\mu}_{\sigma a,b} + \Gamma^{\mu}_{\nu a}\Gamma^{\nu}_{\sigma b} - \Gamma^{\mu}_{\nu b}\Gamma^{\nu}_{\sigma a} \qquad (2.35)$$

Allowing a, b to range over the dimension of the manifold, eq(2.35) defines a rank 4 tensor called the **Riemann tensor**.

2.3.3.b. Commutator of the Covariant Derivatives

Since the parallel transport is closely related to the covariant derivative, (2.33) can also be expressed in terms of the commutator of the latter. Consider the commutator

$$\left[\nabla_\sigma , \nabla_\tau\right] \equiv \nabla_\sigma \nabla_\tau - \nabla_\tau \nabla_\sigma$$

Its effect on a vector is

$$\left[\nabla_\sigma , \nabla_\tau\right] V^\mu = \nabla_\sigma \left(\nabla_\tau V^\mu\right) - \nabla_\tau \left(\nabla_\sigma V^\mu\right)$$

which is a rank 3 tensors. Since $\nabla_\tau V^\mu$ is a $\begin{pmatrix}1\\1\end{pmatrix}$ tensor, we have, by (2.28),

$$\nabla_\sigma \left(\nabla_\tau V^\mu\right) = \partial_\sigma \left(\nabla_\tau V^\mu\right) + \Gamma^\mu_{\nu\sigma} \nabla_\tau V^\nu - \Gamma^\nu_{\tau\sigma} \nabla_\nu V^\mu$$

$$= \partial_\sigma \left(\partial_\tau V^\mu + \Gamma^\mu_{\nu\tau} V^\nu\right) + \Gamma^\mu_{\nu\sigma} \left(\partial_\tau V^\nu + \Gamma^\nu_{\lambda\tau} V^\lambda\right) - \Gamma^\nu_{\tau\sigma} \nabla_\nu V^\mu$$

$$= \partial_\sigma \partial_\tau V^\mu + \Gamma^\mu_{\nu\tau,\sigma} V^\nu + \Gamma^\mu_{\nu\tau} \partial_\sigma V^\nu + \Gamma^\mu_{\nu\sigma} \partial_\tau V^\nu + \Gamma^\mu_{\nu\sigma} \Gamma^\nu_{\lambda\tau} V^\lambda - \Gamma^\nu_{\tau\sigma} \nabla_\nu V^\mu$$

Taking $\sigma \leftrightarrow \tau$, we have

$$\nabla_\tau \left(\nabla_\sigma V^\mu\right) = \partial_\tau \left(\nabla_\sigma V^\mu\right) + \Gamma^\mu_{\nu\tau} \nabla_\sigma V^\nu - \Gamma^\nu_{\tau\sigma} \nabla_\nu V^\mu$$

$$= \partial_\tau \partial_\sigma V^\mu + \Gamma^\mu_{\nu\sigma,\tau} V^\nu + \Gamma^\mu_{\nu\sigma} \partial_\tau V^\nu + \Gamma^\mu_{\nu\tau} \partial_\sigma V^\nu + \Gamma^\mu_{\nu\tau} \Gamma^\nu_{\lambda\sigma} V^\lambda - \Gamma^\nu_{\sigma\tau} \nabla_\nu V^\mu$$

Hence,

$$\left[\nabla_\sigma , \nabla_\tau\right] V^\mu = \Gamma^\mu_{\nu\tau,\sigma} V^\nu + \Gamma^\mu_{\nu\sigma} \Gamma^\nu_{\lambda\tau} V^\lambda - \Gamma^\nu_{\tau\sigma} \nabla_\nu V^\mu$$

$$-\Gamma^\mu_{\nu\sigma,\tau} V^\nu - \Gamma^\mu_{\nu\tau} \Gamma^\nu_{\lambda\sigma} V^\lambda + \Gamma^\nu_{\sigma\tau} \nabla_\nu V^\mu$$

$$= \left(\Gamma^\mu_{\nu\tau,\sigma} - \Gamma^\mu_{\nu\sigma,\tau} + \Gamma^\mu_{\lambda\sigma} \Gamma^\lambda_{\nu\tau} - \Gamma^\mu_{\lambda\tau} \Gamma^\lambda_{\nu\sigma}\right) V^\nu + \left(\Gamma^\nu_{\sigma\tau} - \Gamma^\nu_{\tau\sigma}\right) \nabla_\nu V^\mu$$

$$= R^\mu_{\nu\sigma\tau} V^\nu + \left(\Gamma^\nu_{\sigma\tau} - \Gamma^\nu_{\tau\sigma}\right) \nabla_\nu V^\mu \qquad (2.34)$$

2.3.3.c. Properties of the Riemann Tensor

As a rank 4 tensor, $R^{\mu}_{\nu\sigma\tau}$ has $4^4 = 256$ components in a 4-D space. However, owing to various symmetries, e.g., $R^{\mu}_{\nu\sigma\tau} = -R^{\mu}_{\nu\tau\sigma}$ since $[\nabla_\sigma, \nabla_\tau] = -[\nabla_\tau, \nabla_\sigma]$, it can shown that the number of independent components is only 80. If Γ is a metric connection, the additional symmetries further reduces this number to 20.

Obviously, $R^{\mu}_{\nu\sigma\tau}$ is a complicated object the manipulation of which is always tedious.

Fortunately, in most applications in physics, e.g., general relativity, one needs only deal with its contractions, e.g., the rank 2 **Ricci tensor** defined by

$$R_{\mu\nu} \equiv R^{\lambda}_{\mu\lambda\nu} = \Gamma^{\lambda}_{\mu\nu,\lambda} - \Gamma^{\lambda}_{\mu\lambda,\nu} + \Gamma^{\lambda}_{\sigma\lambda}\,\Gamma^{\sigma}_{\mu\nu} - \Gamma^{\lambda}_{\sigma\nu}\,\Gamma^{\sigma}_{\mu\lambda} \qquad (2.36)$$

or the **Ricci scalar** defined by

$$R = R^{\mu}_{\mu}$$

Although Γ is not a tensor, the difference $\Gamma^{\nu}_{\sigma\tau} - \Gamma^{\nu}_{\tau\sigma}$ is a tensor called the **torsion tensor**. The torsion tensor of the spacetime continuum is often set to zero so that (2.34) becomes much simpler. However, there is not yet any means to test this assumption experimentally.

2.3.4. The Metric

The infinitesimal **distance** *ds* between 2 points at x^μ and $x^\mu + dx^\mu$ is defined as

$$(ds)^2 = ds^2 = g_{\mu\nu}(x)\,dx^\mu dx^\nu \qquad (2.37)$$

where $g_{\mu\nu}(x)$ is the **metric** tensor field. Since the antisymmetric part of *g* cancels out in the summations in (2.37), we can assume *g* to be symmetric without loss of generality. Finite distances between 2 points *P* and *Q* are obtained by integration along a specific path $C(\lambda)$ joining them:

$$s_{PQ} = \int_P^Q d\lambda\, \frac{ds}{d\lambda} = \int_P^Q d\lambda\, \sqrt{g_{\mu\nu}\left[x(\lambda)\right]\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}} \qquad (2.38)$$

In a 3-D Euclidean space described by Cartesian coordinates,

$$ds^2 = \sum_{\mu=1}^{3}\left(dx^\mu\right)^2$$

so that

$$g_{\mu\nu} = \delta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (2.39)$$

which is independent of *x*. This is the prototype of a "flat" space.

One important use of the metric tensor is to lower or raise tensor indices. Thus, for a (contravariant) vector V^μ, we set

$$V_\mu = g_{\mu\nu} V^\nu \qquad \text{and} \qquad V^\mu = g^{\mu\nu} V_\nu \qquad (2.43,5)$$

where V_μ is a 1-form (covariant vector) and $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$, i.e.,

$$g_{\mu\nu} g^{\nu\sigma} = \delta_\mu^\sigma \qquad (2.44)$$

Thus, one can define a scalar product between 2 vectors by

$$\mathbf{U}\cdot\mathbf{V} \equiv g_{\mu\nu} U^\mu V^\nu = U^\mu V_\mu = U_\mu V^\mu \qquad (2.40)$$

Hence, in a metric space, one can regard vectors and 1-forms as different representations of the same object, hence the terms **contra**- and **co- variant vectors**.

In fact, they can be treated as identical in a Euclidean space with cartesian coordinates.

2.3.5. The Metric Connection

For a manifold endowed with both affine connection (parallelism) and metric (lengths and angles), one must make sure they are compatible. Thus, if 2 vectors are parallel transported together along a curve, the angle between them should remain unchanged. As mentioned before, an affine connection that is compatible with a metric is called a **metric connection**.

Consider the parallel transport of 2 vectors **V** and **W** at point $P = x^{\mu}(\lambda)$ along curve $C(\lambda)$ to point $Q = x^{\mu}(\lambda + \delta\lambda)$. We can define a vector field **V**(*x*) such that $\mathbf{V}(Q) = \mathbf{V}(P \to Q)$ for any point *Q* on *C*; and analogously a field **W**(*x*). The derivative of **V** along *C* is therefore

$$\frac{D\mathbf{V}}{d\lambda} = U^{\sigma}\nabla_{\sigma}\mathbf{V} = \lim_{\delta\lambda\to 0}\frac{\mathbf{V}(Q) - \mathbf{V}(P \to Q)}{\delta\lambda} = 0 \qquad (2.46a)$$

where **U** is the tangent to *C*. Similarly, $\frac{D\mathbf{W}}{d\lambda} = 0$.

The compatibility condition for a metric connection can be stated as the conservation of the scalar product $\mathbf{V}\cdot\mathbf{W} = g_{\mu\nu}V^{\mu}W^{\nu}$ under parallel transport, i.e.,

$$\frac{D(\mathbf{V}\cdot\mathbf{W})}{d\lambda} = U^{\sigma}\nabla_{\sigma}\left(g_{\mu\nu}V^{\mu}W^{\nu}\right) = 0 \qquad (2.46)$$

Now, the Leibnitz rule gives

$$\nabla_{\sigma}\left(g_{\mu\nu}V^{\mu}W^{\nu}\right) = \left(\nabla_{\sigma}g_{\mu\nu}\right)V^{\mu}W^{\nu} + g_{\mu\nu}\left(\nabla_{\sigma}V^{\mu}\right)W^{\nu} + g_{\mu\nu}V^{\mu}\left(\nabla_{\sigma}W^{\nu}\right) \qquad (2.47)$$

so that using (2.46a), eq(2.46) becomes

$$\left(U^{\sigma}\nabla_{\sigma}g_{\mu\nu}\right)V^{\mu}W^{\nu} = 0$$

Since this must hold for arbitrary **V** and **W** on arbitrary curve or **U**, we must have

$$\nabla_{\sigma}g_{\mu\nu} = 0$$

i.e.,

$$g_{\mu\nu,\sigma} - \Gamma^{\tau}_{\mu\sigma}\, g_{\tau\nu} - \Gamma^{\tau}_{\nu\sigma}\, g_{\mu\tau} = 0 \qquad (2.48)$$

Therefore, *g* is a constant under the covariant derivative, i.e., it is **covariantly constant**. By interchanging indices in (2.48), we have

$$\mu \to \sigma \to \nu \to \mu: \qquad g_{\sigma\mu,\nu} - \Gamma^{\tau}_{\sigma\nu}\, g_{\tau\mu} - \Gamma^{\tau}_{\mu\nu}\, g_{\sigma\tau} = 0 \qquad (2.48a)$$

$\mu \to \sigma \to \mu$: $$g_{\sigma\nu,\mu} - \Gamma^{\tau}_{\sigma\mu}\, g_{\tau\nu} - \Gamma^{\tau}_{\nu\mu}\, g_{\sigma\tau} = 0 \qquad (2.48b)$$

Hence, (2.48a) + (2.48b) – (2.48) gives

$$g_{\sigma\mu,\nu} + g_{\sigma\nu,\mu} - g_{\mu\nu,\sigma} = \left(\Gamma^{\tau}_{\sigma\nu} - \Gamma^{\tau}_{\nu\sigma}\right) g_{\tau\mu} + \left(\Gamma^{\tau}_{\sigma\mu} - \Gamma^{\tau}_{\mu\sigma}\right) g_{\tau\nu} + \left(\Gamma^{\tau}_{\nu\mu} + \Gamma^{\tau}_{\mu\nu}\right) g_{\sigma\tau} \qquad (2.49)$$

where we've made use of the symmetricity of *g*. If the connection is also symmetric in the lower indices, i.e., $\Gamma^{\mu}_{\sigma\nu} = \Gamma^{\mu}_{\nu\sigma}$, (2.49) simplifies to

$$g_{\sigma\mu,\nu} + g_{\sigma\nu,\mu} - g_{\mu\nu,\sigma} = 2\Gamma^{\tau}_{\mu\nu}\, g_{\sigma\tau} \qquad (2.49a)$$

$$\Rightarrow \qquad \Gamma^{\tau}_{\mu\nu} = \frac{1}{2} g^{\tau\sigma} \left(g_{\sigma\mu,\nu} + g_{\sigma\nu,\mu} - g_{\mu\nu,\sigma}\right) \qquad (2.50)$$

The $\Gamma^{\tau}_{\mu\nu}$ in (2.50) is called the **Christoffel symbol**.

A metric connection expresses the parallelism implied by the metric. However, the manifold can be equipped with other affine connections and their associated parallelism. Indeed, it may even possess several different metrics. In which case, there will be several different kinds of "distance" and meanings of "parallel". Fortunately, our space-time continuum seems to have only one metric.

A measure of the curvature of the manifold is given by the **Ricci curvature scalar** defined by

$$R = g^{\mu\nu} R_{\mu\nu} = R^{\mu}_{\mu} \qquad (2.51)$$

which can be interpreted as a **radius of curvature**.

2.4. What is the Structure of Our Spacetime?

1. Galilean Spacetime

   Space is the 3-D Euclidean space S. Time is an 1-D metric space T. Spacetime is the direct product space $T \otimes S$ called a **fibre bundle**. [see Fig.2.13]

2. Minkowski Spacetime

   Spacetime is a 4-D metric flat space with signature -2.

3. Curved Spacetime

   Spacetime is a 4-D space with a metric tensor that depends on gravity.