Introduction to Differential Geometry & General Relativity

Third Printing January 2002

Lecture Notes
by
Stefan Waner

with a Special Guest Lecture
by Gregory C. Levine

Departments of Mathematics and Physics, Hofstra University

Introduction to Differential Geometry and General Relativity

Lecture Notes by Stefan Waner,
with a Special Guest Lecture by Gregory C. Levine

Department of Mathematics, Hofstra University

These notes are dedicated to the memory of Hanno Rund.

TABLE OF CONTENTS

1. Preliminaries

Distance and Open Sets

Here, we do just enough topology so as to be able to talk about smooth manifolds. We begin with n-dimensional Euclidean space

$$E_n = \{(y_1, y_2, \dots, y_n) \mid y_i \in \mathbb{R}\}.$$

Thus, E_1 is just the real line, E_2 is the Euclidean plane, and E_3 is 3−dimensional Euclidean space.

The **magnitude**, or **norm**, $\|\boldsymbol{y}\|$ of $\boldsymbol{y} = (y_1, y_2, \dots, y_n)$ in E_n is defined to be

$$\|\boldsymbol{y}\| = \sqrt{{y_1}^2 + {y_2}^2 + \dots + {y_n}^2},$$

which we think of as its distance from the origin. Thus, the **distance** between two points $\boldsymbol{y} = (y_1, y_2, \dots, y_n)$ and $\boldsymbol{z} = (z_1, z_2, \dots, z_n)$ in E_n is defined as the norm of $\boldsymbol{z} - \boldsymbol{y}$:

Distance Formula

Distance between $\boldsymbol{y}$ and $\boldsymbol{z} = \|\boldsymbol{z} - \boldsymbol{y}\| = \sqrt{(z_1 - y_1)^2 + (z_2 - y_2)^2 + \dots + (z_n - y_n)^2}$.

Proposition 1.1 (Properties of the norm)
The norm satisfies the following:
(a) $\|\boldsymbol{y}\| \geq 0$, and $\|\boldsymbol{y}\| = 0$ iff $\boldsymbol{y} = 0$ (positive definite)
(b) $\|\lambda\boldsymbol{y}\| = |\lambda|\|\boldsymbol{y}\|$ for every $\lambda \in \mathbb{R}$ and $\boldsymbol{y} \in E_n$.
(c) $\|\boldsymbol{y} + \boldsymbol{z}\| \leq \|\boldsymbol{y}\| + \|\boldsymbol{z}\|$ for every $\boldsymbol{y}, \boldsymbol{z} \in E_n$ (triangle inequality 1)
(d) $\|\boldsymbol{y} - \boldsymbol{z}\| \leq \|\boldsymbol{y} - \boldsymbol{w}\| + \|\boldsymbol{w} - \boldsymbol{z}\|$ for every $\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w} \in E_n$ (triangle inequality 2)

The proof of Proposition 1.1 is an exercise which may require reference to a linear algebra text (see “inner products”).

Definition 1.2 A Subset U of E_n is called **open** if, for every $\boldsymbol{y}$ in U, all points of E_n within some positive distance r of $\boldsymbol{y}$ are also in U. (The size of r may depend on the point $\boldsymbol{y}$ chosen. Illustration in class).

Intuitively, an open set is a solid region minus its boundary. If we include the boundary, we get a **closed set**, which formally is defined as the complement of an open set.

Examples 1.3
(a) If $a \in E_n$, then the **open ball with center $\boldsymbol{a}$ and radius $\boldsymbol{r}$** is the subset

$$B(\boldsymbol{a}, r) = \{x \in E_n \mid \|\boldsymbol{x} - \boldsymbol{a}\| < r\}.$$

Open balls are open sets: If $\boldsymbol{x} \in B(\boldsymbol{a}, r)$, then, with $s = r - \|\boldsymbol{x}-\boldsymbol{a}\|$, one has $B(\boldsymbol{x}, s) \subset B(\boldsymbol{a}, r)$.
(b) E_n is open.
(c) Ø is open.
(d) Unions of open sets are open.
(e) Open sets are unions of open balls. (Proof in class)

Definition 1.4 Now let $M \subset E_s$. A subset $V \subset M$ is called **open in *M*** (or **relatively open**) if, for every $\boldsymbol{y}$ in V, all points of M within some positive distance r of $\boldsymbol{y}$ are also in V.

Examples 1.5
(a) Open balls in *M*
If $M \subset E_s$, $\boldsymbol{m} \in M$, and $r > 0$, define

$$B_M(\boldsymbol{m}, r) = \{x \in M \mid \|\boldsymbol{x}-\boldsymbol{m}\| < r\}.$$

Then

$$B_M(\boldsymbol{m}, r) = B(\boldsymbol{m}, r) \cap M,$$

and so $B_M(\boldsymbol{m}, r)$ is open in M.
(b) M is open in M.
(c) Ø is open in M.
(d) Unions of open sets in M are open in M.
(e) Open sets in M are unions of open balls in M.

Parametric Paths and Surfaces in E_3

From now on, the three coordinates of 3-space will be referred to as y_1, y_2, and y_3.

Definition 1.6 A smooth **path** in E_3 is a set of three smooth (infinitely differentiable) real-valued functions of a single real variable t:

$$y_1 = y_1(t),\ y_2 = y_2(t),\ y_3 = y_3(t).$$

The variable t is called the **parameter** of the curve. The path is **non-singular** if the vector $(\frac{dy_1}{dt}, \frac{dy_2}{dt}, \frac{dy_3}{dt})$ is nowhere zero.

Notes
(a) Instead of writing $y_1 = y_1(t)$, $y_2 = y_2(t)$, $y_3 = y_3(t)$, we shall simply write $y_i = y_i(t)$.
(b) Since there is nothing special about three dimensions, we define a **smooth path in E_n** in exactly the same way: as a collection of smooth functions $y_i = y_i(t)$, where this time i goes from 1 to n.

Examples 1.7
(a) Straight lines in E_3
(b) Curves in E_3 (circles, etc.)

Definition 1.8 A **smooth surface immersed in E_3** is a collection of three smooth real-valued functions of *two* variables x^1 and x^2 (notice that x finally makes a debut).

$$y_1 = y_1(x^1, x^2)$$
$$y_2 = y_2(x^1, x^2)$$
$$y_3 = y_3(x^1, x^2),$$

or just

$$y_i = y_i(x^1, x^2) \quad (i = 1, 2, 3).$$

We also require that the 3×2 matrix whose ij entry is $\dfrac{\partial y_i}{\partial x^j}$ has rank two. We call x^1 and x^2 the **parameters** or **local coordinates**.

Examples 1.9
(a) Planes in E_3
(b) The paraboloid $y_3 = y_1^2 + y_2^2$
(c) The sphere $y_1^2 + y_2^2 + y_3^2 = 1$, using spherical polar coordinates:

$$y_1 = \sin x_1 \cos x_2$$
$$y_2 = \sin x_1 \sin x_2$$
$$y_3 = \cos x_1$$

(d) The ellipsoid $\dfrac{y_1^2}{a^2} + \dfrac{y_2^2}{b^2} + \dfrac{y_3^2}{c^2} = 1$, where a, b and c are positive constants.
(e) We calculate the rank of the Jacobean matrix for spherical polar coordinates.
(f) The torus with radii $a > b$:

$$y_1 = (a+b\cos x^2)\cos x^1$$
$$y_2 = (a+b\cos x^2)\sin x^1$$
$$y_3 = b\sin x^2$$

Question The parametric equations of a surface show us how to obtain a point on the surface once we know the two local coordinates (parameters). In other words, we have specified a function $E_2 \longrightarrow E_3$. How do we obtain the local coordinates from the Cartesian coordinates y_1, y_2, y_3?
Answer We need to solve for the local coordinates x^i as functions of y_j. This we do in one or two examples in class. For instance, in the case of a sphere, we get

$$x^1 = \cos^{-1}(y_3)$$

$$x^2 = \begin{cases} \cos^{-1}(y_1 / \sqrt{y_1{}^2 + y_2{}^2}) & \text{if } y_2 \geq 0 \\ 2\pi - \cos^{-1}(y_1 / \sqrt{y_1{}^2 + y_2{}^2}) & \text{if } y_2 < 0 \end{cases} .$$

This allows us to give each point on much of the sphere *two unique coordinates*, x^1, and x^2. There is a problem with continuity when $y_2 = 0$, since then x^2 switches from 0 to 2π. There is also a problem at the poles ($y_1 = y_2 = 0$), since then the above functions are not even defined. Thus, we restrict to the portion of the sphere given by

$$0 < x^1 < \pi$$
$$0 < x^2 < 2\pi,$$

which is an open subset U of the sphere. (Think of it as the surface of the earth with the Greenwich Meridian removed.) We call x^1 and x^2 the **coordinate functions**. They are functions

$$x^1\colon U \longrightarrow E_1$$

and

$$x^2\colon U \longrightarrow E_1.$$

We can put them together to obtain a single function $\boldsymbol{x}\colon U \longrightarrow E_2$ given by

$$\begin{aligned} \boldsymbol{x}(y_1, y_2, y_3) &= (x^1(y_1, y_2, y_3), x^2(y_1, y_2, y_3)) \\ &= \left(\cos^{-1}(y_3), \begin{cases} \cos^{-1}(y_1 / \sqrt{y_1{}^2 + y_2{}^2}) & \text{if } y_2 \geq 0 \\ 2\pi - \cos^{-1}(y_1 / \sqrt{y_1{}^2 + y_2{}^2}) & \text{if } y_2 < 0 \end{cases} \right) \end{aligned}$$

as specified by the above formulas, as a **chart.**

Definition 1.10 A **chart** of a surface S is a pair of functions $\boldsymbol{x} = (x^1(y_1, y_2, y_3), x^2(y_1, y_2, y_3))$ which specify each of the **local coordinates** (parameters) x^1 and x^2 as smooth functions of a general point (**global** or **ambient coordinates**) (y_1, y_2, y_3) on the surface.

Question Why are these functions called a chart?
Answer The chart above assigns to each point on the sphere (away from the meridian) two coordinates. So, we can think of it as giving a two-dimensional map of the surface of the sphere, just like a geographic chart.

Question Our chart for the sphere is very nice, but is only appears to chart a portion of the sphere. What about the missing meridian?
Answer We can use another chart to get those by using different paramaterization that places the poles on the equator. (Diagram in class.)

In general, we chart an entire manifold M by "covering" it with open sets U which become the domains of coordinate charts.

Exercise Set 1

1. Prove Proposition 1.1.(Consult a linear algebra text.)

2. Prove the claim in Example 1.3 (d).

3. Prove that finite intersection of open sets in E_n are open.

4. Parametrize the following curves in E_3.

(a) a circle with center (1, 2, 3) and radius 4

(b) the curve $x = y^2;\ z = 3$

(c) the intersection of the planes $3x-3y+z=0$ and $4x+y+z=1$.

5. Express the following planes parametrically:

  (a) $y_1 + y_2 - 2y_3 = 0.$

  (b) $2y_1 + y_2 - y_3 = 12.$

6. Express the following quadratic surfaces parametrically: [Hint. For the hyperboloids, refer to parameterizations of the ellipsoid, and use the identity $\cosh^2 x - \sinh^2 x = 1$. For the double cone, use $y_3 = cx^1$, and x^1 as a factor of y_1 and y_2.]

  (a) Hyperboloid of One Sheet: $\dfrac{y_1^2}{a^2} + \dfrac{y_2^2}{b^2} - \dfrac{y_3^2}{c^2} = 1.$

  (b) Hyperboloid of Two Sheets: $\dfrac{y_1^2}{a^2} - \dfrac{y_2^2}{b^2} - \dfrac{y_3^2}{c^2} = 1$

  (c) Cone: $\dfrac{y_3^2}{c^2} = \dfrac{y_1^2}{a^2} + \dfrac{y_2^2}{b^2}$.

  (d) Hyperbolic Paraboloid: $\dfrac{y_3}{c} = \dfrac{y_1^2}{a^2} - \dfrac{y_2^2}{b^2}$

7. Solve the parametric equations you obtained in 5(a) and 6(b) for x^1 and x^2 as smooth functions of a general point (y_1, y_2, y_3) on the surface in question.

---

2. Smooth Manifolds and Scalar Fields

We now formalize the above ideas.

Definition 2.1 An **open cover** of $M \subset E_s$ is a collection $\{U_\alpha\}$ of open sets in M such that $M = \cup_\alpha U_\alpha$.

Examples

(a) E_s can be covered by open balls.

(b) E_s can be covered by the single set E_s.

(c) The unit sphere in E_s can be covered by the collection $\{U_1, U_2\}$ where

$$U_1 = \{(y_1, y_2, y_3) \mid y_3 > -1/2\}$$
$$U_2 = \{(y_1, y_2, y_3) \mid y_3 < 1/2\}.$$

Definition 2.2 A subset M of E_s is called an ***n*-dimensional smooth manifold** if we are given a collection $\{U_\alpha;\ x_\alpha^{\ 1}, x_\alpha^{\ 2}, \ldots, x_\alpha^{\ n}\}$ where:
(a) The U_α form an open cover of M.
(b) Each $x_\alpha^{\ r}$ is a C^∞ real-valued function defined on U (that is, $x_\alpha^{\ r}$: $U_\alpha \longrightarrow E_1$), and extending to an open set of E_s, called the ***r*-th coordinate**, such that the map x: $U_\alpha \longrightarrow E_n$ given by $x(u) = (x_\alpha^{\ 1}(u), x_\alpha^{\ 2}(u), \ldots, x_\alpha^{\ n}(u))$ is one-to-one. (That is, to each point in U_α, we are assigned a *unique* set of n coordinates.) The tuple $(U_\alpha;\ x_\alpha^{\ 1}, x_\alpha^{\ 2}, \ldots, x_\alpha^{\ n})$ is called a **local chart of *M***. The collection of all charts is called a **smooth atlas of *M***. Further, U_α is called a **coordinate neighborhood.**
(c) If (U, x^i), and $(V, \bar{x}^j)$ are two local charts of M, and if $U \cap V \neq \emptyset$, then we can write

$$x^i = x^i(\bar{x}^j)$$

with inverse

$$\bar{x}^k = \bar{x}^k(x^l)$$

for each i and k, where all functions in sight are C^∞. These functions are called the **change-of-coordinates** transformations.
By the way, we call the "big" space E_s in which the manifold M is embedded the **ambient space**.

Notes
1. Always think of the x^i as the **local coordinates** (or parameters) of the manifold. We can paramaterize each of the open sets U by using the inverse function x^{-1} of x, which assigns to each point in some neighborhood of E_n a corresponding point in the manifold.
2. Condition **(c)** implies that

$$\det\left(\frac{\partial \bar{x}^i}{\partial x^j}\right) \neq 0,$$

and

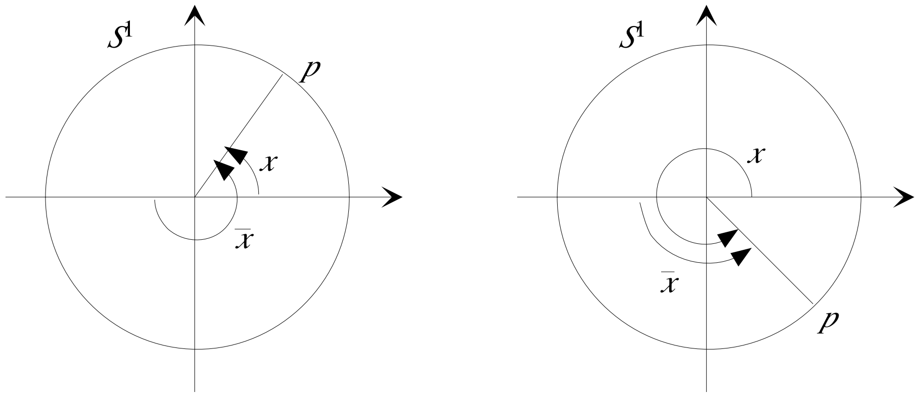
$$\det\left(\frac{\partial x^i}{\partial \bar{x}^j}\right) \neq 0,$$

since the associated matrices must be invertible.
3. The ambient space need not be present in the general theory of manifolds; that is, it is possible to define a smooth manifold M without any reference to an ambient space at all—see any text on differential topology or differential geometry (or look at Rund's appendix).
4. More terminology: We shall sometimes refer to the x^i as the **local coordinates**, and to the y^j as the **ambient coordinates**. Thus, a point in an n-dimensional manifold M in E_s has n local coordinates, but s ambient coordinates.

Examples 2.3
(a) E_n is an n-dimensional manifold, with the single identity chart defined by

$$x^i(y_1, \ldots, y_n) = y_i.$$

(b) S^1, the unit circle, with the exponential map, is a 1-dimensional manifold. Here is a possible structure:with two charts as show in in the following figure.



One has

$$x: S^1 - \{(1, 0)\} \longrightarrow E_1$$
$$\bar{x}: S^1 - \{(-1, 0)\} \longrightarrow E_1,$$

with $0 < x, \bar{x} < 2\pi$, and the change-of-coordinate maps are given by

$$\bar{x} = \begin{cases} x+\pi & \text{if } x < \pi \\ x-\pi & \text{if } x > \pi \end{cases}$$ (See the figure for the two cases.)

and

$$x = \begin{cases} \bar{x}+\pi & \text{if } \bar{x} < \pi \\ \bar{x}-\pi & \text{if } \bar{x} > \pi \end{cases} .$$

Notice the symmetry between x and $\bar{x}$. Also notice that these change-of-coordinate functions are only defined when $\theta \neq 0, \pi$. Further,

$$\partial\bar{x}/\partial x = \partial x/\partial\bar{x} = 1.$$

Note also that, in terms of complex numbers, we can write, for a point $p = e^{iz} \in S^1$,

$$x = \arg(z), \quad \bar{x} = \arg(-z).$$

(c) Generalized Polar Coordinates

Let us take $M = S^n$, the unit $n-$sphere,

$$S^n = \{(y_1, y_2, \dots , y_n, y_{n+1}) \in E_{n+1} \mid \Sigma_i y_i^2 = 1\},$$

with coordinates $(x^1, x^2, \dots , x^n)$ with

$$0 < x^1, x^2, \dots , x^{n-1} < \pi$$

and

$$0 < x^n < 2\pi,$$

given by

$$\begin{aligned}
y_1 &= \cos x^1 \\
y_2 &= \sin x^1 \cos x^2 \\
y_3 &= \sin x^1 \sin x^2 \cos x^3 \\
&\dots \\
y_{n-1} &= \sin x^1 \sin x^2 \sin x^3 \sin x^4 \dots \cos x^{n-1} \\
y_n &= \sin x^1 \sin x^2 \sin x^3 \sin x^4 \dots \sin x^{n-1} \cos x^n \\
y_{n+1} &= \sin x^1 \sin x^2 \sin x^3 \sin x^4 \dots \sin x^{n-1} \sin x^n
\end{aligned}$$

In the homework, you will be asked to obtain the associated chart by solving for the x^i. Note that if the sphere has radius r, then we can multiply all the above expressions by r, getting

$$\begin{aligned}
y_1 &= r \cos x^1 \\
y_2 &= r \sin x^1 \cos x^2 \\
y_3 &= r \sin x^1 \sin x^2 \cos x^3 \\
&\dots \\
y_{n-1} &= r \sin x^1 \sin x^2 \sin x^3 \sin x^4 \dots \cos x^{n-1} \\
y_n &= r \sin x^1 \sin x^2 \sin x^3 \sin x^4 \dots \sin x^{n-1} \cos x^n \\
y_{n+1} &= r \sin x^1 \sin x^2 \sin x^3 \sin x^4 \dots \sin x^{n-1} \sin x^n.
\end{aligned}$$

(d) The torus $T = S^1 \times S^1$, with the following four charts:

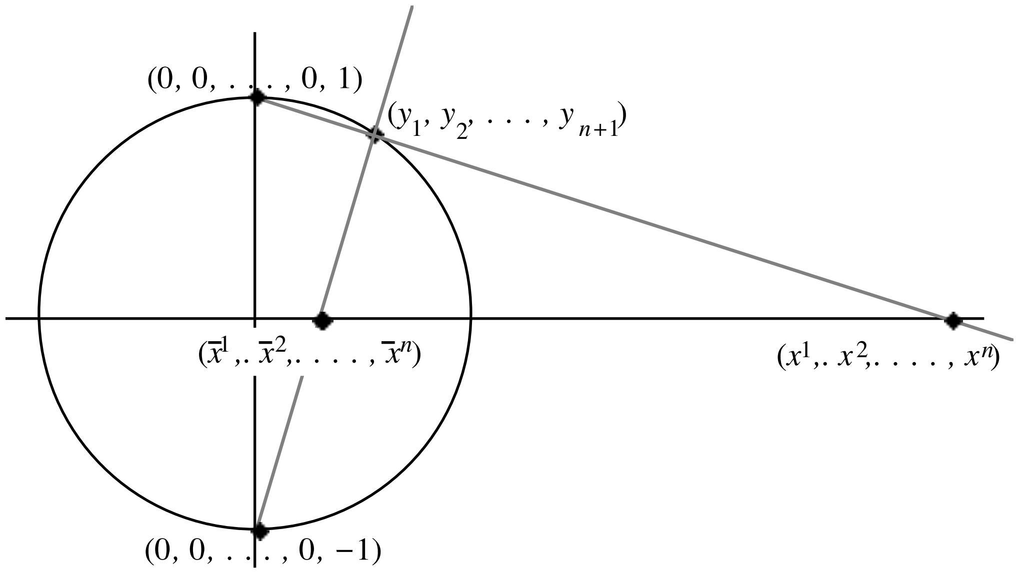
$\boldsymbol{x}$: $(S^1 - \{(1, 0)\}) \times (S^1 - \{(1, 0)\}) \longrightarrow E_2$, given by

$$\begin{aligned}
x^1((\cos\theta, \sin\theta), (\cos\phi, \sin\phi)) &= \theta \\
x^2((\cos\theta, \sin\theta), (\cos\phi, \sin\phi)) &= \phi.
\end{aligned}$$

The remaining charts are defined similarly, and the change-of-coordinate maps are omitted.

(e) The cylinder (homework)

(f) S^n, with (again) stereographic projection, is an n-manifold; the two charts are given as follows. Let P be the point $(0, 0, \ldots, 0, 1)$ and let Q be the point $(0, 0, \ldots, 0, -1)$. Then define two charts $(S^n - P, x^i)$ and $(S^n - Q, \bar{x}^i)$ as follows. (See the figure.)



If $(y_1, y_2, \ldots, y_n, y_{n+1})$ is a point in S^n, let

$$x^1 = \frac{y_1}{1-y_{n+1}}; \qquad \bar{x}^1 = \frac{y_1}{1+y_{n+1}};$$

$$x^2 = \frac{y_2}{1-y_{n+1}}; \qquad \bar{x}^2 = \frac{y_2}{1+y_{n+1}};$$

$$\ldots \qquad \ldots$$

$$x^n = \frac{y_n}{1-y_{n+1}}. \qquad \bar{x}^n = \frac{y_n}{1+y_{n+1}}.$$

We can invert these maps as follows: Let $r^2 = \Sigma_i x^i x^i$, and $\bar{r}^2 = \Sigma_i \bar{x}^i \bar{x}^i$. Then:

$$y_1 = \frac{2x^1}{r^2+1}; \qquad y_1 = \frac{2\bar{x}^1}{1+\bar{r}^2};$$

$$y_2 = \frac{2x^2}{r^2+1}; \qquad y_2 = \frac{2\bar{x}^2}{1+\bar{r}^2};$$

$$\ldots \qquad \ldots$$

$$y_n = \frac{2x^n}{r^2+1}; \qquad y_n = \frac{2\bar{x}^n}{1+\bar{r}^2};$$

$$y_{n+1} = \frac{r^2-1}{r^2+1}; \qquad y_{n+1} = \frac{1-\bar{r}^2}{1+\bar{r}^2}.$$

The change-of-coordinate maps are therefore:

$$x^1 = \frac{y_1}{1-y_{n+1}} = \frac{\dfrac{2\bar{x}^1}{1+\bar{r}^2}}{1 - \dfrac{1-\bar{r}^2}{1+\bar{r}^2}} = \frac{\bar{x}^1}{\bar{r}^2};$$

$$x^2 = \frac{\bar{x}^2}{\bar{r}^2};$$

$$\ldots$$

$$x^n = \frac{\bar{x}^n}{\bar{r}^2}.$$

This makes sense, since the maps are not defined when $\bar{x}^i = 0$ for all i, corresponding to the north pole.

Note

Since $\bar{r}$ is the distance from $\bar{x}^i$ to the origin, this map is hyperbolic reflection in the unit circle:

$$x^i = \frac{1}{\bar{r}} \frac{\bar{x}^i}{\bar{r}};$$

and squaring and adding gives

$$r = \frac{1}{\bar{r}}.$$

That is, project it to the circle, and invert the distance from the origin. This also gives the inverse relations, since we can write

$$\bar{x}^i = \bar{r}^2 x^i = \frac{x^i}{r^2}.$$

In other words, we have the following transformation rules.

Change of Coordinate Transformations for Stereographic Projection

Let $r^2 = \Sigma_i x^i x^i$, and $\bar{r}^2 = \Sigma_i \bar{x}^i \bar{x}^i$. Then

$$\bar{x}^i = \frac{x^i}{r^2}$$

$$x^i = \frac{\bar{x}^i}{\bar{r}^2}$$

$$r\bar{r} = 1$$

Note

We can put all the coordinate functions $x_\alpha{}^r$: $U_\alpha \longrightarrow E_1$ together to get a single map

$$x_\alpha: U_\alpha \longrightarrow W_\alpha \subset E_n.$$

A more precise formulation of condition (c) in the definition of a manifold is then the following: each W_α is an open subset of E_n, each x_α is invertible, and each composite

$$W_\alpha \xrightarrow{x_\alpha^{-1}} E_n \xrightarrow{x_\beta} W_\beta$$

is defined on an open subset and smooth.

We now want to discuss scalar and vector fields on manifolds, but how do we specify such things? First, a scalar field.

Definition 2.4 A **smooth scalar field** on a smooth manifold M is just a smooth real-valued map $\Phi\colon M \longrightarrow E_1$. (In other words, it is a smooth function of the coordinates of M as a subset of E_r.) Thus, Φ associates to each point m of M a unique scalar $\Phi(m)$. If U is a subset of M, then a **smooth scalar field on *U*** is smooth real-valued map $\Phi\colon U \longrightarrow E_1$. If $U \neq M$, we sometimes call such a scalar field **local**.

If Φ is a scalar field on M and x is a chart, then we can express Φ as a smooth function ϕ of the associated parameters $x^1, x^2, \ldots, x^n$. If the chart is $\bar{x}$, we shall write $\bar{\phi}$ for the function of the other parameters $\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^n$. Note that we must have $\phi = \bar{\phi}$ at each point of the manifold (see the transformation rule below).

Examples 2.5

(a) Let $M = E_n$ (with its usual structure) and let Φ be any smooth real-valued function in the usual sense. Then, using the identity chart, we have $\Phi = \phi$.

(b) Let $M = S^2$, and define $\Phi(y_1, y_2, y_3) = y_3$. Using stereographic projection, we find both ϕ and $\bar{\phi}$:

$$\phi(x^1, x^2) = y_3(x^1, x^2) = \frac{r^2-1}{r^2+1} = \frac{(x^1)^2 + (x^2)^2 - 1}{(x^1)^2 + (x^2)^2 + 1}$$

$$\bar{\phi}(\bar{x}^1, \bar{x}^2) = y_3(\bar{x}^1, \bar{x}^2) = \frac{1-\bar{r}^2}{1+\bar{r}^2} = \frac{1 - (\bar{x}^1)^2 - (\bar{x}^2)^2}{1 + (\bar{x}^1)^2 + (\bar{x}^2)^2}$$

(c) Local Scalar Field The most obvious candidate for local fields are the coordinate functions themselves. If U is a coordinate neighborhood, and $\boldsymbol{x} = \{x^i\}$ is a chart on U, then the maps x^i are local scalar fields.

Sometimes, as in the above example, we may wish to specify a scalar field purely by specifying it in terms of its local parameters; that is, by specifying the various functions ϕ instead of the single function Φ. The problem is, we can't just specify it any way we want, since it must give a value to each point in the manifold independently of local coordinates. That is, if a point $p \in M$ has local coordinates (x^j) with one chart and $(\bar{x}^h)$ with another, they must be related via the relationship

$$\bar{x}^j = \bar{x}^j(x^h).$$

Transformation Rule for Scalar Fields

$$\bar{\phi}(\bar{x}^j) = \phi(x^h)$$

Example 2.6 Look at Example 2.5(b) above. If you substituted $\bar{x}^i$ as a function of the x^j, you would get $\bar{\phi}(\bar{x}^1, \bar{x}^2) = \phi(x^1, x^2)$.

Exercise Set 2

1. Give the paraboloid $z = x^2 + y^2$ the structure of a smooth manifold.

2. Find a smooth atlas of E_2 consisting of three charts.

3. (a) Extend the method in Exercise 1 to show that the graph of any smooth function f: $E_2 \longrightarrow E_1$ can be given the structure of a smooth manifold.

(b) Generalize part (a) to the graph of a smooth function f: $E_n \longrightarrow E_1$.

4. Two atlases of the manifold M **give the same smooth structure** if their union is again a smooth atlas of M.

(a) Show that the smooth atlases (E_1, f), and (E_1, g), where $f(x) = x$ and $g(x) = x^3$ are incompatible.

(b) Find a third smooth atlas of E_1 that is incompatible with both the atlases in part (a).

5. Consider the ellipsoid $L \subset E_3$ specified by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \qquad (a, b, c \neq 0).$$

Define f: $L \longrightarrow S^2$ by $f(x, y, z) = \left(\frac{x}{a}, \frac{y}{b}, \frac{z}{c}\right)$.

(a) Verify that f is invertible (by finding its inverse).

(b) Use the map f, together with a smooth atlas of S^2, to construct a smooth atlas of L.

6. Find the chart associated with the generalized spherical polar coordinates described in Example 2.3(c) by inverting the coordinates. How many additional charts are needed to get an atlas? Give an example.

7. Obtain the equations in Example 2.3(f).

---

3. Tangent Vectors and the Tangent Space

We now turn to vectors tangent to smooth manifolds. We must first talk about smooth paths on M.

Definition 3.1 A **smooth path** on M is a smooth map $\boldsymbol{r}$: $(-1, 1) \longrightarrow M$, where $\boldsymbol{r}(t) = (y_1(t), y_2(t), \ldots, y_s(t))$. We say that r is a smooth path **through** $\boldsymbol{m} \in \boldsymbol{M}$ if $\boldsymbol{r}(t_0) = m$ for some $t_0 \in (-1, 1)$. We can specify a path in M at m by its coordinates:

$y_1 = y_1(t),$
$y_2 = y_2(t),$
$. . .$
$y_s = y_s(t),$

where m is the point $(y_1(t_0), y_2(t_0), . . . , y_s(t_0))$. Equivalently, since the ambient and local coordinates are functions of each other, we can also express a path—at least that part of it inside a coordinate neighborhood—in terms of its local coordinates:

$x^1 = x^1(),$
$x^2 = x^2(t),$
$. . .$
$x^n = x^n(t).$

Examples 3.2
(a) Smooth paths in E_n
(b) A smooth path in S^1, and S^n

Definition 3.3 A **tangent vector** at $m \in M \subset E_r$ is a vector $\boldsymbol{v}$ in E_r of the form

$$\boldsymbol{v} = \boldsymbol{y}'(t_0)$$

for some path $\boldsymbol{y} = \boldsymbol{y}(t)$ in M through m and $\boldsymbol{y}(t_0) = m$.

Examples 3.4
(a) Let M be the surface $y_3 = y_1^2 + y_2^2$, which we paramaterize by

$y_1 = x^1$
$y_2 = x^2$
$y_3 = (x^1)^2 + (x^2)^2$

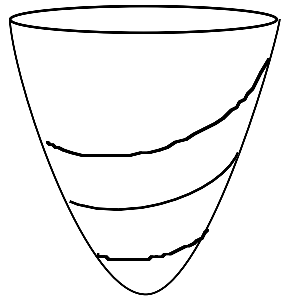
This corresponds to the single chart $(U=M; x^1, x^2)$, where

$x^1 = y_1$ and $x^2 = y_2$.

To specify a tangent vector, let us first specify a path in M, such as

$y_1 = \sqrt{t}\ \sin t$
$y_2 = \sqrt{t}\ \cos t$
$y_3 = t$

(Check that the equation of the surface is satisfied.) This gives the path shown in the figure.



Now we obtain a tangent vector field along the path by taking the derivative:

$$\left(\frac{dy_1}{dt}, \frac{dy_2}{dt}, \frac{dy_3}{dt}\right) = \left(\sqrt{t}\,\cos t + \frac{\sin t}{2\sqrt{t}},\ -\sqrt{t}\,\sin t + \frac{\cos t}{2\sqrt{t}},\ 1\right).$$

(To get actual tangent vectors at points in M, evaluate this at a fixed point t_0.)

Note We can also express the coordinates x^i in terms of t:

$$x^1 = y_1 = \sqrt{t}\,\sin t$$
$$x^2 = y_2 = \sqrt{t}\,\cos t$$

This descibed a path in some chart (that is, in coordinate space) rather than on the mnanifold itself. We can also take the derivative,

$$\left(\frac{dx^1}{dt}, \frac{dx^2}{dt}\right) = \left(\sqrt{t}\,\cos t + \frac{\sin t}{2\sqrt{t}},\ -\sqrt{t}\,\sin t + \frac{\cos t}{2\sqrt{t}}\right).$$

We also think of this as the tangent vector, given in terms of the local coordinates. A lot more will be said about the relationship between the above two forms of the tangent vector below.

Algebra of Tangent Vectors: Addition and Scalar Multiplication

The sum of two tangent vectors is, geometrically, also a tangent vector, and the same goes for scalar multiples of tangent vectors. However, we have defined tangent vectors using paths in M, and we cannot produce these new vectors by simply adding or scalar-multiplying the corresponding paths: if $\boldsymbol{y} = \boldsymbol{f}(t)$ and $\boldsymbol{y} = \boldsymbol{g}(t)$ are two paths through $m \in M$ where $\boldsymbol{f}(t_0) = \boldsymbol{g}(t_0) = m$, then adding them coordinate-wise need not produce a path in M. However, we *can* add these paths using some chart as follows.

Choose a chart x at m, with the property (for convenience) that $x(m) = \mathbf{0}$. Then the paths $x(f(t))$ and $x(g(t))$ (defined as in the note above) give two paths through the origin in coordinate space. *Now* we can add these paths or multiply them by a scalar without leaving coordinate space and then use the chart map to lift the result back up to M. In other words, define

$$(\boldsymbol{f}+\boldsymbol{g})(t) = x^{-1}(x(\boldsymbol{f}(t)) + x(\boldsymbol{g}(t)))$$

and $$(\lambda\boldsymbol{f})(t) = x^{-1}(\lambda x(\boldsymbol{f}(t))).$$

Taking their derivatives at the point t_0 will, by the chain rule, produce the sum and scalar multiples of the corresponding tangent vectors. Since we can add and scalar-multiply tangent vectors

Definition 3.5 If M is an n-dimensional manifold, and $m \in M$, then the **tangent space at *m*** is the set T_m of all tangent vectors at m.

The above constructions turn T_m into a **vector space**.

Let us return to the issue of the two ways of describing the coordinates of a tangent vector at a point $m \in M$: writing the path as $y_i = y_i(t)$ we get the **ambient coordinates** of the tangent vector:

$$\boldsymbol{y}'(t_0) = \left(\frac{dy_1}{dt}, \frac{dy_2}{dt}, \ldots\ldots, \frac{dy_s}{dt}\right)_{t=t_0} \qquad \text{Ambient coordinates}$$

and, using some chart x at m, we get the **local coordinates**

$$\boldsymbol{x}'(t_0) = \left(\frac{dx^1}{dt}, \frac{dx^2}{dt}, \ldots, \frac{dx^n}{dt}\right)_{t=t_0}.$$

Question In general, how are the dx^i/dt related to the dy_i/dt?
Answer By the chain rule,

$$\frac{dy_1}{dt} = \frac{\partial y_1}{\partial x^1}\frac{dx^1}{dt} + \frac{\partial y_1}{\partial x^2}\frac{dx^2}{dt},$$

and similarly for dy_2/dt and dy_3/dt. Thus, *we can recover the original three ambient vector coordinates from the local coordinates.* In other words, *the local vector coordinates completely specify the tangent vector.*

Note The chain rule as used above shows us how to convert local coordinates to ambient coordinates and vice-versa:

Converting Between Local and Ambient Coordinates of a Tangent Vector
If the tangent vector V has ambient coordinates $(v_1, v_2, \ldots, v_s)$ and local coordinates $(v^1, v^2, \ldots, v^n)$, then they are related by the formulæ

$$v_i = \sum_{k=1}^{n} \frac{\partial y_i}{\partial x^k} v^k$$

and

$$v^i = \sum_{k=1}^{s} \frac{\partial x^i}{\partial y_k} v_k$$

Note To obtain the coordinates of sums or scalar multiples of tangent vectors, simply take the corresponding sums and scalar multiples of the coordinates. In other words:

$$(v+w)^i = v^i + w^i$$

and $$(\lambda v)^i = \lambda v^I$$

just as we would expect to do for ambient coordinates. (Why can we do this?)

Examples 3.4 Continued:
(b) Take $M = E_n$, and let $\boldsymbol{v}$ be any vector in the usual sense with coordinates α^i. Choose x to be the usual chart $x^i = y_i$. If $\boldsymbol{p} = (p^1, p^2, \ldots, p^n)$ is a point in M, then $\boldsymbol{v}$ is the derivative of the path

$$x^1 = p^1 + t\alpha^1$$
$$x^2 = p^2 + t\alpha^2;$$
$$\ldots$$
$$x^n = p^n + t\alpha^n$$

at $t = 0$. Thus this vector has local and ambient coordinates equal to each other, and equal to

$$\frac{dx^i}{dt} = \alpha^i,$$

which are the same as the original coordinates. In other words, the tangent vectors are "the same" as ordinary vectors in E_n.

(c) Let $M = S^2$, and the path in S^2 given by

$$y_1 = \sin t$$
$$y_2 = 0$$
$$y_3 = \cos t$$

This is a path (circle) through $m = (0, 0, 1)$ following the line of longitude $\phi = x^2 = 0$, and has tangent vector

$$\left(\frac{dy_1}{dt}, \frac{dy_2}{dt}, \frac{dy_3}{dt}\right) = (\cos t, 0, -\sin t) = (1, 0, 0) \text{ at the point } m.$$

(d) We can also use the local coordinates to describe a path; for instance, the path in part (b) can be described using spherical polar coordinates by

$$x^1 = t$$
$$x^2 = 0$$

The derivative

$$\left(\frac{dx^1}{dt}, \frac{dx^2}{dt}\right) = (1, 0)$$

gives the local coordinates of the tangent vector itself (the coordinates of its image in coordinate Euclidean space).

(e) In general, if $(U; x^1, x^2, \ldots, x^n)$ is a coordinate system near m, then we can obtain paths $y_i(t)$ by setting

$$x^j(t) = \begin{cases} t + \text{const.} & \text{if } j = i \\ \text{const.} & \text{if } j \neq i \end{cases},$$

where the constants are chosen to make $x^i(t_0)$ correspond to m for some t_0. (The paths in (c) and (d) are an example of this.) To view this as a path in M, we just apply the parametric equations $y_i = y_i(x^j)$, giving the y_i as functions of t.

The associated tangent vector at the point where $t = t_0$ is called $\partial/\partial x^i$. It has local coordinates

$$v^j = \left(\frac{dx^j}{dt}\right)_{t = t_0} = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} = \delta_i^j$$

δ_i^j is called the **Kronecker Delta**, and is defined by

$$\delta_i^j = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases}.$$

We can now get the ambient coordinates by the above conversion:

$$v_j = \sum_{k=1}^{n} \frac{\partial y_j}{\partial x^k} v^k = \sum_{k=1}^{n} \frac{\partial y_j}{\partial x^k} \delta_i^k = \frac{\partial y_j}{\partial x^i} .$$

We call this vector $\frac{\partial}{\partial x^i}$. Summarizing,

Definition of $\boldsymbol{\frac{\partial}{\partial x^i}}$

Pick a point $m \in M$. Then $\frac{\partial}{\partial x^i}$ is the vector at m whose **local coordinates** are given by

$$\begin{aligned} j \text{ th coordinate} &= \left(\frac{\partial}{\partial x^i}\right)^j \\ &= \delta_i^j = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} \\ &= \frac{\partial x^j}{\partial x^i} \end{aligned}$$

Its **ambient coordinates** are given by

$$j \text{ th coordinate} = \frac{\partial y_j}{\partial x^i}$$

(everything evaluated at t_0) Notice that the path itself has disappeared from the definition...

Now that we have a better feel for local and ambeinet coordinates of vectors, let us state some more "general nonsense": Let M be an n-dimensional manifold, and let $m \in M$.

Proposition 3.6 ***(The Tangent Space)***
There is a one-to-one correspondence between tangent vectors at m and plain old vectors in E_n. In other words, the tangent space "looks like" E_n. Technically, this correspondnece is a linear ismorphism.

Proof *(and this will explain why local coordinates are better than ambient ones)*
Let T_m be the set of tangent vectors at m (that is, the tangent space), and define

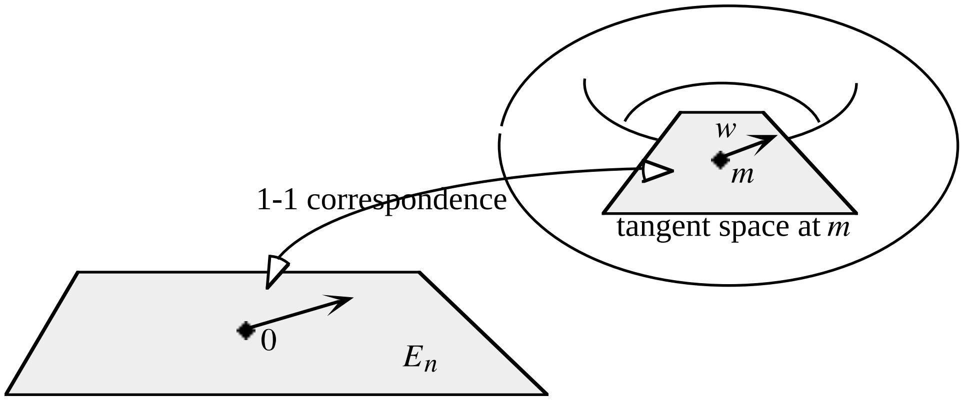
$$F: T_m \longrightarrow E_n$$

by assigning to a typical tangent vector its n local coordinates. Define an inverse

$$G: E_n \longrightarrow T_m$$

by the formula $G(v^1, v^2, \ldots, v^n) = v^1\frac{\partial}{\partial x^1} + v^2\frac{\partial}{\partial x^2} + \ldots + v^n\frac{\partial}{\partial x^n}$

$$= \Sigma_i v^i \frac{\partial}{\partial x^i} .$$



Then we can verify that F and G are inverses as follows:

$$F(G(v^1, v^2, \ldots, v^n)) = F(\Sigma_i v^i \frac{\partial}{\partial x^i})$$

$$= \text{local coordinates of the vector } v^1\frac{\partial}{\partial x^1} + v^2\frac{\partial}{\partial x^2} + \ldots + v^n\frac{\partial}{\partial x^n} .$$

But, in view of the simple local coordinate structure of the vectors $\frac{\partial}{\partial x^i}$, the i th coordinate of this field is

$$v^1(0) + \ldots + v^{i-1}(0) + v^i(1) + v^{i+1}(0) = \ldots + v^n(0) = v^i.$$

In other words,

$$i \text{ th coordinate of } F(G(v)) = F(G(v))^i = v^i,$$

so that $F(G(v)) = v$. Conversely,

$$G(F(w)) = w^1\frac{\partial}{\partial x^1} + w^2\frac{\partial}{\partial x^2} + \ldots + w^n\frac{\partial}{\partial x^n} ,$$

where w^i are the local coordinates of the vector w. Is this the same vector as w? Well, let us look at the ambient coordinates; since if two vectors have the same *ambient* coordinates, they are certainly the same vector! But we know how to find the ambient coordinates of each term in the sum. So, the j th ambient coordinate of $G(F(w))$ is

$$G(F(w))_j = w^1\frac{\partial y_j}{\partial x^1} + w^2\frac{\partial y_j}{\partial x^2} + \ . \ . \ . + w^n\frac{\partial y_j}{\partial x^n}$$

(using the formula for the ambient coordinates of the $\partial/\partial x^i$)

$= w_j$ (using the conversion formulas)

Therefore, $G(F(w)) = w$, and we are done. ✪

That is why we use local coordinates; there is no need to specify a path every time we want a tangent vector!

Note Under the one-to-one correspondence in the proposition, the standard basis vectors in E_n correspond to the tangent vectors $\partial/\partial x^1, \partial/\partial x^2, \ldots, \partial/\partial x^n$. Therefore, the latter vectors are a basis of the tangent space T_m.

---

1. Suppose that $\boldsymbol{v}$ is a tangent vector at $m \in M$ with the property that there exists a local coordinate system x^i at m with $v^i = 0$ for every i. Show that $\boldsymbol{v}$ has zero coordinates in every coefficient system, and that, in fact, $\boldsymbol{v} = \mathbf{0}$.

2. **(a)** Calculate the ambient coordinates of the vectors $\partial/\partial\theta$ and $\partial/\partial\phi$ at a general point on S^2, where θ and ϕ are spherical polar coordinates ($\theta = x^1$, $\phi = x^2$).
(b) Sketch these vectors at some point on the sphere.

3. Prove that $\dfrac{\partial}{\partial \bar{x}^i} = \dfrac{\partial x^j}{\partial \bar{x}^i}\dfrac{\partial}{\partial x^j}$.

4. Consider the torus T^2 with the chart x given by

$y_1 = (a+b \cos x^1)\cos x^2$

$y_2 = (a+b \cos x^1)\sin x^2$

$y_3 = b \sin x^1$

$0 < x^i < 2\pi$. Find the ambeint coordinates of the two orthogonal tangent vectors at a general point, and sketch the resulting vectors.

---

4. Contravariant and Covariant Vector Fields

Question How are the local coordinates of a given tangent vector for one chart related to those for another?

Answer Again, we use the chain rule. The formula

$$\frac{d\bar{x}^i}{dt} = \frac{\partial \bar{x}^i}{\partial x^j}\frac{dx^j}{dt}$$

(Note: we are using the **Einstein Summation Convention**: repeated index implies summation) tells us how the coordinates transform. In other words, a tangent vector through a point m in M is a collection of n numbers $v^i = dx^i/dt$ (specified for each chart x at m) where the quantities for one chart are related to those for another according to the formula

$$\bar{v}^i = \frac{\partial \bar{x}^i}{\partial x^j} v^j.$$

This leads to the following definition.

Definition 4.1 A **contravariant vector** at $m \in M$ is a collection v^i of n quantities (defined for each chart at m) which transform according to the formula

$$\bar{v}^i = \frac{\partial \bar{x}^i}{\partial x^j} v^j.$$

It follows that contravariant vectors "are" just tangent vectors: the contravariant vector v^i corresponds to the tangent vector given by

$$\boldsymbol{v} = v^i \frac{\partial}{\partial x^i},$$

so we shall henceforth refer to tangent vectors and contravariant vectors.

A **contravariant vector field** V **on** $\boldsymbol{M}$ associates with each chart x a collection of n smooth real-valued **coordinate functions** V^i of the n variables $(x^1, x^2, \ldots, x^n)$, such that evaluating V^i at any point gives a vector at that point. Further, the domain of the V^i is the whole of the range of $\boldsymbol{x}$. Similarly, a **contravariant vector field** V **on** $\boldsymbol{U} \subset \boldsymbol{M}$ is defined in the same way, but its domain is restricted to $x(\boldsymbol{U})$.

Thus, the coordinates of a smooth vector field transform the same way:

Contravariant Vector Transformation Rule

$$\bar{V}^i = \frac{\partial \bar{x}^i}{\partial x^j} V^j$$

where now the V^i and $\bar{V}^j$ are functions of the associated coordinates $(x^1, x^2, \ldots, x^n)$, rather than real numbers.

Notes 4.2

1. The above formula is reminiscent of matrix multiplication: In fact, if $\bar{D}$ is the matrix whose ij th entry is $\frac{\partial \bar{x}^i}{\partial x^j}$, then the above equation becomes, in matrix form:

$$\bar{V} = \bar{D}V,$$

where we think of V and $\bar{V}$ as column vectors.

2. By "transform," we mean that the above relationship holds between the coordinate functions V^i of the x^i associated with the chart x, and the functions $\bar{V}^i$ of the $\bar{x}^i$, associated with the chart $\bar{x}$.

3. Note the formal symbol cancellation: if we cancel the ∂'s, the x's, and the superscripts on the right, we are left with the symbols on the left!
4. From the proof of 3.6, we saw that, if $\boldsymbol{V}$ is any smooth contravariant vector field on M, then

$$\boldsymbol{V} = V^j \frac{\partial}{\partial x^j} .$$

Examples 4.3
(a) Take $M = E_n$, and let $\boldsymbol{F}$ be any (tangent) vector field in the usual sense with coordinates F^i. If $\boldsymbol{p} = (p^1, p^2, \ldots, p^n)$ is a point in M, then $\boldsymbol{v}$ is the derivative of the path

$$\begin{aligned} x^1 &= p^1 + tF^1 \\ x^2 &= p^2 + tF^2; \\ &\ldots \\ x^n &= p^n + tF^n \end{aligned}$$

at $t = 0$. Thus this vector field has (ambient and local) coordinate functions

$$\frac{dx^i}{dt} = F^i,$$

which are the same as the original coordinates. In other words, the tangent vectors fields are "the same" as ordinary vector fields in E_n.

(b) An Important Local Vector Field Recall from Examples 3.4 (e) above the definition of the vectors $\partial/\partial x^i$: At each point m in a manifold M, we have the n vectors $\partial/\partial x^1, \partial/\partial x^2, \ldots, \partial/\partial x^n$, where the typical vector $\partial/\partial x^i$ was obtained by taking the derivative of the path:

$$\frac{\partial}{\partial x^i} = \text{vector obtained by differentiating the path } x^j(t) = \begin{cases} t + \text{const.} & \text{if } j = i \\ \text{const.} & \text{if } j \neq i \end{cases} ,$$

where the constants are chosen to make $x^i(t_0)$ correspond to m for some t_0. This gave

$$\left(\frac{\partial}{\partial x^i}\right)^j = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} .$$

Now, there is nothing to stop us from defining n different *vector fields* $\partial/\partial x^1, \partial/\partial x^2, \ldots, \partial/\partial x^n$, in exactly the same way: at each point in the coordinate neighborhood of the chart x, associate the vector above.

Note: $\dfrac{\partial}{\partial x^i}$ is a *field*, and not the *i*th coordinate of a field. Its *j*th coordinate under the chart *x* is given by

$$\left(\frac{\partial}{\partial x^i}\right)^j = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} \quad = \delta_i^j = \frac{\partial x^j}{\partial x^i}\,.$$

at every point in the image of *x,* and is called the **Kronecker Delta**, δ_i^j. More about that later.

Question Since the coordinates do not depend on *x*, does it mean that the vector field $\partial/\partial x^i$ is constant?
Answer No. Remember that a tangent field is a field on (part of) a *manifold*, and as such, it is not, in general, constant. The only thing that is constant are its coordinates under the *specific chart x.* The corresponding coordinates under another chart $\bar{x}$ are $\partial\bar{x}^j/\partial x^i$ (which are not constant in general).

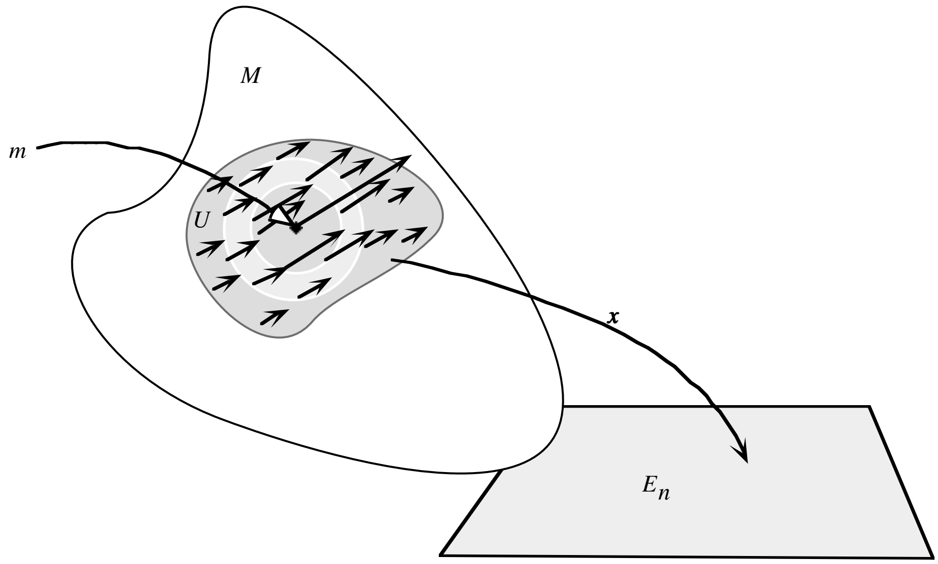
(c) Patching Together Local Vector Fields The vector field in the above example has the disadvantage that is local. We can "extend" it to the whole of *M* by making it zero near the boundary of the coordinate patch, as follows. If $m \in M$ and $\boldsymbol{x}$ is any chart of *M*, lat $x(m) = y$ and let *D* be a disc or some radius *r* centered at *y* entirely contained in the image of *x*. Now define a vector field on the whole of *M* by

$$\boldsymbol{w}(p) = \begin{cases} \dfrac{\partial}{\partial x^j}\, e^{-R^2} & \text{if } p \text{ is in } D \\ 0 & \text{otherwise} \end{cases}$$

where

$$R = \frac{|x(p) - y|}{r - |x(p) - y|}\,.$$

The following figure shows what this field looks like on *M*.

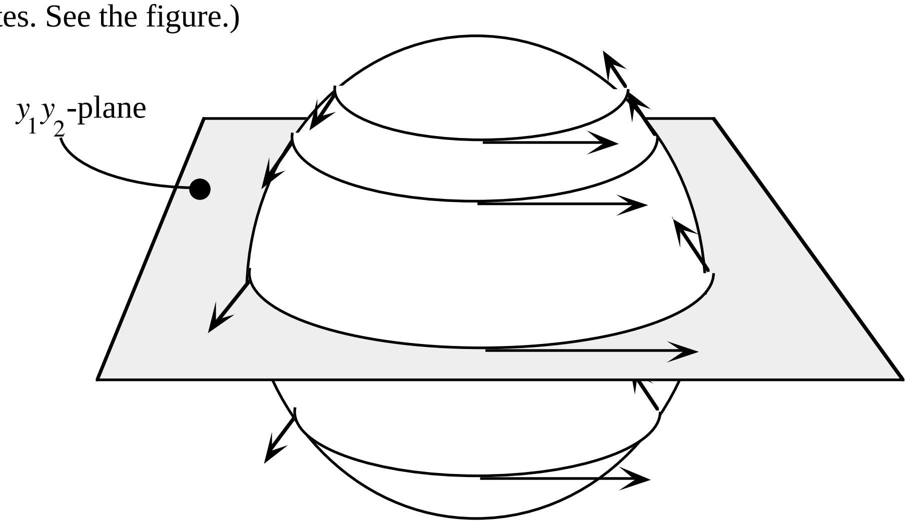


The fact that $\overline{V}^i$ is a smooth function of the $\overline{x}^i$ now follows from the fact that all the partial derivatives of all orders vanish as you leave the domain of $\boldsymbol{x}$.

(d) Take $M = S^n$, with stereographic projection given by the two charts discussed earlier. Consider the circulating vector field on S^n defined at the point $y = (y_1, y_2, \ldots, y_n, y_{n+1})$ by the paths

t ♋ $(y_1\cos t - y_2\sin t,\ y_1\sin t + y_2\cos t,\ y_3, \ldots, y_{n+1})$.

(For fixed $y = (y_1, y_2, \ldots, y_n, y_{n+1})$ this defines a path at the point y—see Example 3.2(c) in the web site) This is a circulating field in the y_1y_2-plane—look at spherical polar coordinates. See the figure.)



Note: Length of tangent vector = radius of circle

In terms of the charts, the local coordinates of this field are:

$$x^1 = \frac{y_1}{1-y_{n+1}} = \frac{y_1\cos t - y_2\sin t}{1-y_{n+1}}; \qquad \text{so } V^1 = \frac{dx^1}{dt} = -\frac{y_1\sin t + y_2\cos t}{1-y_{n+1}} = -x^2$$
$$x^2 = \frac{y_2}{1-y_{n+1}} = \frac{y_1\sin t + y_2\cos t}{1-y_{n+1}}; \qquad \text{so } V^2 = \frac{dx^2}{dt} = \frac{y_1\cos t - y_2\sin t}{1-y_{n+1}} = x^1$$
$$x^3 = \frac{y_3}{1-y_{n+1}}; \qquad \text{so } V^3 = \frac{dx^3}{dt} = 0$$
$$\ldots$$
$$x^n = \frac{y_n}{1-y_{n+1}}; \qquad \text{so } V^n = \frac{dx^n}{dt} = 0.$$

and

$$\bar{x}^1 = \frac{y_1}{1+y_{n+1}} = \frac{y_1\cos t - y_2\sin t}{1+y_{n+1}}; \qquad \text{so } \bar{V}^1 = \frac{d\bar{x}^1}{dt} = -\frac{y_1\sin t + y_2\cos t}{1+y_{n+1}} = -\bar{x}^2$$
$$\bar{x}^2 = \frac{y_2}{1+y_{n+1}} = \frac{y_1\sin t + y_2\cos t}{1+y_{n+1}}; \qquad \text{so } \bar{V}^2 = \frac{d\bar{x}^2}{dt} = \frac{y_1\cos t - y_2\sin t}{1+y_{n+1}} = \bar{x}^1$$
$$\bar{x}^3 = \frac{y_3}{1+y_{n+1}}; \qquad \text{so } \bar{V}^3 = \frac{d\bar{x}^3}{dt} = 0$$
$$\ldots$$
$$\bar{x}^n = \frac{y_n}{1+y_{n+1}}; \qquad \text{so } \bar{V}^n = \frac{d\bar{x}^n}{dt} = 0.$$

Now let us check that they transform according to the contravariant vector transformation rule. First, we saw above that

$$\bar{x}^i = \frac{x^i}{r^2},$$

and hence

$$\frac{\partial \bar{x}^i}{\partial x^j} = \begin{cases} \dfrac{r^2-2(x^i)^2}{r^4} & \text{if } j = i \\ \dfrac{-2x^i x^j}{r^4} & \text{if } j \neq i \end{cases}.$$

In matrix form, this is:

$$\bar{D} = \frac{1}{r^4}\begin{bmatrix} r^2-2(x^1)^2 & -2x^1x^2 & -2x^1x^3 & \ldots & -2x^1x^n \\ -2x^2x^1 & r^2-2(x^2)^2 & -2x^2x^3 & \ldots & -2x^2x^n \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -2x^nx^1 & -2x^nx^2 & -2x^nx^3 & \ldots & r^2-2(x^n)^2 \end{bmatrix}$$

Thus,

$$\bar{D}V = \frac{1}{r^4}\begin{bmatrix} r^2-2(x^1)^2 & -2x^1x^2 & -2x^1x^3 & \dots & -2x^1x^n \\ -2x^2x^1 & r^2-2(x^2)^2 & -2x^2x^3 & \dots & -2x^2x^n \\ \dots & \dots & \dots & \dots & \dots \\ -2x^nx^1 & -2x^nx^2 & -2x^nx^3 & \dots & r^2-2(x^n)^2 \end{bmatrix}\begin{bmatrix} -x^2 \\ x^1 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

$$= \frac{1}{r^4}\begin{bmatrix} -x^2r^2 + 2(x^1)^2x^2 - 2(x^1)^2x^2 \\ 2(x^2)^2x^1 + r^2x^1 - 2(x^2)^2x^1 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -x^2/r^2 \\ x^1/r^2 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} = \begin{bmatrix} -\bar{x}^2 \\ \bar{x}^1 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} = \bar{V}.$$

Covariant Vector Fields

We now look at the (local) gradient. If ϕ is a smooth scalar field on M, and if x is a chart, then we obtain the locally defined vector field $\partial\phi/\partial x^i$. By the chain rule, these functions transform as follows:

$$\frac{\partial\phi}{\partial\bar{x}^i} = \frac{\partial\phi}{\partial x^j}\frac{\partial x^j}{\partial\bar{x}^i},$$

or, writing $C_i = \partial\phi/\partial x^i$,

$$\bar{C}_i = \frac{\partial x^j}{\partial\bar{x}^i}C_j .$$

This leads to the following definition.

Definition 4.4 A **covariant vector field** C on M associates with each chart x a collection of n smooth functions $C_i(x^1, x^2, \dots, x^n)$ which satisfy:

Covariant Vector Transformation Rule

$$\bar{C}_i = C_j\frac{\partial x^j}{\partial\bar{x}^i}$$

Notes 4.5

1. If $\underline{D}$ is the matrix whose ij th entry is $\dfrac{\partial x^i}{\partial\bar{x}^j}$, then the above equation becomes, in matrix form:

$$\bar{C} = C\underline{D},$$

where now we think of C and $\bar{C}$ as *row* vectors.

2. Note that

$$(\underline{D}\bar{D})^i{}_j = \frac{\partial x^i}{\partial \bar{x}^k}\frac{\partial \bar{x}^k}{\partial x^j} = \frac{\partial x^i}{\partial x^j} = \delta^i{}_j,$$

and similarly for $\bar{D}\underline{D}$. Thus, $\bar{D}$ and $\underline{D}$ are inverses of each other.

3. Note again the formal symbol cancellation: if we cancel the ∂'s, the x's, and the superscripts on the right, we are left with the symbols on the left!

4. *Guide to memor*y: In the contravariant objects, the barred x goes on top; in covariant vectors, on the bottom. In both cases, the non-barred indices matches.

Question Geometrically, a contravariant vector is a vector that is tangent to the manifold. How do we think of a covariant vector?

Answer The key to the answer is this:

Note From now on, all scalar and vector fields are assumed **smooth.**

Definition 4.6 A smooth **1-form**, or smooth **cotangent vector field** on the manifold M (or on an open subset U of M) is a function F that assigns to each smooth tangent vector field $\boldsymbol{V}$ on M (or on an open subset U) a smooth scalar field $F(\boldsymbol{V})$, which has the following properties:

$$F(\boldsymbol{V}+\boldsymbol{W}) = F(\boldsymbol{V}) + F(\boldsymbol{W})$$
$$F(\alpha\boldsymbol{V}) = \alpha F(\boldsymbol{V}).$$

for every pair of tangent vector fields $\boldsymbol{V}$ and $\boldsymbol{W}$, and every scalar α. (In the language of linear algebra, this says that F is a linear transformation.)

Proposition 4.7 ***(Covariant Fields are One-Form Fields)***
There is a one-to-one correspondence between covariant vector fields on M (or U) and 1-forms on M (or U). Thus, we can think of covariant tangent fields as nothing more than 1-forms.

Proof Here is the one-to-one correspondence. Let $\mathcal{F}$ be the family of 1-forms on M (or U) and let C be the family of covariant vector fields on M (or U). Define

$$\Phi: C \longrightarrow \mathcal{F}$$

by

$$\Phi(C_i)(V^j) = C_k V^k.$$

In the homework, we see that $C_k V^k$ is indeed a scalar by checking the transformation rule:

$$\bar{C}_k \bar{V}^k = C_l V^l.$$

The linearity property of Φ now follows from the distributive laws of arithmetic. We now define the inverse

$$\Psi: \mathcal{F} \longrightarrow C$$

by

$$(\Psi(F))_i = F(\partial/\partial x^i).$$

We need to check that this is a covariant vector field; that is, that it transforms in the correct way. But, it x and $\bar{x}$ are two charts, then

$$F(\frac{\partial}{\partial \bar{x}^i}) = F(\frac{\partial x^j}{\partial \bar{x}^i}\frac{\partial}{\partial x^j}) \text{ (if you don't believe this, look at the ambient coordinates)}$$

$$= \frac{\partial x^j}{\partial \bar{x}^i} F(\frac{\partial}{\partial x^j}),$$

by linearity.

That Ψ and Φ are in fact inverses is left to the exercise set. ❇

Examples 4.8

(a) Let $M = S^1$ with the charts:

$$x = \arg(z),\ \bar{x} = \arg(-z)$$

discussed in §2. There, we saw that the change-of-coordinate maps are given by

$$x = \begin{cases} \bar{x}+\pi & \text{if } \bar{x} \le \pi \\ \bar{x}-\pi & \text{if } \bar{x} \ge \pi \end{cases} \qquad \bar{x} = \begin{cases} x+\pi & \text{if } x \le \pi \\ x-\pi & \text{if } x \ge \pi \end{cases},$$

with

$$\partial\bar{x}/\partial x = \partial x/\partial\bar{x} = 1,$$

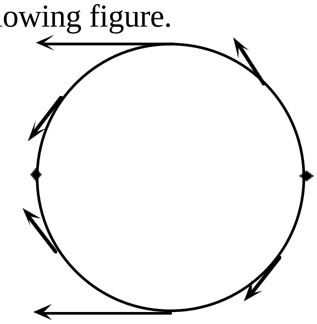
so that the change-of-coordinates do nothing. It follows that functions C and $\bar{C}$ specify a covariant vector field iff $C = \bar{C}$. (Then they are automatically a contravariant field as well). For example, let

$$C(x) = 1 = \bar{C}(\bar{x}).$$

This field circulates around S^1. On the other hand, we could define

$$C(x) = \sin x \text{ and } \bar{C}(\bar{x}) = -\sin\bar{x} = \sin x.$$

This field is illustrated in the following figure.



(The length of the vector at the point $e^{i\theta}$ is given by $\sin\theta$.)

(b) Let ϕ be a scalar field. Its ambient gradient, **grad** ϕ, is given by

$$\mathbf{grad}\,\phi = [\frac{\partial\phi}{\partial y_1}, , \ldots , \frac{\partial\phi}{\partial y_s}],$$

that is, the garden-variety gradient you learned about in calculus. This gradient is, in general, neither covariant or contravariant. However, we can use it to obtain a 1-form as follows: If $\boldsymbol{V}$ is any contravariant vector field, then the rate of change of ϕ along $\boldsymbol{V}$ is given by $\boldsymbol{V}.\mathbf{grad}\,\phi$. (If $\boldsymbol{V}$ happens to be a unit vector at some point, then this is the directional derivative at that point.) In other words, dotting with **grad** ϕ assigns to each contravariant vector field the scalar field $F(\boldsymbol{v}) = \boldsymbol{V}.\mathbf{grad}\,\phi$ which tells it how fast ϕ is changing along $\boldsymbol{V}$. We also get the 1-form identities:

$$F(\boldsymbol{V}+\boldsymbol{W}) = F(\boldsymbol{V}) + F(\boldsymbol{W})$$
$$F(\alpha\boldsymbol{V}) = \alpha F(\boldsymbol{V}).$$

The coordinates of the corresponding covariant vector field are

$$\begin{aligned} F(\partial/\partial x^i) &= (\partial/\partial x^i).\mathbf{grad}\,\phi \\ &= [\frac{\partial y_1}{\partial x^i}, \frac{\partial y_2}{\partial x^i}, \ldots , \frac{\partial y_s}{\partial x^i}] \cdot [\frac{\partial\phi}{\partial y_1}, , \ldots , \frac{\partial\phi}{\partial y_s}], \\ &= \frac{\partial\phi}{\partial x^i}, \end{aligned}$$

which is the example that first motivated the definition.

(c) Generalizing (b), let $\boldsymbol{\Sigma}$ be any smooth vector field (in E_s) defined on M. Then the operation of dotting with $\boldsymbol{\Sigma}$ is a linear function from smooth tangent fields on M to smooth scalar fields. Thus, it is a cotangent field on M with local coordinates given by applying the linear function to the canonical charts $\partial/\partial x^i$:

$$C_i = \frac{\partial}{\partial x^i} \cdot \boldsymbol{\Sigma}.$$

The gradient is an example of this, since we are taking

$$\boldsymbol{\Sigma} = \mathbf{grad}\,\phi$$

in the preceding example.

Note that, in general, dotting with $\boldsymbol{\Sigma}$ depends only on the tangent component of $\boldsymbol{\Sigma}$. This leads us to the next example.

(d) If V is any tangent (contravariant) field, then we can appeal to (c) above and obtain an associated covariant field. The coordinates of this field are not the same as those of V. To find them, we write:

$$\boldsymbol{V} = V^i \frac{\partial}{\partial x^i} \qquad \text{(See Note 4.2 (4).)}$$

Hence,

$$C_j = \frac{\partial}{\partial x^j} \cdot V^i \frac{\partial}{\partial x^i} = V^i \frac{\partial}{\partial x^j} \cdot \frac{\partial}{\partial x^i}.$$

Note that the tangent vectors $\partial/\partial x^i$ are not necessarily orthogonal, so the dot products don't behave as simply as we might suspect. We let $g_{ij} = \frac{\partial}{\partial x^j} \cdot \frac{\partial}{\partial x^i}$, so that

$$C_j = g_{ij} V^i.$$

We shall see the quantities g_{ij} again presently.

Definition 4.9 If V and W are contravariant (or covariant) vector fields on M, and if α is a real number, we can define new fields $V+W$ and αV by

$$(V + W)^i = V^i + W^i$$

and $$(\alpha V)^i = \alpha V^i.$$

It is easily verified that the resulting quantities are again contravariant (or covariant) fields. (Exercise Set 4). For contravariant fields, these operations coincide with addition and scalar multiplication as we defined them before.

These operations turn the set of all smooth contravariant (or covariant) fields on M into a vector space. Note that we cannot expect to obtain a vector field by adding a covariant field to a contravariant field.

Exercise Set 4

1. Suppose that X^j is a contravariant vector field on the manifold M with the following property: at every point m of M, there exists a local coordinate system x^i at m with $X^j(x^1, x^2, \ldots, x^n) = 0$. Show that X^i is identically zero in any coordinate system.

2. Give and example of a contravariant vector field that is not covariant. Justify your claim.

3. Verify the following claim If V and W are contravariant (or covariant) vector fields on M, and if α is a real number, then $V+W$ and αV are again contravariant (or covariant) vector fields on M.

4. Verify the following claim in the proof of Proposition 4.7: If C_i is covariant and V^j is contravariant, then $C_k V^k$ is a scalar.

5. Let ϕ: $S^n \longrightarrow E_1$ be the scalar field defined by $\phi(p_1, p_2, \ldots, p_{n+1}) = p_{n+1}$.

(a) Express ϕ as a function of the x^i and as a function of the $\bar{x}^j$.

(b) Calculate $C_i = \partial\phi/\partial x^i$ and $\bar{C}_j = \partial\phi/\partial\bar{x}^j$.

(c) Verify that C_i and $\bar{C}_j$ transform according to the covariant vector transformation rules.

6. Is it true that the quantities x^i themselves form a contravariant vector field? Prove or give a counterexample.

7. Prove that Ψ and Φ in Proposition 4.7 are inverse functions.

8. Prove: Every covariant vector field is of the type given in Example 4.8(d). That is, obtained from the dot product with some contrravariant field.

---

5. Tensor Fields

Suppose that $\boldsymbol{v} = \langle v_1, v_2, v_3\rangle$ and $\boldsymbol{w} = \langle w_1, w_2, w_3\rangle$ are vector fields on E_3. Then their **tensor product** is defined to consist of the nine quantities v_iw_j. Let us see how such things transform. Thus, let V and W be contravariant, and let C and D be covariant. Then:

$$\overline{V}^i\overline{W}^j = \frac{\partial\overline{x}^i}{\partial x^k}V^k\,\frac{\partial\overline{x}^j}{\partial x^l}W^l = \frac{\partial\overline{x}^i}{\partial x^k}\frac{\partial\overline{x}^j}{\partial x^l}V^kW^l,$$

and similarly,

$$\overline{V}^i\overline{C}_j = \frac{\partial\overline{x}^i}{\partial x^k}\frac{\partial x^l}{\partial\overline{x}^j}V^kC_l\,,$$

and

$$\overline{C}_i\overline{D}_j = \frac{\partial x^k}{\partial\overline{x}^i}\frac{\partial x^l}{\partial\overline{x}^j}C_kD_l\,.$$

We call these fields "tensors" of type (2, 0), (1, 1), and (0, 2) respectively.

Definition 5.1 A **tensor field of type (2, 0)** on the n-dimensional smooth manifold M associates with each chart x a collection of n^2 smooth functions $T^{ij}(x^1, x^2, \ldots, x^n)$ which satisfy the transformation rules shown below. Similarly, we define tensor fields of type (0, 2), (1, 1), and, more generally, a tensor field of type (m, n).

Some Tensor Transformation Rules

Type (2, 0): $$\overline{T}^{ij} = \frac{\partial\overline{x}^i}{\partial x^k}\frac{\partial\overline{x}^j}{\partial x^l}T^{kl}$$

Type (1, 1): $$\overline{M}^i_j = \frac{\partial\overline{x}^i}{\partial x^k}\frac{\partial x^l}{\partial\overline{x}^j}M^k_l$$

Type (0, 2): $$\overline{S}_{ij} = \frac{\partial x^k}{\partial\overline{x}^i}\frac{\partial x^l}{\partial\overline{x}^j}S_{kl}$$

Notes

(1) A tensor field of type (1, 0) is just a contravariant vector field, while a tensor field of type (0, 1) is a covariant vector field. Similarly, a tensor field of type (0, 0) is a scalar field. Type (1, 1) tensors correspond to linear transformations in linear algebra.

(2) We add and scalar multiply tensor fields in a manner similar to the way we do these things to vector fields. For instant, if A and B are type (1,2) tensors, then their sum is given by

$$(A+B)_{ab}{}^c = A_{ab}{}^c + B_{ab}{}^c.$$

Examples 5.2

(a) Of course, by definition, we can take **tensor products** of vector fields to obtain tensor fields, as we did above in Definition 4.1.

(b) The Kronecker Delta Tensor, given by

$$\delta^i_j = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases}$$

is, in fact a tensor field of type (1, 1). Indeed, one has

$$\delta^i_j = \frac{\partial x^i}{\partial x^j},$$

and the latter quantities transform according to the rule

$$\bar{\delta}^i_j = \frac{\partial \bar{x}^i}{\partial \bar{x}^j} = \frac{\partial \bar{x}^i}{\partial x^k}\frac{\partial x^k}{\partial x^l}\frac{\partial x^l}{\partial \bar{x}^j} = \frac{\partial \bar{x}^i}{\partial x^k}\frac{\partial x^l}{\partial \bar{x}^j}\delta^k_l,$$

whence they constitute a tensor field of type (1, 1).

Notes

1. $\delta^i_j = \bar{\delta}^i_j$ as functions on E_n. Also, $\delta^i_j = \delta^j_i$. That is, it is a **symmetric tensor**.

2. $\dfrac{\partial \bar{x}^i}{\partial x^j}\dfrac{\partial x^j}{\partial \bar{x}^k} = \dfrac{\partial \bar{x}^i}{\partial \bar{x}^k} = \delta^i_k$.

Question OK, so is this how it works: Given a point p of the manifold and a chart x at p this strange object assigns the n^2 quantities δ^i_j; that is, *the identity matrix, regardless of the chart we chose*?

Answer Yes.

Question But how can we interpret this strange object?

Answer Just as a covariant vector field converts contravariant fields into scalars (see Section 3) we shall see that a type (1,1) tensor converts contravariant fields to other contravariant fields. This particular tensor does nothing: put in a specific vector field V, out comes the same vector field. In other words, it is the *identity transformation.*

(c) We can make new tensor fields out of old ones by taking **products** of existing tensor fields in various ways. For example,

$M^i_{jk} N^{pq}_{rs}$ is a tensor of type (3, 4),

while

$M^i_{jk} N^{jk}_{rs}$ is a tensor of type (1, 2).

Specific examples of these involve the Kronecker delta, and are in the homework.

(d) If X is a contravariant vector field, then the functions $\dfrac{\partial X^i}{\partial x^j}$ do *not* define a tensor.

Indeed, let us check the transformation rule directly:

$$\begin{aligned}
\frac{\partial \bar{X}^i}{\partial \bar{x}^j} &= \frac{\partial}{\partial \bar{x}^j}\left(X^k \frac{\partial \bar{x}^i}{\partial x^k}\right) \\
&= \frac{\partial}{\partial x^h}\left(X^k \frac{\partial \bar{x}^i}{\partial x^k}\right)\frac{\partial x^h}{\partial \bar{x}^j} \\
&= \frac{\partial X^k}{\partial x^h}\frac{\partial \bar{x}^i}{\partial x^k}\frac{\partial x^h}{\partial \bar{x}^j} + X^k \frac{\partial^2 \bar{x}^i}{\partial x^h \partial x^K}
\end{aligned}$$

The extra term on the right violates the transformation rules.

We will see more interesting examples later.

Proposition 5.3 ***(If It Looks Like a Tensor, It Is a Tensor)***
Suppose that we are given smooth local functions g_{ii} with the property that for every pair of contravariant vector fields X^i and Y^i, the smooth functions $g_{ij}X^iY^j$ determine a scalar field, then the g_{ij} determine a smooth tensor field of type (0, 2).

Proof Since the $g_{ij}X^iY^j$ form a scalar field, we must have

$$\bar{g}_{ij}\bar{X}^i\bar{Y}^j = g_{hk}X^hY^k.$$

On the other hand,

$$\bar{g}_{ij}\bar{X}^i\bar{Y}^j = \bar{g}_{ij}X^hY^k\frac{\partial\bar{x}^i}{\partial x^h}\frac{\partial\bar{x}^j}{\partial x^k}.$$

Equating the right-hand sides gives

$$g_{hk}X^hY^k = \bar{g}_{ij}\frac{\partial\bar{x}^i}{\partial x^h}\frac{\partial\bar{x}^j}{\partial x^k}X^hY^k \quad \text{-----------------------} \quad \text{(I)}$$

Now, if we could only cancel the terms X^hY^k. Well, choose a point $m \in M$. It suffices to show that $g_{hk} = \bar{g}_{ij}\frac{\partial\bar{x}^i}{\partial x^h}\frac{\partial\bar{x}^j}{\partial x^k}$, when evaluated at the coordinates of m. By Example 4.3(c), we can arrange for vector fields X and Y such that

$$X^i(\text{coordinates of } m) = \begin{cases} 1 & \text{if } i = h \\ 0 & \text{otherwise} \end{cases},$$

and

$$Y^i(\text{coordinates of } m) = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{otherwise} \end{cases}.$$

Substituting these into equation (I) now gives the required transformation rule. ◆

Example 5.4 Metric Tensor
Define a set of quantities g_{ij} by

$$g_{ij} = \frac{\partial}{\partial x^j}\cdot\frac{\partial}{\partial x^i}.$$

If X^i and Y^j are any contravariant fields on M, then $\boldsymbol{X}\cdot\boldsymbol{Y}$ is a scalar, and

$$\boldsymbol{X}\cdot\boldsymbol{Y} = X^i\frac{\partial}{\partial x^i}\cdot Y^j\frac{\partial}{\partial x^j} = g_{ij}X^iY^j.$$

Thus, by proposition 4.3, it is a type (0, 2) tensor. We call this tensor "the metric tensor inherited from the imbedding of M in E_s."

Exercise Set 5
1. Compute the transformation rules for each of the following, and hence decide whether or not they are tensors. Sub-and superscripted quantities (other than coordinates) are understood to be tensors.

(a) $\dfrac{dX^i_j}{dt}$ **(b)** $\dfrac{\partial x^i}{\partial x^j}$ **(c)** $\dfrac{\partial X^i}{\partial x^j}$ **(d)** $\dfrac{\partial^2 \phi}{\partial x^i \partial x^j}$ **(e)** $\dfrac{\partial^2 x^l}{\partial x^i \partial x^j}$

2. (Rund, p. 95 #3.4) Show that if A_j is a type (0, 1) tensor, then

$$\frac{\partial A_j}{\partial x^h} - \frac{\partial A_h}{\partial x^j}$$

is a type (0, 2) tensor.

3. Show that, if M and N are tensors of type (1, 1), then:

(a) $M^i_j N^p_q$ is a tensor of type (2, 2)

(b) $M^i_j N^j_q$ is a tensor of type (1, 1)

(c) $M^i_j N^j_i$ is a tensor of type (0, 0) (that is, a scalar field)

4. Let X be a contravariant vector field, and suppose that M is such that all change-of-coordinate maps have the form $\bar{x}^i = a^{ij}x^j + k^i$ for certain constants a^{ij} and k^j. (We call such a manifold *affine*.) Show that the functions $\dfrac{\partial X^i}{\partial x^j}$ define a tensor field of type (1, 1).

5. (Rund, p. 96, 3.12) If $B^{ijk} = -B^{jki}$, show that $B^{ijk} = 0$. Deduce that any type (3, 0) tensor that is symmetric on the first pair of indices and skew-symmetric on the last pair of indices vanishes.

6. (Rund, p. 96, 3.16) If A_{kj} is a skew-symmetric tensor of type (0, 2), show that the quantities B_{rst} defined by

$$B_{rst} = \frac{\partial A_{st}}{\partial x^r} + \frac{\partial A_{tr}}{\partial x^s} + \frac{\partial A_{rs}}{\partial x^t}$$

(a) are the components of a tensor; and

(b) are skew-symmetric in all pairs in indices.

(c) How many independent components does B_{rst} have?

7. Cross Product

(a) If X and Y are contravariant vectors, then their **cross-product** is defined as the tensor of type (2, 0) given by

$$(X \wedge Y)^{ij} = X^iY^j - X^jY^i.$$

Show that it is a skew-symmetric tensor of type (2, 0).

(b) If $M = E_3$, then the **totally antisymmetric third order tensor** is defined by

$$\varepsilon_{ijk} = \begin{cases} 1 & \text{if } (i, j, k) \text{ is an even permutation of } (1, 2, 3) \\ -1 & \text{if it is an odd permutation of } (1, 2, 3) \end{cases}$$

(or equivalently, $\varepsilon_{123} = +1$, and ε_{ijk} is skew-symmetric in every pair of indices.) Then, the (usual) cross product on E_3 is defined by

$$(X \times Y)_i = \varepsilon_{ijk}(X \wedge Y)^{jk}.$$

(c) What goes wrong when you try to define the "usual" cross product of two vectors on E_4? Is there any analogue of (b) for E_4?

8. Suppose that C^{ij} is a type (2, 0) tensor, and that, regarded as an $n \times n$ matrix C, it happens to be invertible in every coordinate system. Define a new collection of functions, D_{ij} by taking

$$D_{ij} = {C^{-1}}_{ij},$$

the ij the entry of C^{-1} in every coordinate system. Show that D_{ij}, is a type (0, 2) tensor. [Hint: Write down the transformation equation for C^{ij} and invert everything in sight.]

9. What is wrong with the following "proof" that $\dfrac{\partial^2 \bar{x}^j}{\partial x^h \partial x^k} = 0$ regardless of what smooth functions $\bar{x}^j(x^h)$ we use:

$$\frac{\partial^2 \bar{x}^j}{\partial x^h\, \partial x^k} = \frac{\partial}{\partial x^h}\left(\frac{\partial \bar{x}^j}{\partial x^k}\right) \qquad \text{Definition of the second derivative}$$

$$= \frac{\partial}{\partial \bar{x}^l}\left(\frac{\partial \bar{x}^j}{\partial x^k}\right)\frac{\partial \bar{x}^l}{\partial x^h} \qquad \text{Chain rule}$$

$$= \frac{\partial^2 \bar{x}^j}{\partial \bar{x}^l\, \partial x^k}\frac{\partial \bar{x}^l}{\partial x^h} \qquad \text{Definition of the second derivative}$$

$$= \frac{\partial^2 \bar{x}^j}{\partial x^k\, \partial \bar{x}^l}\frac{\partial \bar{x}^l}{\partial x^h} \qquad \text{Changing the order of differentiation}$$

$$= \frac{\partial}{\partial x^k}\left(\frac{\partial \bar{x}^j}{\partial \bar{x}^l}\right)\frac{\partial \bar{x}^l}{\partial x^h} \qquad \text{Definition of the second derivative}$$

$$= \frac{\partial}{\partial x^k}\left(\delta^i_l\right)\frac{\partial \bar{x}^l}{\partial x^h} \qquad \text{Since } \frac{\partial \bar{x}^i}{\partial \bar{x}^l} = \delta^i_l$$

$$= 0 \qquad \text{Since } \delta^i_l \text{ is constant!}$$

---

6. Riemannian Manifolds

Definition 6.1 A **smooth inner product** on a manifold M is a function $\langle -,- \rangle$ that associates to each pair of smooth contravariant vector fields X and Y a scalar (field) $\langle X, Y\rangle$, satisfying the following properties.

Symmetry: $\langle X, Y\rangle = \langle Y, X\rangle$ for all X and Y,

Bilinearity: $\langle \alpha X, \beta Y\rangle = \alpha\beta\langle X, Y\rangle$ for all X and Y, and scalars α and β

$\langle X, Y+Z\rangle = \langle X, Y\rangle + \langle X, Z\rangle$

$\langle X+Y, Z\rangle = \langle X, Z\rangle + \langle Y, Z\rangle$.

Non-degeneracy: If $\langle X, Y\rangle = 0$ for every Y, then $X = 0$.

We also call such a gizmo a **symmetric bilinear form**. A manifold endowed with a smooth inner product is called a **Riemannian manifold.**

Before we look at some examples, let us see how these things can be specified. First, notice that, if $\boldsymbol{x}$ is any chart, and p is any point in the domain of $\boldsymbol{x}$, then

$$\langle X, Y\rangle = X^i Y^j \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \rangle.$$

This gives us smooth functions

$$g_{ij} = \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \rangle$$

such that

$$\langle X, Y\rangle = g_{ij} X^i Y^j$$

and which, by Proposition 5.3, constitute the coefficients of a type (0, 2) symmetric tensor. We call this tensor the **fundamental tensor** or **metric tensor** of the Riemannian manifold.

Examples 6.2

(a) $M = E_n$, with the usual inner product; $g_{ij} = \delta_{ij}$.

(b) (Minkowski Metric) $M = E_4$, with g_{ij} given by the matrix

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -c^2 \end{bmatrix},$$

where c is the speed of light.

Question How does this effect the length of vectors?

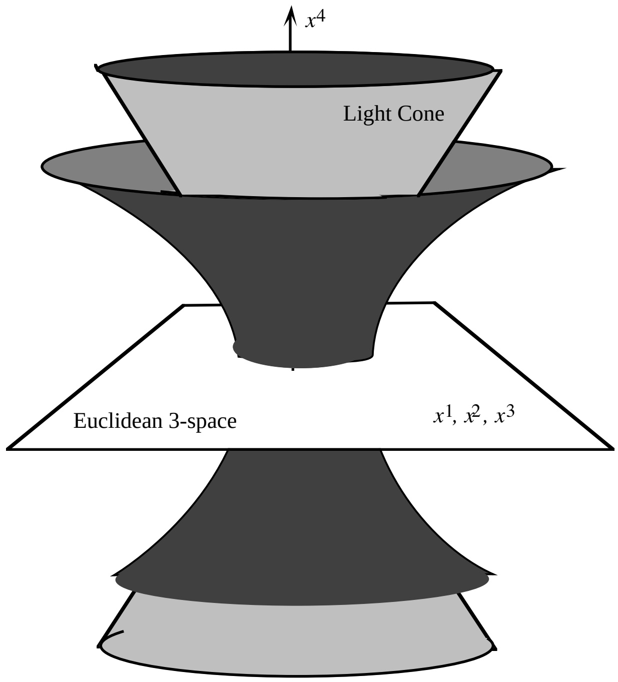
Answer We saw in Section 3 that, in E_n, we could think of tangent vectors in the usual way; as directed line segments starting at the origin. The role that the metric plays is that it tells you the length of a vector; in other words, it gives you a new distance formula:

Euclidean 3- space: $d(x, y) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2}$

Minkowski 4-space: $d(x, y) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2 - c^2(y_4 - x_4)^2}$

.

Geometrically, the set of all points in Euclidean 3-space at a distance r from the origin (or any other point) is a sphere of radius r. In Minkowski space, it is a hyperbolic surface. In Euclidean space, the set of all points a distance of 0 from the origin is just a single point; in M, it is a cone, called the **light cone**. (See the figure.)



(c) If M is any manifold embedded in E_s, then we have seen above that M inherits the structure of a Riemannian metric from a given inner product on E_s. In particular, if M is any 3-dimensional manifold embedded in E_4 with the metric shown above, then M inherits such a inner product.

(d) As a particular example of (c), let us calculate the metric of the two-sphere $M = S^2$, with radius r, using polar coordinates $x^1 = \theta$, $x^2 = \phi$. To find the coordinates of g_{**} we need to calculate the inner product of the basis vectors $\partial/\partial x^1$, $\partial/\partial x^2$. We saw in Section 3 that the ambient coordinates of $\partial/\partial x^i$ are given by

$$j \text{ th coordinate} = \frac{\partial y_j}{\partial x^i},$$

where

$$y_1 = r\ \sin(x^1)\cos(x^2)$$
$$y_2 = r\ \sin(x^1)\sin(x^2)$$
$$y_3 = r\ \cos(x^1)$$

Thus,

$$\frac{\partial}{\partial x^1} = r(\cos(x^1)\cos(x^2),\ \cos(x^1)\sin(x^2),\ -\sin(x^1))$$
$$\frac{\partial}{\partial x^2} = r(-\sin(x^1)\sin(x^2),\ \sin(x^1)\cos(x^2),\ 0)$$

This gives

$$g_{11} = \langle \partial/\partial x^1,\ \partial/\partial x^1\rangle = r^2$$
$$g_{22} = \langle \partial/\partial x^2,\ \partial/\partial x^2\rangle = r^2 \sin^2(x^1)$$
$$g_{12} = \langle \partial/\partial x^1, \partial/\partial x^2 \rangle = 0,$$

so that

$$g_{**} = \begin{bmatrix} r^2 & 0 \\ 0 & r^2\sin^2(x^1) \end{bmatrix}.$$

(e) The *n*-Dimensional Sphere Let M be the n-sphere of radius r with the followihg generalized polar coordinates.

$$y_1 = r\cos x^1$$
$$y_2 = r\sin x^1 \cos x^2$$
$$y_3 = r\sin x^1 \sin x^2 \cos x^3$$
$$\dots$$
$$y_{n-1} = r\sin x^1 \sin x^2 \sin x^3 \sin x^4 \dots \cos x^{n-1}$$
$$y_n = r\sin x^1 \sin x^2 \sin x^3 \sin x^4 \dots \sin x^{n-1} \cos x^n$$
$$y_{n+1} = r\sin x^1 \sin x^2 \sin x^3 \sin x^4 \dots \sin x^{n-1} \sin x^n.$$

(Notice that x^1 is playing the role of ϕ and the $x^2, x^3, \dots, x^{n-1}$ the role of θ.) Following the line of reasoning in the previous example, we have

$$\frac{\partial}{\partial x^1} = (-r\sin x^1, r\cos x^1 \cos x^2, r\cos x^1 \sin x^2 \cos x^3, \dots,$$
$$r\cos x^1 \sin x^2 \dots \sin x^{n-1} \cos x^n, r\cos x^1 \sin x^2 \dots \sin x^{n-1} \sin x^n)$$

$$\frac{\partial}{\partial x^2} = (0, -r\sin x^1 \sin x^2, \dots, r\sin x^1 \cos x^2 \sin x^3 \dots \sin x^{n-1} \cos x^n,$$
$$r\sin x^1 \cos x^2 \sin x^3 \dots \sin x^{n-1} \sin x^n).$$

$$\frac{\partial}{\partial x^3} = (0, 0, -r\sin x^1 \sin x^2 \sin x^3, r\sin x^1 \sin x^2 \cos x^3 \cos x^4 \dots,$$
$$r\sin x^1 \sin x^2 \cos x^3 \sin x^4 \dots \sin x^{n-1} \cos x^n, r\sin x^1 \sin x^2 \cos x^3 \sin x^4 \dots \sin x^{n-1} \sin x^n),$$

and so on.

$$g_{11} = \langle \partial/\partial x^1,\ \partial/\partial x^1\rangle = r^2$$
$$g_{22} = \langle \partial/\partial x^2,\ \partial/\partial x^2\rangle = r^2\sin^2 x^1$$
$$g_{33} = \langle \partial/\partial x^3,\ \partial/\partial x^3\rangle = r^2\sin^2 x^1 \sin^2 x^2$$

$$\ldots$$

$$g_{nn} = \langle \partial/\partial x^n,\, \partial/\partial x^n \rangle = r^2 \sin^2 x^1 \sin^2 x^2 \ldots \sin^2 x^{n-1}$$

$$g_{ij} = 0 \text{ if } i \neq j$$

so that

$$g_{**} = \begin{bmatrix} r^2 & 0 & 0 & \ldots & 0 \\ 0 & r^2\sin^2 x^1 & 0 & \ldots & 0 \\ 0 & 0 & r^2\sin^2 x^1 \sin^2 x^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & r^2\sin^2 x^1 \sin^2 x^2 \ldots \sin^2 x^{n-1} \end{bmatrix}.$$

(f) Diagonalizing the Metric Let G be the matrix of g_{**} in some local coordinate system, evaluated at some point p on a Riemannian manifold. Since G is symmetric, it follows from linear algebra that there is an invertible matrix $P = (P_{ji})$ such that

$$PGP^T = \begin{bmatrix} \pm 1 & 0 & 0 & 0 \\ 0 & \pm 1 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \pm 1 \end{bmatrix}$$

at the point p. Let us call the sequence $(\pm 1, \pm 1, \ldots, \pm 1)$ the **signature** of the metric at p. (Thus, in particular, a Minkowski metric has signature $(1, 1, 1, -1)$.) If we now define new coordinates $\bar{x}^j$ by

$$x^i = P_{ji} \bar{x}^j,$$

(so that we are using the inverse of P for this) then $\partial x^i / \partial \bar{x}^j = P_{ji}$, and so

$$\bar{g}_{ij} = \frac{\partial x^a}{\partial \bar{x}^i} g_{ab} \frac{\partial x^b}{\partial \bar{x}^j} = P_{ia} g_{ab} P_{jb}$$
$$= P_{ia} g_{ab} (P^T)_{bj} = (PGP^T)_{ij}$$

showing that, at the point p,

$$\bar{g}_{**} = \begin{bmatrix} \pm 1 & 0 & 0 & 0 \\ 0 & \pm 1 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \pm 1 \end{bmatrix}.$$

Thus, in the eyes of the metric, the unit basis vectors $e_i = \partial/\partial \bar{x}^i$ are **orthogonal**; that is,

$$\langle e_i, e_j \rangle = \pm\delta_{ij}.$$

Note The non-degeneracy condition in Definition 6.1 is equivalent to the requirement that the locally defined quantities

$$g = \det(g_{ij})$$

are nowhere zero.

Here are some things we can do with a Riemannian manifold.

Definition 6.3 If X is a contravariant vector field on M, then define the **square norm norm of *X*** by

$$\|X\|^2 = \langle X, X \rangle = g_{ij}X^iX^j.$$

Note that $\|X\|^2$ may be negative. If $\|X\|^2 < 0$, we call X **timelike**; if $\|X\|^2 > 0$, we call X **spacelike**, and if $\|X\|^2 = 0$, we call X **null.** If X is not spacelike, then we can define

$$\|X\| = \sqrt{\|X\|^2} = \sqrt{g_{ij}X^iX^j}\ .$$

In the exercise set you will show that null need not imply zero.

Note Since $\langle X, X \rangle$ is a scalar field, so is $\|X\|$ is a scalar field, if it exists, and satisfies $\|\phi X\| = |\phi|\cdot\|X\|$ for every contravariant vector field X and every scalar field ϕ. The expected inequality

$$\|X + Y\| \leq \|X\| + \|Y\|$$

need *not* hold. (See the exercises.)

Arc Length One of the things we can do with a metric is the following. A path C given by $x^i = x^i(t)$ is **non-null** if $\|dx^i/dt\|^2 \neq 0$. It follows that $\|dx^i/dt\|^2$ is either always positive (**"spacelike"**) or negative (**"timelike"**).

Definition 6.4 If C is a non-null path in M, then define its **length** as follows: Break the path into segments S each of which lie in some coordinate neighborhood, and define the length of S by

$$L(a, b) = \int_a^b \sqrt{\pm\, g_{ij}\frac{dx^i}{dt}\frac{dx^j}{dt}}\ dt,$$

where the sign ±1 is chosen as +1 if the curve is space-like and −1 if it is time-like. In other words, we are defining the arc-length differential form by

$$ds^2 = \pm g_{ij} dx^i dx^j.$$

To show (as we must) that this is independent of the choice of chart x, all we need observe is that the quantity under the square root sign, being a contraction product of a type (0, 2) tensor with a type (2, 0) tensor, is a scalar.

Proposition 6.5 (Paramaterization by Arc Length)
Let C be a non-null path $x^i = x^i(t)$ in M. Fix a point $t = a$ on this path, and define a new function s (arc length) by

$$s(t) = L(a, t) = \text{length of path from } t = a \text{ to } t.$$

Then s is an invertible function of t, and, using s as a parameter, $\|dx^i/ds\|^2$ is constant, and equals 1 if C is space-like and −1 if it is time-like.

Conversely, if t is any parameter with the property that $\|dx^i/dt\|^2 = \pm 1$, then, choosing any parameter value $t = a$ in the above definition of arc-length s, we have

$$t = \pm s + C$$

for some constant C. (In other words, t must be, up to a constant, arc length. Physicists call the parameter $\tau = s/c$, where c is the speed of light, **proper time** for reasons we shall see below.)

Proof Inverting $s(t)$ requires $s'(t) \neq 0$. But, by the Fundamental theorem of Calculus and the definition of $L(a, t)$,

$$\left(\frac{ds}{dt}\right)^2 = \pm\, g_{ij} \frac{dx^i}{dt}\frac{dx^j}{dt} \neq 0$$

for all parameter values t. In other words,

$$\langle \frac{dx^i}{dt}, \frac{dx^i}{dt} \rangle \neq 0.$$

But this is the never null condition which we have assumed. Also,

$$\langle \frac{dx^i}{ds}, \frac{dx^i}{ds} \rangle = g_{ij}\frac{dx^i dx^j}{ds\, ds} = g_{ij}\frac{dx^i dx^j}{dt\, dt}\left(\frac{dt}{ds}\right)^2 = \pm\left(\frac{ds}{dt}\right)^2\left(\frac{dt}{ds}\right)^2 = \pm 1$$

For the converse, we are given a parameter t such that

$$\langle \frac{dx^i}{dt}, \frac{dx^i}{dt} \rangle = \pm 1.$$

in other words,

$$g_{ij}\frac{dx^i dx^j}{dt\ dt} = \pm 1.$$

But now, with s defined to be arc-length from $t = a$, we have

$$\left(\frac{ds}{dt}\right)^2 = \pm\, g_{ij}\frac{dx^i dx^j}{dt\ dt} = +1$$

(the signs cancel for time-like curves) so that

$$\left(\frac{ds}{dt}\right)^2 = 1,$$

meaning of course that $t = \pm s + C$. ❇

Exercise Set 6

1. Give an example of a Riemannian metric on E_2 such that the corresponding metric tensor g_{ij} is not constant.

2. Let a_{ij} be the components of any symmetric tensor of type (0, 2) such that $\det(a_{ij})$ is never zero. Define

$$\langle X, Y\rangle_a = a_{ij}X^iY^j.$$

Show that this is a smooth inner product on M.

3. Give an example to show that the "triangle inequality" $\|X+Y\| \le \|X\| + \|Y\|$ is not always true on a Riemannian manifold.

4. Give an example of a Riemannian manifold M and a nowhere zero vector field X on M with the property that $\|X\| = 0$. We call such a field **a null field.**

5. Show that if g is any smooth type (0, 2) tensor field, and if $g = \det(g_{ij}) \neq 0$ for some chart x, then $\bar{g} = \det(\bar{g}_{ij}) \neq 0$ for every other chart $\bar{x}$ (at points where the change-of-coordinates is defined). [Use the property that, if A and B are matrices, then $\det(AB) = \det(A)\det(B)$.]

6. Suppose that g_{ij} is a type (0, 2) tensor with the property that $g = \det(g_{ij})$ is nowhere zero. Show that the resulting inverse (of matrices) g^{ij} is a type (2, 0) tensor. (Note that it must satisfy $g_{ij}g^{kl} = \delta^k_i\ \delta^l_j$.)

7. (Index lowering and raising) Show that, if R_{abc} is a type (0, 3) tensor, then $R_a{}^i{}_c$ given by

$$R_a{}^i{}_c = g^{ib}R_{abc},$$

is a type (1, 2) tensor. (Here, g^{**} is the inverse of g_{**}.) What is the inverse operation?

---

8. A type (1, 1) tensor field T is **orthogonal** in the Riemannian manifold M if, for all pairs of contravariant vector fields X and Y on M, one has

$$\langle TX, TY\rangle = \langle X, Y\rangle,$$

where $(TX)^i = T^i_k X^k$. What can be said about the columns of T in a given coordinate system x? (Note that the i^{th} column of T is the local vector field given by $T(\partial/\partial x^i)$.)

---

7. Locally Minkowskian Manifolds: An Introduction to Relativity

First a general comment: We said in the last section that, at any point p in a Riemannian manifold M, we can find a local chart at p with the property that the metric tensor g_{**} is diagonal, with diagonal terms ±1. In particular, we said that Minkowski space comes with a such a metric tensor having signature (1, 1, 1, −1). Now there is nothing special about the number 1 in the discussion: we can also find a local chart at any point p with the property that the metric tensor g_{**} is diagonal, with diagonal terms any non-zero numbers we like (although we cannot choose the signs).

In relativity, we take deal with 4-dimensional manifolds, and take the first three coordinates x^1, x^2, x^3 to be spatial (measuring distance), and the fourth one, x^4, to be temporal (measuring time). Let us postulate that we are living in some kind of 4-dimensional manifold M (since we want to include time as a coordinate. By the way, we refer to a chart x at the point p as a **frame of reference**, or just **frame**). Suppose now we have a particle—perhaps moving, perhaps not—in M. Assuming it persists for a period of time, we can give it spatial coordinates (x^1, x^2, x^3) at every instant of time (x^4). Since the first three coordinates are then functions of the fourth, it follows that the particle determines a path in M given by

$$\begin{aligned} x^1 &= x^1(x^4) \\ x^2 &= x^2(x^4) \\ x^3 &= x^3(x^4) \\ x^4 &= x^4, \end{aligned}$$

so that x^4 is the parameter. This path is called the **world line** of the particle. Mathematically, there is no need to use x^4 as the parameter, and so we can describe the world line as a path of the form

$$x^i = x^i(t),$$

where t is some parameter. (Note: t is *not* time; it's just a parameter. x^4 is time). Conversely, if t is any parameter, and $x^i = x^i(t)$ is a path in M, then, if x^4 is an invertible function of t, that is, $dx^4/dt \neq 0$ (so that, at each time x^4, we can solve for the other coordinates uniquely) then we can solve for x^1, x^2, x^3 as smooth functions of x^4, and hence picture the situation as a *particle moving through space.*

Now, let's assume our particle is moving through M with world line $x^i = x^i(t)$ as seen in our frame (local coordinate system). The velocity and speed of this particle (as measured in our frame) are given by

$$\boldsymbol{v} = \left(\frac{dx^1}{dx^4}, \frac{dx^2}{dx^4}, \frac{dx^3}{dx^4}\right)$$

$$\text{Speed}^2 = \left(\frac{dx^1}{dx^4}\right)^2 + \left(\frac{dx^2}{dx^4}\right)^2 + \left(\frac{dx^3}{dx^4}\right)^2 .$$

The problem is, we cannot expect $\boldsymbol{v}$ to be a vector—that is, satisfy the correct transformation laws. But we *do* have a contravariant 4-vector

$$T^i = \frac{dx^i}{dt}$$

(T stands for tangent vector. Also, remember that t is not time). If the particle is moving at the speed of light c, then

$$\left(\frac{dx^1}{dx^4}\right)^2 + \left(\frac{dx^2}{dx^4}\right)^2 + \left(\frac{dx^3}{dx^4}\right)^2 = c^2 \qquad \ldots\ldots\ldots \qquad \text{(I)}$$

$$\Leftrightarrow \quad \left(\frac{dx^1}{dt}\right)^2 + \left(\frac{dx^2}{dt}\right)^2 + \left(\frac{dx^3}{dt}\right)^2 = c^2\left(\frac{dx^4}{dt}\right)^2 \qquad \text{(using the chain rule)}$$

$$\Leftrightarrow \quad \left(\frac{dx^1}{dt}\right)^2 + \left(\frac{dx^2}{dt}\right)^2 + \left(\frac{dx^3}{dt}\right)^2 - c^2\left(\frac{dx^4}{dt}\right)^2 = 0.$$

Now this looks like the norm-squared $||\boldsymbol{T}||^2$ of the vector T under the metric whose matrix is

$$g_{**} = \text{diag}[1,\ 1,\ 1,\ -c^2] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -c^2 \end{bmatrix}$$

In other words, the particle is moving at light-speed $\Leftrightarrow\ ||\boldsymbol{T}||^2 = 0$
$\Leftrightarrow\ ||\boldsymbol{T}||$ is null

under this rather interesting local metric. So, to check whether a particle is moving at light speed, just check whether $\boldsymbol{T}$ is null.

Question What's the $-c^2$ doing in place of -1 in the metric?
Answer Since physical units of time are (usually) not the same as physical units of space, we would like to convert the units of x_4 (the units of time) to match the units of the other axes. Now, to convert units of time to units of distance, we need to multiply by something with units of distance/time; that is, by a non-zero *speed*. Since relativity holds that the speed of light c is a universal constant, it seems logical to use c as this conversion factor.

Now, *if* we happen to be living in a Riemannian 4-manifold whose metric diagonalizes to something with signature $(1,\ 1,\ 1,\ -c^2)$, then the physical property of traveling at the speed of light is measured by $\|T\|^2$, which is a *scalar*, and thus *independent of the frame of reference*. In other words, we have discovered a metric signature that is consistent with the requirement that *the speed of light is constant in all frames*(in which g_{**} has the above diagoal form, so that ita makes sense to say what the speed if light *is*).

Definition 7.1 A Riemannian 4-manifold M is called **locally Minkowskian** if its metric has signature $(1, 1, 1, -c^2)$.

For the rest of this section, we will be in a locally Minkowskian manifold M.

Note If we now choose a chart x in locally Minkowskian space where the metric has the diagonal form diag$[1, 1, 1, -c^2]$ shown above at a given point p, then we have, at the point p:

(a) If any path C has $\|T\|^2 = 0$, then

$$\left(\frac{dx^1}{dt}\right)^2 + \left(\frac{dx^2}{dt}\right)^2 + \left(\frac{dx^3}{dt}\right)^2 - c^2\left(\frac{dx^4}{dt}\right)^2 = 0 \quad \text{(because this is how we calculate } \|T\|^2\text{)}$$

(b) If V is any contravariant vector with zero x^4-coordinate, then

$$\|V\|^2 = (V^1)^2 + (V^2)^2 + (V^3)^2 \qquad \text{(for the same reason as above)}$$

(a) says that we measure the world line C as representing a particle traveling with light speed, and (b) says that we measure ordinary length in the usual way. This motivates the following definition.

Definition 7.2 A **Lorentz frame at the point $p \in M$** is any coordinate system $\bar{x}^i$ with the following properties:

(a) If any path C has the scalar $\|T\|^2 = 0$, then, at p,

$$\left(\frac{d\bar{x}^1}{dt}\right)^2 + \left(\frac{d\bar{x}^2}{dt}\right)^2 + \left(\frac{d\bar{x}^3}{dt}\right)^2 - c^2\left(\frac{d\bar{x}^4}{dt}\right)^2 = 0 \;\ldots\ldots \qquad \text{(II)}$$

(Note: In general, $\langle \bar{T}, \bar{T}\rangle$ is not of this form, since $\bar{g}_{ij}$ may not be be diagonal)

(b) If $\boldsymbol{V}$ is a contravariant vector at p with zero $\bar{x}^4$-coordinate, then

$$\|\boldsymbol{V}\|^2 = (\bar{V}^1)^2 + (\bar{V}^2)^2 + (\bar{V}^3)^2 \qquad \ldots\ldots \quad \text{(III)}$$

(Again, this need not be $\|\bar{\boldsymbol{V}}\|^2$.)

It follows from the remark preceding the defintion that if x is any chart such that, at the point p, the metric has the nice form diag$[1, 1, 1, -c^2]$, then x is a Lorentz frame at the point p. Note that in general, the coordinates of $\boldsymbol{T}$ in the system $\bar{x}^i$ are given by matrix multiplication with some possibly complicated change-of-coordinates matrix, and to further complicate things, the metric may look messy in the new coordinate system. Thus, very few frames are going to be Lorentz.

Physical Interpretation of a Lorentz Frame
What the definition means physically is that an observer in the $\bar{x}$-frame who measures a particle traveling at light speed in the x-frame will also reach the conclusion that its speed is c, because he makes the decision based on (I), which is equivalent to (II). In other words:

A Lorentz frame in locally Minkowskian space is any frame in which light appears to be traveling at light speed, and where we measure length in the usual way.

Question Do all Lorentz frames at p have the property that metric has the nice form $\text{diag}[1, 1, 1, -c^2]$?
Answer Yes, as we shall see below.

Question OK. But if x and $\bar{x}$ are two Lorentz frames at the point p, how are they related?
Answer Here is an answer. First, continue to denote a specific Lorentz frame at the point p by x.

Theorem 7.3 (Criterion for Lorentz Frames)
The following are equivalent for a locally Minkowskian manifold M
(a) A coordinate system $\bar{x}^i$ is Lorentz at the point p
(b) If x is any frame such that, at p, $G = \text{diag}[1, 1, 1, -c^2]$, then the columns of the change-of-coordinate matrix

$$D^i_j = \frac{\partial \bar{x}^i}{\partial x^j}$$

satisfy

$$\langle \text{column } i, \text{ column } j \rangle = \langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle,$$

where the inner product is defined by the matrix G.
(c) $\bar{G} = \text{diag}[1, 1, 1, -c^2]$

Proof
(a) ⇒ (b) Suppose the coordinate system $\bar{x}^i$ is Lorentz at p, and let x be as hypothesized in (b). We proceed by invoking condition (a) of Definition 7.2 for several paths. (These paths will correspond to sending out light rays in various directions.)

Path *C*: $x^1 = ct$; $x^2 = x^3 = 0$, $x^4 = t$ (a photon traveling along the x^1-axis in E_4). This gives

$$T = (c, 0, 0, 1),$$

and hence $\|T\|^2 = 0$, and hence Definition 7.2 (a) applies. Let D be the change-of-basis matrix to the (other) inertial frame $\bar{x}^i$;

$$D_k{}^i = \frac{\partial \bar{x}^i}{\partial x^k},$$

so that

$$\bar{T}^i = D_k{}^i T^k$$

$$= \begin{bmatrix} D^1{}_1 & D^1{}_2 & D^1{}_3 & D^1{}_4 \\ D^2{}_1 & D^2{}_2 & D^2{}_3 & D^2{}_4 \\ D^3{}_1 & D^3{}_2 & D^3{}_3 & D^3{}_4 \\ D^4{}_1 & D^4{}_2 & D^4{}_3 & D^4{}_4 \end{bmatrix} \begin{bmatrix} c \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

By property (a) of Definition 7.2,

$$(\bar{T}^1)^2 + (\bar{T}^1)^2 + (\bar{T}^1)^2 - c^2(\bar{T}^1)^2 = 0,$$

so that

$$(cD_1{}^1 + D^1{}_4)^2 + (cD_1{}^2 + D^2{}_4)^2 + (cD_1{}^3 + D^3{}_4)^2 - c^2(cD_1{}^4 + D^4{}_4)^2 = 0 \qquad \ldots \quad (*)$$

If we reverse the direction of the photon, we similarly get

$$(-cD_1{}^1 + D^1{}_4)^2 + (-cD_1{}^2 + D^2{}_4)^2 + (-cD_1{}^3 + D^3{}_4)^2 - c^2(-cD_1{}^4 + D^4{}_4)^2 = 0 \qquad \ldots(**)$$

Noting that this only effects cross-terms, subtracting and dividing by $4c$ gives

$$D_1{}^1 D^1{}_4 + D_1{}^2 D^2{}_4 + D_1{}^3 D^3{}_4 - c^2 D_1{}^4 D^4{}_4 = 0;$$

that is,

$$\langle column\ 1, column\ 4\rangle = 0 = \langle e_1, e_4\rangle.$$

In other words, the first and fourth columns of D are orthogonal under the Minkowskian inner product. Similarly, by sending light beams in the other directions, we see that the other columns of D are orthogonal to the fourth column.

If, instead of subtracting, we now add (*) and (**), and divide by 2, we get

$$c^2[D_1{}^1 D_1{}^1 + D_1{}^2\, D_1{}^2 + D_1{}^3\, D_1{}^3 - c^2 D_1{}^4\, D_1{}^4] \\ + [D_4{}^1 D_4{}^1 + D_4{}^2\, D_4{}^2 + D_4{}^3\, D_4{}^3 - c^2 D_4{}^4\, D_4{}^4] = 0,$$

showing that

$$c^2\langle \text{column 1, column 1}\rangle = -\langle \text{column 4, column 4}\rangle.$$

So, if we write

$$\langle \text{column 1, column 1} \rangle = k,$$

then

$$\langle \text{column 4, column 4} \rangle = -c^2 k \qquad \ldots \; (***)$$

Similarly (by choosing other photons) we can replace column 1 by either column 2 or column 3, showing that if we take

$$\langle \text{column 1, column 1} \rangle = k,$$

we have

$$\langle \text{column } i, \text{column } i \rangle = \begin{cases} k & \text{if } 1 \le i \le 3 \\ -kc^2 & \text{if } i = 4 \end{cases} .$$

Let us now take another, more interesting, photon given by

Path *D*: $x^1 = (c/\sqrt{2}\,)t;\;\; x^2 = -(c/\sqrt{2}\,)t;\;\; x^3 = 0;\;\; x^4 = t$, with

$$\boldsymbol{T} = \langle c/\sqrt{2}\,, -c/\sqrt{2}\,, 0, 1 \rangle.$$

(You can check to see that $\|T\|^2 = 0$, so that it does indeed represent a photon.) Since $\|\overline{T}\|^2 = 0$, we get

$$(D_1{}^1 c/\sqrt{2} - D_2{}^1 c/\sqrt{2} + D_4{}^1)^2 + (D_1{}^2 c/\sqrt{2} - D_2{}^2 c/\sqrt{2} + D_4{}^2)^2$$
$$+ (D_1{}^3 c/\sqrt{2} - D_2{}^3 c/\sqrt{2} + D_4{}^3)^2 - c^2(D_1{}^4 c/\sqrt{2} - D_2{}^4 c/\sqrt{2} + D_4{}^4)^2 = 0$$

and, looking at a similar photon traveling in the opposite x^2-direction,

$$(D_1{}^1 c/\sqrt{2} + D_2{}^1 c/\sqrt{2} + D_4{}^1)^2 + (D_1{}^2 c/\sqrt{2} + D_2{}^2 c/\sqrt{2} + D_4{}^2)^2$$
$$+ (D_1{}^3 c/\sqrt{2} + D_2{}^3 c/\sqrt{2} + D_4{}^3)^2 - c^2(D_1{}^4 c/\sqrt{2} + D_2{}^4 c/\sqrt{2} + D_4{}^4)^2 = 0$$

Subtracting these gives

$$2c^2[D_1{}^1 D_2{}^1 + D_1{}^2 D_2{}^2 + D_1{}^3 D_2{}^3 - c^2 D_1{}^4 D_2{}^4]$$
$$+ 4c/\sqrt{2}\,[D_2{}^1 D_4{}^1 + D_2{}^2 D_4{}^2 + D_2{}^3 D_4{}^3 - c^2 D_2{}^4 D_4{}^4] = 0.$$

But we already know that the second term vanishes, so we are left with

$$D_1{}^1 D_2{}^1 + D_1{}^2 D_2{}^2 + D_1{}^3 D_2{}^3 - c^2 D_1{}^4 D_2{}^4 = 0,$$

showing that columns 1 and 2 are also orthogonal.

Choosing similar photons now shows us that columns 1, 2, and 3 are mutually orthogonal. Therefore, we have

$$\langle \text{column } i,\ \text{column} j \rangle = \begin{cases} 0 & \text{if } i \neq j \\ k & \text{if } 1 \leq i = j \leq 3 \\ -kc^2 & \text{if } i = j = 4 \end{cases} \qquad \text{. ..... (IV)}$$

But, what is k? Let us invoke condition (b) of Defintion 7.2. To measure the length of a vector in the new frame, we need to transform the metric tensor using this coordinate change. Recall that, using matrix notation, the metric G transforms to $\bar{G} = P^T G P$, where P is the matrix inverse of D above. In the exercise set, you will see that the columns of P have the same property (IV) above, but with k replaced by $1/k$. But,

$$\bar{G} = P^T G P$$

Now, since G is just a constant multiple of an elementary matrix, all it does is multiply the last row of P by c^2. So, when we take $P^T(GP)$, we are really getting the funny dot product of the columns of P back again, which just gives a multiple of G. In other words, we get

$$\bar{G} = P^T G P = G/k.$$

Now we invoke condition (b) in Definition 7.2: Take the vector $\bar{V} = (1, 0, 0, 0)$ in the $\bar{x}$-frame. (Recognize it? It is the vector $\partial/\partial\bar{x}^1$.) Since its 4th coordinate is zero, condition (b) says that its norm-squared must be given by the usual length formula:

$$\|\bar{V}\|^2 = 1.$$

On the other hand, we can also use $\bar{G}$ to compuate $\|\bar{V}\|^2$, and we get

$$\|\bar{V}\|^2 = \frac{1}{k},$$

showing that $k = 1$. Hence, $\bar{G} = G$, and also D has the desired form. This proves (b) (and also (c), by the way).

(b) ⇒ (c) If the change of coordinate matrix has the above orthogonality property,

$$D_i^1 D_j^1 + D_i^2 D_j^2 + D_i^3 D_j^3 - c^2 D_i^4 D_j^4 = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } 1 \leq i = j \leq 3 \\ -c^2 & \text{if } i = j = 4 \end{cases}$$

then the argument in **(a)** ⇒ **(b)** shows that $\bar{G} = G$ (since $k = 1/k = 1$ here).

(c) ⇒ **(a)** If $\bar{G} = \text{diag}[1, 1, 1, -c^2]$ at the point p, then $\bar{x}$ is Lorentz at p, by the remarks preceding Definition 7.2.
❊

We will call the transformation from one Lorentz frame to another a **generalized Lorentz transformation.**

An Example of a Lorentz Transformation
We would like to give a simple example of such a transformation matrix D, so we look for a matrix D whose first column has the general form $\langle a, 0, 0, b\rangle$, with a and b non-zero constants. (Why? If we take $b = 0$, we will wind up with a less interesting transformation: a rotation in 3-space.) There is no loss of generality in taking $a = 1$, so let us use $\langle 1, 0, 0, -\beta/c\rangle$. Here, c is the speed of light, and β is a certain constant. (The meaning of β will emerge in due course). Its norm-squared is $(1 - \beta^2)$, and we want this to be 1, so we replace the vector by

$$\left\langle \frac{1}{\sqrt{1-\beta^2}}, 0, 0, -\frac{\beta/c}{\sqrt{1-\beta^2}} \right\rangle.$$

This is the first column of D. To keep things simple, let us take the next two columns to be the corresponding basis vectors e_2, e_3. Now we might be tempted to take the forth vector to be e_4, but that would not be orthogonal to the above first vector. By symmetry (to get a zero inner product) we are *forced* to take the last vector to be

$$\left\langle -\frac{\beta c}{\sqrt{1-\beta^2}}, 0, 0, \frac{1}{\sqrt{1-\beta^2}} \right\rangle$$

This gives the transformation matrix as

$$D = \begin{bmatrix} \frac{1}{\sqrt{1-\beta^2}} & 0 & 0 & -\frac{\beta c}{\sqrt{1-\beta^2}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\frac{\beta/c}{\sqrt{1-\beta^2}} & 0 & 0 & \frac{1}{\sqrt{1-\beta^2}} \end{bmatrix}.$$

and hence the new coordinates (by integrating everything in sight; using the boundary conditions $\bar{x}^i = 0$ when $x^i = 0$) as

$$\bar{x}^1 = \frac{x^1 - \beta c x^4}{\sqrt{1-\beta^2}}; \qquad \bar{x}^2 = x^2; \qquad \bar{x}^3 = x^3; \qquad \bar{x}^4 = \frac{x^4 - \beta x^1/c}{\sqrt{1-\beta^2}}.$$

Notice that solving the first equation for x^1 gives

$$x^1 = \bar{x}^1\sqrt{1-\beta^2} + \beta c x^4.$$

Since x^4 is just time t here, it means that the origin of the $\bar{x}$-system has coordinates (βct, 0, 0) in terms of the original coordinates. In other words, it is moving in the x-direction with a velocity of

$$v = \beta c,$$

so we must interpret β as the speed in "warp;"

$$\beta = \frac{v}{c}.$$

This gives us the famous

Lorentz Transformations of Special Relativity

If two Lorentz frames x and $\bar{x}$ have the same coordinates at $(x, y, z, t) = (0, 0, 0, 0)$, and if the $\bar{x}$-frame is moving in the x-direction with a speed of v, then the $\bar{x}$-coordinates of an event are given by

$$\bar{x} = \frac{x - vt}{\sqrt{1-v^2/c^2}}; \qquad \bar{y} = y; \qquad \bar{z} = z; \qquad \bar{t} = \frac{t - vx/c^2}{\sqrt{1-v^2/c^2}}$$

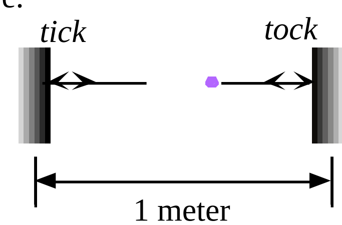
Exercise Set 7

1. What can be said about the scalar $\|dx^i/dt\|^2$ in a Lorentz frame for a particle traveling at **(a)** sub-light speed **(b)** super-light speed.

2. (a) Show that, if $x^i(t)$ is a timelike path in the Minkowskian manifold M so that $dx^4/dt \neq 0$, then $d\bar{x}^4/dt \neq 0$ in every Lorentz frame $\bar{x}$. In other words, if a particle is moving at sub-light speed in any one Lorentz frame, then it is moving at sub-light speed in all Lorentz frames.

(b) Conclude that, if a particle is traveling at super-light speed in one Lorentz frame, then it is traveling at super-light speeds in all such frames.

3. Referring to the Lorentz transformations for special relativity, consider a "photon clock" constructed by bouncing a single photon back and forth bewtwwen two parallel mirrors as shown in in the following figure.



Now place this clock in a train moving in the x-direction with velocity v. By comparing the time it takes between a *tick* and a *tock* for a stationary observer and one on the train, obtain the time contraction formula ($\Delta\bar{t}$ in terms Δt) from the length contraction one.

4. Prove the claim in the proof of 7.3, that if D is a 4×4 matrix whose columns satisfy

$$\langle \text{column } i, \text{ column} j\rangle = \begin{cases} 0 & \text{if } i \neq j \\ k & \text{if } 1 \leq i = j \leq 3 \\ -kc^2 & \text{if } i = j = 4 \end{cases},$$

using the Minkowski inner product G (*not* the standard inner product), then D^{-1} has its columns satisfying

$$\langle \text{column } i, \text{ column} j\rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1/k & \text{if } 1 \leq i = j \leq 3 \\ -c^2/k & \text{if } i = j = 4 \end{cases}.$$

[Hint: use the given property of D to write down the entries of its inverse P in terms of the entries of D.]

5. Invariance of the Minkowski Form

Show that, if $P = x^i_0$ and $Q = x^i_0 + \Delta x^i$ are any two events in the Lorentz frame x^i, then, for all Lorenz frames $\bar{x}^i$, one has

$$(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 - c^2(\Delta x^4)^2 = (\Delta\bar{x}^1)^2 + (\Delta\bar{x}^2)^2 + (\Delta\bar{x}^3)^2 - c^2(\Delta\bar{x}^4)^2$$

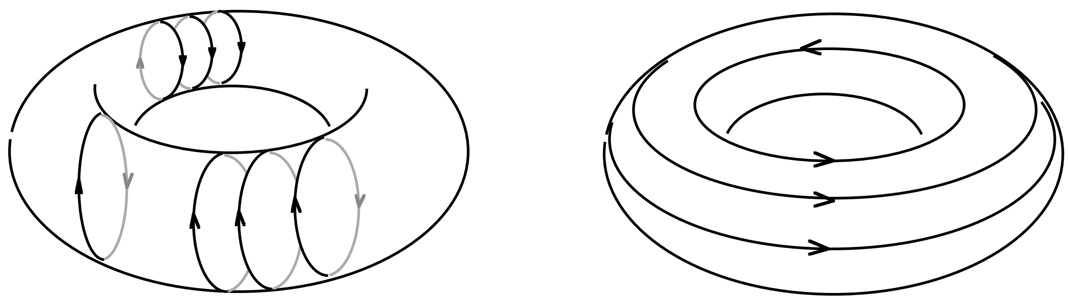
[Hint: Consider the path $x^i(t) = x_0{}^i + \Delta x^i t$, so that dx^i/dt is independent of t. Now use the transformation formula to conclude that $d\bar{x}^i/dt$ is also independent of t. (You might have to transpose a matrix before multiplying…) Deduce that $\bar{x}^i(t) = z^i + r^i t$ for some constants r^i and s^i. Finally, set $t = 0$ and $t = 1$ to conclude that $\bar{x}^i(t) = \bar{x}_0{}^i + \Delta\bar{x}^i t$, and apply (c) above.]

6. If the $\bar{x}^i$-system is moving with a velocity v in a certain direction with resepct to the x^i-system, we call this a **boost** in the given direction. Show that successive boosts in two perpendicular directions do not give a "pure" boost (the spatial axes are rotated—no longer parallel to the original axes). Now do some reading to find the transformation for a pure boost in an arbitrary direction.

---

8. Covariant Differentiation

Intuitively, by a parallel vector field, we mean a vector field with the property that the vectors at different points are parallel. Is there a notion of a parallel field on a manifold? For instance, in E_n, there is an obvious notion: just take a fixed vector $\boldsymbol{v}$ and translate it around. On the torus, there are good candidates for parallel fields (see the figure) but not on the 2-sphere. (There are, however, parallel fields on the 3-sphere…)



Let us restrict attention to parallel fields *of constant length.* Usually, we can recognize such a field by taking the derivatives of its coordinates, or by following a path, and taking the derivative of the vector field with respect to t: we should come up with zero. The problem is, we won't always come up with zero if the coordinates are not rectilinear, since the vector field may change direction as we move along the curved coordinate axes.

Technically, this says that, if X^j was such a field, we should check for its parallelism by taking the derivatives dX^j/dt along some path $x^i = x^i(t)$. However, there are two catches to this approach: one geometric and one algebraic.

Geometric Look, for example, at the filed on either torus in the above figure. Since it is circulating and hence non-constant, $dX/dt \neq 0$, which is not what we want. However, the *projection* of dX/dt parallel to the manifold does vanish—we will make this precise below.

Algebraic Since

$$\bar{X}^j = \frac{\partial \bar{x}^j}{\partial x^h} X^h,$$

one has, by the product rule,

$$\frac{d\bar{X}^j}{dt} = \frac{\partial^2 \bar{x}^j}{\partial x^k \partial x^h} X^h \frac{dx^k}{dt} + \frac{\partial \bar{x}^j}{\partial x^h}\frac{dX^h}{dt}\ , \quad \text{.........................} \text{(I)}$$

showing that, unless the second derivatives vanish, dX/dt does not transform as a vector field. What this means in practical terms is that we cannot check for parallelism at present—even in E_3 if the coordinates are not linear.

The projection of dX/dt along M will be called the **covariant derivative** of X (with respect to t), and written DX/dt. To compute it, we need to do a little work. First, some linear algebra.

Lemma 8.1 (Projection onto the Tangent Space)
Let M be a Riemannian n-manifold with metric g, and let V be a vector in E_s. The projection πV of V onto T_m has (local) coordinates given by

$$(\pi V)^i = g^{ik}(V.\partial/\partial x^k),$$

where $[g^{ij}]$ is the matrix inverse of $[g_{ij}]$, and $g_{ij} = (\partial/\partial x^i).(\partial/\partial x^j)$ as usual.

Proof
We can represent V as a sum,

$$V = \pi V + V^\perp,$$

where $V^\perp$ is the component of V normal to T_m. Now write $\partial/\partial x^k$ as e_k, and write

$$\pi V = a^1 e_1 + \ldots + a^n e_n,$$

where the a^i are the desired local coordinates. Then

$$\begin{aligned} V &= \pi V + V^\perp \\ &= a^1 e_1 + \ldots + a^n e_n + V^\perp \end{aligned}$$

and so

$$\begin{aligned} V\cdot e_1 &= a^1 e_1\cdot e_1 + \ldots + a^n e_n\cdot e_1 + 0 \\ V\cdot e_2 &= a^1 e_1\cdot e_2 + \ldots + a^n e_n\cdot e_2 \\ &\ldots \\ V\cdot e_n &= a^1 e_1\cdot e_n + \ldots + a^n e_n\cdot e_n \end{aligned}$$

whci we can write in matrix form as

$$[V.e_i] = [a^i]g_{**}$$

whence

$$[a^i] = [V.e_i]g^{**}.$$

Finally, since g^{**} is symmetric, we can transpose everything in sight to get

$$[a^i] = g^{**}[V.e_i],$$

as required. ☒

For reasons that will become clear later, let us now look at some partial derivatives of the fundamental matrix $[g_{**}]$ in terms of ambeint coordinates.

$$\frac{\partial}{\partial x^p}[g_{qr}] = \frac{\partial}{\partial x^p}\left[\frac{\partial y_s}{\partial x^q}\frac{\partial y_s}{\partial x^r}\right]$$
$$= \frac{\partial^2 y_s}{\partial x^p \partial x^q}\frac{\partial y_s}{\partial x^r} + \frac{\partial^2 y_s}{\partial x^r \partial x^p}\frac{\partial y_s}{\partial x^q}$$

or

$$g_{qr,p} = y_{s,pq}\, y_{s,r} + y_{s,rp}\, y_{s,q}$$

Look now at what happens to the indices q, r, and p if we permute them (they're just letters, after all) cyclically in the above formula (that is, $p \to q \to r$), we get two more formulas.

$$g_{qr,p} = \boxed{y_{s,pq}\, y_{s,r}} + \boxed{y_{s,rp}\, y_{s,q}} \qquad \text{(Original formula)}$$
$$g_{rp,q} = y_{s,qr}\, y_{s,p} + \boxed{y_{s,pq}\, y_{s,r}}$$
$$g_{pq,r} = \boxed{y_{s,rp}\, y_{s,q}} + y_{s,qr}\, y_{s,p}$$

Note that each term on the right occurs twice altogether as shown by the boxes. This permits us to solve for the completely boxed term $y_{s,pq}\, y_{s,r}$ by addin gthe first two equations and subtracting the third:

$$y_{s,pq}\, y_{s,r} = \frac{1}{2}[\, g_{qr,p} + g_{rp,q} - g_{pq,r} \,].$$

Definition 8.2 Christoffel Symbols
We make the following definitions.

$$[pq,\, r] = \frac{1}{2}[\, g_{qr,p} + g_{rp,q} - g_{pq,r} \,] \qquad \textbf{Christoffel Symbols of the First Kind}$$
$$\begin{Bmatrix} i \\ pq \end{Bmatrix} = g^{ir}\,[pq,\, r] \qquad \textbf{Christoffel Symbols of the Second Kind}$$
$$= \frac{1}{2} g^{ir}[\, g_{qr,p} + g_{rp,q} - g_{pq,r} \,]$$

Neither of these gizmos are tensors, but instead transform as follows (Which you will prove in the exercises!)

Transformation Law for Christoffel Symbols of the First Kind

$$[hk,\, l] = \overline{[ri,\, j]}\ \frac{\partial\bar{x}^r}{\partial x^h}\frac{\partial\bar{x}^i}{\partial x^k}\frac{\partial\bar{x}^j}{\partial x^l} + \bar{g}_{ij}\frac{\partial^2\bar{x}^i}{\partial x^h \partial x^k}\frac{\partial\bar{x}^j}{\partial x^l}$$

Transformation Law for Christoffel Symbols of the Second Kind

$$\begin{Bmatrix} p \\ hk \end{Bmatrix} = \overline{\begin{Bmatrix} t \\ ri \end{Bmatrix}}\ \frac{\partial x^p}{\partial\bar{x}^t}\frac{\partial\bar{x}^r}{\partial x^h}\frac{\partial\bar{x}^i}{\partial x^k} + \frac{\partial x^p}{\partial\bar{x}^t}\frac{\partial^2\bar{x}^t}{\partial x^h \partial x^k}$$

(Look at how the patterns of indices match those in the Christoffel symbols...)

Proposition 8.2 (Formula for Coavariant Derivative)

$$\frac{DX^i}{dt} = \frac{dX^i}{dt} + \begin{Bmatrix} i \\ pq \end{Bmatrix} X^p \frac{dx^q}{dt}$$

Proof By definition,

$$\frac{DX}{dt} = \pi\frac{dX}{dt}\ ,$$

which, by the lemma, has local coordinates given by

$$\frac{DX^i}{dt} = g^{ir}\left(\frac{dX}{dt} \cdot \frac{\partial}{\partial x^r}\right).$$

To evaluate the term in parentheses, we use ambeint coordinates. *dX*/*dt* has ambient coordinates

$$\frac{d}{dt}\left(X^p\frac{\partial y_s}{\partial x^p}\right) = \frac{dX^p}{dt}\frac{\partial y_s}{\partial x^p} + X^p\frac{\partial^2 y_s}{\partial x^p \partial x^q}\frac{dx^q}{dt}\ .$$

Thus, dotting with $\partial/\partial x^k = \partial y_s/\partial x^r$ gives

$$\frac{dX^p}{dt}\frac{\partial y_s}{\partial x^p}\frac{\partial y_s}{\partial x^r} + X^p\frac{\partial^2 y_s}{\partial x^p \partial x^q}\frac{\partial y_s}{\partial x^r}\frac{dx^q}{dt}$$

$$= \frac{dX^p}{dt}g_{pr} + X^p[pq,\, r]\frac{dx^q}{dt}\ .$$

Finally,

$$
\begin{aligned}
\frac{DX^i}{dt} &= g^{ir}\left(\frac{dX}{dt} \cdot \frac{\partial}{\partial x^r}\right) \\
&= g^{ir}\left(\frac{dX^p}{dt} g_{pr} + X^p[pq,\, r]\, \frac{dx^q}{dt}\right) \\
&= \delta^i_p \frac{dX^p}{dt} + \left\{ {i \atop pq} \right\} X^p \frac{dx^q}{dt} \quad \text{(Defn of Christoffel symbols of the 2nd Kind)} \\
&= \frac{dX^i}{dt} + \left\{ {i \atop pq} \right\} X^p \frac{dx^q}{dt}
\end{aligned}
$$

as required. ☒

In the exercises, you will check directly that the covariant derivative transforms correctly.

This allows us to say whether a field is parallel and of constant length by seeing whether this quantity vanishes. This claim is motivated by the following.

Proposition 8.3 (Parallel Fields of Constant Length)
X^i is a parallel field of constant length in E_n iff $DX^i/dt = 0$ for all paths in E_n.

Proof Designate the usual coordinate system by x^i. Then X^i is parallel and of constant length iff its coordinates with respect to the chart x are constant; that is, iff

$$\frac{dX^i}{dt} = 0.$$

But, since for this coordinate system, $g_{ij} = \delta_{ij}$, the Christoffel symbols clearly vanish, and so

$$\frac{DX^i}{dt} = \frac{dX^i}{dt} = 0.$$

But, if the contravariant vector DX^i/dt vanishes under one coordinate system (whose domain happens to be the whole manifold) it must vanish under all of them. (Notice that we can't say that about things that are not vectors, such as dX^i/dt.) ◆

Partial Derivatives

Write

$$
\begin{aligned}
\frac{DX^i}{dt} &= \frac{dX^i}{dt} + \left\{ {i \atop pq} \right\} X^p \frac{dx^q}{dt} \\
&= \frac{\partial X^i}{\partial x^q} \frac{dx^q}{dt} + \left\{ {i \atop pq} \right\} X^p \frac{dx^q}{dt}
\end{aligned}
$$

$$= \left[\frac{\partial X^i}{\partial x^q} + \begin{Bmatrix} i \\ pq \end{Bmatrix} X^p \right] \frac{dx^q}{dt}$$

The quantity in brackets converts the vector dx^q/dt into the vector DX^i/dt. Moreover, since every contravariant vector has the form dx^q/dt (recall the definition of tangent vectors in terms of paths), it follows that the quantity in brackets "looks like" a tensor of type (1, 1), and we call it the q^{th} **covariant partial derivative** of X^i:

Definition 8.4 The **covariant partial derivative** of the contravariant field X^p is the type (1, 1) tensor given by

Covariant Partial Derivative of X^i

$$X^i{}_{|q} = \frac{\partial X^i}{\partial x^q} + \begin{Bmatrix} i \\ pq \end{Bmatrix} X^p$$

(Some texts use $\nabla_q X^i$.) Do you see now why it is called the "covariant" derivative? Similarly, we can obtain the type (0, 2) tensor (check that it transforms corectly)

Covariant Partial Derivative of Y_p

$$Y_{p|q} = \frac{\partial Y_p}{\partial x_q} - \begin{Bmatrix} i \\ pq \end{Bmatrix} Y_i$$

Notes

1. All these forms of derivatives satisfy the expected rules for sums and also products. (See the exercises.)

2. If C is a path on M, then we obtain the following analogue of the chain rule:

$$\frac{DX^i}{dt} = X^p{}_{|k} \frac{dx^k}{dt} .$$

(See the definitions.)

Exercise Set 8

1. (a) Show that $\begin{Bmatrix} i \\ jk \end{Bmatrix} = \begin{Bmatrix} i \\ kj \end{Bmatrix}$.

(b) If Γ^i_{jk} are functions that transform in the same way as Christoffel symbols of the second kind (called a **connection**) show that $\Gamma^i_{jk} - \Gamma^i_{kj}$ is always a type (1, 2) tensor (called the associated **torsion** tensor).

(c) If a_{ij} and g_{ij} are any two symmetric non-degenerate type (0, 2) tensor fields with associated Christoffel symbols $\begin{Bmatrix} i \\ jk \end{Bmatrix}_a$ and $\begin{Bmatrix} i \\ jk \end{Bmatrix}_g$ respectively. Show that

$$\begin{Bmatrix} i \\ jk \end{Bmatrix}_a - \begin{Bmatrix} i \\ jk \end{Bmatrix}_g$$

is a type (1, 2) tensor.

2. Covariant Differential of a Covariant Vector Field Show that, if Y_i is a covariant vector, then

$$DY_p = dY_p - \begin{Bmatrix} i \\ pq \end{Bmatrix} Y_i \, dx^q.$$

are the components of a covariant vector field. (That is, check that it transforms correctly.)

3. Covariant Differential of a Tensor Field Show that, it we define

$$DT^h_p = dT^h_p + \begin{Bmatrix} h \\ rq \end{Bmatrix} T^r_p \, dx^q - \begin{Bmatrix} i \\ pq \end{Bmatrix} T^h_i \, dx^q.$$

then the coordinates transform like a (1, 1) tensor.

4. Obtain the transformation equations for Chritstoffel symbols of the first and second kind. (You might wish to consult an earlier printing of these notes or the Internet site...)

5. Show directly that the coordinates of DX^p/dt transform as a contravariant vector.

6. Show that, if X^i is any vector field on E_n, then its ordinary partial derivatives agree with $X^p{}_{|k}$.

7. Show that, if X^i and Y^j are any two (contravariant) vector fields on M, then

$$(X^i + Y^i)_{|k} = X^i{}_{|k} + Y^i{}_{|k}$$
$$(X^iY^j)_{|k} = X^i{}_{|k}Y^j + X^iY^j{}_{|k}.$$

8. Show that, if C is a path on M, then

$$\frac{DX^i}{dt} = X^i{}_{|k}\frac{dx^k}{dt}.$$

9. Show that, if X and Y are vector fields, then

$$\frac{d}{dt}\langle X, Y\rangle = \langle \frac{DX}{dt}, Y\rangle + \langle X, \frac{DY}{dt}\rangle,$$

where the big D's denote covariant differentiation.

10. **(a)** What is $\phi_{|i}$ if ϕ is a scalar field?

(b) Give a definition of the "contravariant" derivative, $X^{a|b}$ of X^a with respect to x^b, and show that $X^{a|b} = 0$ if and only if $X^a{}_{|b} = 0$.

---

9. Geodesics and Local Inertial Frames

Let us now apply some of this theory to curves on manifolds. If a non-null curve C on M is paramaterized by $x^i(t)$, then we can reparamaterize the curve using arc length,

$$s(t) = \int_a^t \sqrt{\pm g_{ij}\frac{dx^i dx^j}{du\,du}}\, du,$$

(starting at some arbitrary point) as the parameter. The reason for wanting to do this is that the tangent vector $T^i = dx^i/ds$ is then a unit vector (see the exercises) and also independent of the paramaterization.

If we were talking about a curve in E_3, then the derivative of the unit tangent vector (again with respect to s to make it independent of the paramaterization) is normal to the curve, and its magnitude is a measure of how fast the curve is "turning," and so we call the derivative of T^i the **curvature** of C.

If C happens to be on a manifold, then the unit tangent vector is still

$$T^i = \frac{dx^i}{ds} = \frac{dx^i}{dt} \Big/ \frac{ds}{dt} = \frac{dx^i/dt}{\sqrt{\pm g_{pq}\frac{dx^p}{dt}\frac{dx^q}{dt}}}$$

(the last formula is there if you want to actually compute it). But, to get the curvature, we need to take the covariant derivative:

$$\begin{aligned} P^i &= \frac{DT^i}{ds} \\ &= \frac{D(dx^i/ds)}{ds} \\ &= \frac{d^2x^i}{ds^2} + \begin{Bmatrix} i \\ pq \end{Bmatrix} \frac{dx^p}{ds}\frac{dx^q}{ds} \end{aligned}$$

Definitions 9.1 The **first curvature vector** ***P*** of the curve C is

$$P^i = \frac{d^2x^i}{ds^2} + \begin{Bmatrix} i \\ pq \end{Bmatrix} \frac{dx^p}{ds}\frac{dx^q}{ds} .$$

A curve on M whose first curvature is zero is called a **geodesic**. Thus, a geodesic is a curve that satisfies the system of second order differential equations

$$\frac{d^2x^i}{ds^2} + \begin{Bmatrix} i \\ pq \end{Bmatrix} \frac{dx^p}{ds}\frac{dx^q}{ds} = 0.$$

In terms of the parameter t, this becomes (see the exercises)

$$\frac{d^2x^i}{dt^2}\frac{ds}{dt} - \frac{dx^i}{dt}\frac{d^2s}{dt^2} + \begin{Bmatrix} i \\ pq \end{Bmatrix} \frac{dx^p}{dt}\frac{dx^q}{dt}\frac{ds}{dt} = 0,$$

where

$$\frac{ds}{dt} = \sqrt{\pm g_{ij}\frac{dx^i dx^j}{dt\ dt}} .$$

Note that P is a tangent vector at right angles to the curve C which measures its change *relative to M*.

Question Why is P at right angles to C?
Answer This can be checked as follows:

$$\frac{d}{ds}\langle T, T\rangle = \langle \frac{DT}{ds}, T\rangle + \langle T, \frac{DT}{ds}\rangle \qquad \text{(Exercise Set 8 \#9)}$$

$$= 2\langle \frac{DT}{ds}, T\rangle \qquad \text{(symmetry of the scalar product)}$$

$$= 2\langle P, T\rangle \qquad \text{(definition of } P\text{)}$$

so that $\langle P, T\rangle = \frac{1}{2}\frac{d}{ds}\langle T, T\rangle.$

But $\langle T, T\rangle = \pm 1$ (refer back to the Proof of 6.5 to check this)

whence $\langle P, T\rangle = \frac{1}{2}\frac{d}{ds}(\pm 1) = 0,$

as asserted.

Local Flatness, or "Local Inertial Frames"

In "flat space" E_s all the Christoffel symbols vanish, so the following question arises:

Question Can we find a chart (local coordinate system) such that the Christoffel symbols vanish—at least in the domain of the chart?
Answer This is asking too much; we shall see later that the derivatives of the Christoffel symbols give an invariant tensor (called the **curvature**) which does not vanish in general. However, we *do* have the following.

Proposition 9.2 (Existence of a Local Inertial Frame)
If m is any point in the Riemannian manifold M, then there exists a local coordinate system x^i at m such that:

(a) $g_{ij}(m) = \begin{cases} \pm 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} = \pm\delta_{ij}$

(b) $\frac{\partial g_{ij}}{\partial x^k}(m) = 0$

We call such a coordinate system a **local inertial frame** or a **normal frame.**

(It follows that $\Gamma^j_{i\,k}(m) = 0$.) Note that, if M is locally Minkowskian, then local intertial frames are automatically Lorentz frames.

Before proving the proposition, we need a lemma.

Lemma 9.3 (Some Equivalent Things)
Let $m \in M$. Then the following are equivalent:
(a) $g_{pq,r}(m) = 0$ for all p, q, r.
(b) $[pq, r]_m = 0$ for all p, q, r.
(c) $\left\{ \begin{matrix} r \\ pq \end{matrix} \right\}_m = 0$ for all p, q, r.

Proof
(a) ⇒ (b) follows from the definition of Christoffel symbols of the first kind.
(b) ⇒ (a) follows from the identity

$$g_{pq,r} = [qr, p] - [rp, q] \qquad \text{(Check it!)}$$

(b) ⇒ (c) follows from the definition of Christoffel symbols of the second kind.
(c) ⇒ (b) follows from the inverse identity

$$[pq, r] = g_{sr} \left\{ \begin{matrix} r \\ pq \end{matrix} \right\} .$$

⌘

Proof of Proposition 9.2[‡] First, we need a fact from linear algebra: if $\langle -,- \rangle$ is an inner product on the vector space L, then there exists a basis $\{V(1), V(2), \ldots, V(n)\}$ for L such that

$$\langle V(i), V(j) \rangle = \begin{cases} \pm 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} \qquad = \pm\delta_{ij}$$

(To prove this, use the fact that any symmetric matrix can be diagonalized using a $P{-}P^T$ type operation.)

To start the proof, fix any chart x^i near m with $x^i(m) = 0$ for all i, and choose a basis $\{V(i)\}$ of the tangent space at m such that they satisfy the above condition. With our bare hands, we are now going to specify a new coordinate system be $\bar{x}^i = \bar{x}^i(x^j)$ such that

$$\bar{g}_{ij} = \langle V(i), V(j) \rangle \qquad \text{(showing part (a))}.$$

The functions $\bar{x}^i = \bar{x}^i(x^j)$ will be specified by constructing their inverse $x^i = x^i(\bar{x}^j)$ using a quadratic expression of the form:

[‡] This is my own version of the proof. There is a version in Bernard Schutz's book, but the proof there seems overly complicated and also has some gaps relating to consistncy of the systems of linear equations.

$x^i = \bar{x}^j A(i,j) + \frac{1}{2}\bar{x}^j\bar{x}^k B(i,j,k)$where $A(i,j)$ and $B(i,j,k)$ are constants. It will follow from Taylor's theorem (and the fact that $x^i(m) = 0$) that $A(i,j) = \left(\frac{\partial x^i}{\partial \bar{x}^j}\right)_m$ and $B(i,j,k)$ $= \left(\frac{\partial^2 x^i}{\partial \bar{x}^j \partial \bar{x}^k}\right)_m$

so that

$$x^i = \bar{x}^j \left(\frac{\partial x^i}{\partial \bar{x}^j}\right)_m + \frac{1}{2}\bar{x}^j\bar{x}^k \left(\frac{\partial^2 x^i}{\partial \bar{x}^j \partial \bar{x}^k}\right)_m$$

where all the partial derivatives are evaluated at m.

Note These partial derivatives are just (yet to be determined) numbers which, if we differentiate the above quadratic expression, turn out to be its actual partial derivatives evaluated at m.

In order to specify this inverse, all we need to do is specify the terms $A(i,j)$ and $B(i,j,k)$ above. In order to make the map invertible, we must also guarantee that the Jacobean $(\partial x^i/\partial \bar{x}^j)_m = A(i,j)$ is invertible, and this we shall do.

We also have the transformation equations

$$\bar{g}_{ij} = \frac{\partial x^k}{\partial \bar{x}^i} \frac{\partial x^l}{\partial \bar{x}^j} g_{kl} \qquad \ldots\ldots\ldots \qquad \text{(I)}$$

and we want these to be specified and equal to $\langle V(i),\ V(j)\rangle$ when evaluated at m. This is easy enough to do: Just set

$$A(i,j) = \left(\frac{\partial x^i}{\partial \bar{x}^j}\right)_m = V(j)^i.$$

For then, no matter how we choose the $B(i,j,k)$ we have

$$\begin{aligned}\bar{g}_{ij}(m) &= \left(\frac{\partial x^k}{\partial \bar{x}^i}\right)_m \left(\frac{\partial x^l}{\partial \bar{x}^j}\right)_m g_{kl} \\ &= V(i)^k\ V(j)^l g_{kl} \\ &= \langle V(i),\ V(j)\rangle,\end{aligned}$$

as desired. Notice also that, since the $\{V(i)\}$ are a basis for the tangent space, the change-of-coordinates Jacobean, whose columns are the $V(i)$, is automatically invertible. Also, the $V(i)$ are the coordinate axes of the new system.

(**An Aside** This is not the only choice we can make: We are solving the system of equations (I) for the n^2 unknowns $\partial x^i/\partial \bar{x}^j|_m$. The number of equations in (I) is not the expected n^2, since switching i and j results in the same equation (due to symmetry of the g's). The number of *distinct* equations is

$$n + \binom{n}{2} = \frac{n(n+1)}{2},$$

leaving us with a total of

$$n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}$$

of the partial derivatives $\partial x^i/\partial \bar{x}^j$ that we can choose arbitrarily.[†])

Next, we want to kill the partial derivatives $\partial \bar{g}_{ij}/\partial \bar{x}^a$ by choosing appropriate values for the $B(i, j, k)$ (that is, the second-order partial derivatives $\partial^2 x^i/\partial \bar{x}^j \partial \bar{x}^k$). By the lemma, it suffices to arrange that

$$\overline{\begin{Bmatrix} p \\ hk \end{Bmatrix}}(m) = 0.$$

But
$$\overline{\begin{Bmatrix} p \\ hk \end{Bmatrix}}(m) = \left(\begin{Bmatrix} t \\ ri \end{Bmatrix} \frac{\partial \bar{x}^p}{\partial x^t} \frac{\partial x^r}{\partial \bar{x}^h} \frac{\partial x^i}{\partial \bar{x}^k} + \frac{\partial \bar{x}^p}{\partial x^t} \frac{\partial^2 x^t}{\partial \bar{x}^h \partial \bar{x}^k} \right)(m)$$
$$\frac{\partial \bar{x}^p}{\partial x^t} \left(\begin{Bmatrix} t \\ ri \end{Bmatrix} \frac{\partial x^r}{\partial \bar{x}^h} \frac{\partial x^i}{\partial \bar{x}^k} + \frac{\partial^2 x^t}{\partial \bar{x}^h \partial \bar{x}^k} \right)(m)$$

so it suffices to arrange that

$$\frac{\partial^2 x^t}{\partial \bar{x}^h \partial \bar{x}^k}(m) = -\begin{Bmatrix} t \\ ri \end{Bmatrix} \frac{\partial x^r}{\partial \bar{x}^h} \frac{\partial x^i}{\partial \bar{x}^k} \ (m),$$

That is, all we need to do is to define

$$B(t, h, k) = -\begin{Bmatrix} t \\ ri \end{Bmatrix} \frac{\partial x^r}{\partial \bar{x}^h} \frac{\partial x^i}{\partial \bar{x}^k} \ (m),$$

and we are done. ◆

[†] In the real world, where $n = 4$, this is interpreted as saying that we are left with 6 degrees of freedom in choosing local coordinates to be in an inertial frame. Three of these correspond to changing the coordinates by a constant velocity (3 degrees of freedom) or rotating about some axis (3 degrees of freedom: two angles to specify the axis, and a third to specify the rotation).

Corollary 9.4 (Partial Derivatives Look Nice in Inertial Frames)
Given any point $m \in M$, there exist local coordinates such that

$$X^p{}_{|k}(m) = \left(\frac{\partial X^p}{\partial x^k}\right)_m$$

Also, the coordinates of $\left(\frac{\partial X^p}{\partial x^k}\right)_m$ in an inertial frame transform to those of $X^p{}_{|k}(m)$ in every frame.

Corollary 9.5 (Geodesics are Locally Straight in Inertial Frames)
If C is a geodesic passing through $m \in M$, then, in any inertial frame, it has zero classical curvature at m. (that is, $d^2x^I/ds^2 = 0$).

This is the reason we call them "inertial" frames: freely falling particles fall in straight lines in such frames (that is, with zero ciurvature, at least near the origin).

Question Is there a local coordinate system such that all geodesics are in fact straight lines?
Answer Not in general; if you make some geodesics straight, then others wind up curved. It is the *curvature tensor* that is responsible for this. This involves the derivatives of the Christoffel symbols, and we can't make it vanish.

Question If I throw a ball in the air, then the path is curved and also a geodesic. Does this mean that our earthly coordinates are not inertial?
Answer Yes. At each instant in time, we can construct a local inertial frame corresponding to that event. But this frame varies from point to point along our world line if our world line is not a geodesic (more about this below), and the only way our world line can be a geodesic is if we were freely falling (and therefore felt no gravity). Technically speaking, the "earthly" coordinates we use constitute a **momentary comoving reference frame;** it is inertial at each point along our world line, but the direction of the axes are constantly changing in space-time.

Proposition 9.6 (Changing Inertial Frames)
If x and $\bar{x}$ are inertial frames at $m \in M$, then, recalling that $\underline{D}$ is the matrix whose ij th entry is $(\partial x^i/\partial \bar{x}^j)$, one has

$$\det \underline{D} = \det \bar{D} = \pm 1$$

Proof By definition of inertial frames,

$$g_{ij}(m) = \begin{cases} \pm 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} = \pm\delta_{ij}$$

and similarly for $\bar{g}^{ij}$, so that $\bar{g}^{ij} = \pm g^{ij}$, whence $\det(g_{**}) = \pm \det(\bar{g}_{**}) = \pm 1$. On the other hand,

$$\bar{g}_{ij} = \frac{\partial x^k}{\partial \bar{x}^i}\frac{\partial x^l}{\partial \bar{x}^j}\, g_{kl},$$

which, in matrix form, becomes

$$\bar{g}_{**} = \underline{D}^T g_{**}\underline{D}.$$

Taking determinants gives

$$\det(\bar{g}_{**}) = \det(\underline{D}^T)\det(g_{**})\det(\underline{D}) = \det(\underline{D})^2 \det(g_{**}),$$

giving

$$\pm 1 = \pm\det(\underline{D})^2,$$

which must mean that $\det(\underline{D})^2 = +1$, so that $\det(\underline{D}) = \pm 1$ as claimed. ❄

Note that the above theorem also workds if we use units in which $\det g = -c^2$ as in Lorentz frames.

Definition 9.7 Two (not necessarily inertial) frames x and $\bar{x}$ have the **same parity** if $\det \bar{D} > 0$. An **orientation** of M is an atlas of M such that all the charts have the same parity. M is called **orientable** if it has such an atlas, and **oriented** if it is equipped with one.

Notes

1. Reversing the direction of any one of the axes reverses the orientation.
2. It follows that every orientable manifold has two orientations; one corresponding to each choice of equivalence class of orientations.
3. If M is an oriented manifold and $m \in M$, then we can choose an *oriented* inertial frame $\bar{x}$ at m, so that the change-of-coordinates matrix $\underline{D}$ has positive determinant. Further, if $\underline{D}$ happens to be the change-of-coordinates from one oriented inertial frame to another, then $\det(\underline{D}) = +1$.
4. E_3 has two orientations: one given by any left-handed system, and the other given by any right-handed system.
5. In the homework, you will see that spheres are orientable, whereas Klein bottles are not.

We now show how we can use inertial frames to construct a tensor field.

Definition 9.8 Let M be an *oriented* n-dimensional Riemannian manifold. The **Levi-Civita tensor ε of type $(0, n)$** or **volume form** is defined as follows. If $\bar{x}$ is any coordinate system and $m \in M$, then define

$$\begin{aligned}\bar{\varepsilon}_{i_1 i_2 \dots i_n}(m) &= \det\left(\underline{D}_{i_1}\ \underline{D}_{i_2}\ \dots\ \underline{D}_{i_n}\right)\\ &= \text{determinant of } \underline{D} \text{ with columns permuted according to the indices}\end{aligned}$$

where $\underline{D}_j$ is the j th column of the change-of-coordinates matrix $\partial x^k/\partial\bar{x}^l$, and where x is any *oriented* inertial frame at m.[††]

Note ε is a completely antisymmetric tensor. If $\bar{x}$ is itself an inertial frame, then, since $\det(\underline{D}) = +1$ (see Note 2 above) the coordinates of $\varepsilon(m)$ are given by

$$\bar{\varepsilon}_{i_1 i_2 \dots i_n}(m) = \begin{cases} 1 & \text{if } (i_1, i_2, \dots, i_n) \text{ is an even permutation of } (1, 2, \dots, n) \\ -1 & \text{if if } (i_1, i_2, \dots, i_n) \text{ is an odd permutation of } (1, 2, \dots, n) \end{cases}$$

.

(Compare this with the metric tensor, which is also "nice" in inertial frames.)

Proposition 9.9 (Levi-Civita Tensor)
The Levi-Civita tensor is a well-defined, smooth tensor field.

Proof To show that it is well-defined, we must show independence of the choice of inertial frames. But, if ε and μ are defined at $m \in M$ as above by using two different inertial frames, with corresponding change-of-coordinates matrices $\underline{D}$ and $\underline{E}$, then $\bar{\underline{D}}\underline{E}$ is the change-of coordinates from one inertial frame to another, and therefore has determinant 1. Now,

$$\begin{aligned}\bar{\varepsilon}_{i_1 i_2 \dots i_n}(m) &= \det\left(\underline{D}_{i_1}\ \underline{D}_{i_2}\ \dots\ \underline{D}_{i_n}\right)\\ &= \det \underline{D}\ \mathcal{E}_{i_1 i_2 \dots i_n}\end{aligned}$$

(where $\mathcal{E}_{i_1 i_2 \dots i_n}$ is the identity matrix with columns ordered as shown in the indices)

$$= \det \underline{D}\bar{\underline{D}}\underline{E}\ \mathcal{E}_{i_1 i_2 \dots i_n}$$

(since $\bar{\underline{D}}\underline{E}$ has determinant 1; this being where we use the fact that things are oriented!)

$$= \det \underline{E}\ \mathcal{E}_{i_1 i_2 \dots i_n} = \bar{\mu}_{i_1 i_2 \dots i_n},$$

showing it is well-defined at each point. We now show that it is a tensor. If $\bar{x}$ and $\bar{y}$ are any two oriented coordinate systems at m and change-of-coordinate matrices $\underline{D}$ and $\underline{E}$ with respect to some inertial frame x at m, and if the coordinates of the tensor with respect to

[††] Note that this tensor cannot be defined without a metric being present. In the absence of a metric, the best you can do is define a "relative tensor," which is not quite the same, and what Rund calls the "Levi-Civita symbols" in his book. Wheeler, *et al.* just define it for Minkowski space.

these coordinates are $\bar{\varepsilon}_{k_1 k_2 \dots k_n}$ and $\bar{\mu}_{r_1 r_2 \dots r_n} = \det\ (\underline{E}_{r_1}\ \underline{E}_{r_2}\ \dots\ \underline{E}_{r_n})$ respectively, then at the point m,

$$\bar{\varepsilon}_{k_1 k_2 \dots k_n} = \det\,(\underline{D}_{k_1}\ \underline{D}_{k_2}\ \dots\ \underline{D}_{k_n})$$

$$= \varepsilon_{i_1 i_2 \dots i_n} \frac{\partial x^{i_1}}{\partial \bar{x}^{k_1}} \frac{\partial x^{i_2}}{\partial \bar{x}^{k_2}} \dots \frac{\partial x^{i_n}}{\partial \bar{x}^{k_n}}$$

(by definition of the determinant! since $\varepsilon_{i_1 i_2 \dots i_n}$ is just the sign of the permutation!)

$$= \varepsilon_{i_1 i_2 \dots i_n} \frac{\partial x^{i_1}}{\partial \bar{y}^{r_1}} \frac{\partial x^{i_2}}{\partial \bar{y}^{r_2}} \dots \frac{\partial x^{i_n}}{\partial \bar{y}^{r_n}} \frac{\partial \bar{y}^{r_1}}{\partial \bar{x}^{k_1}} \frac{\partial \bar{y}^{r_2}}{\partial \bar{x}^{k_2}} \dots \frac{\partial \bar{y}^{r_n}}{\partial \bar{x}^{k_n}}$$

$$= \bar{\mu}_{r_1 r_2 \dots r_n} \frac{\partial \bar{y}^{r_1}}{\partial \bar{x}^{k_1}} \frac{\partial \bar{y}^{r_2}}{\partial \bar{x}^{k_2}} \dots \frac{\partial \bar{y}^{r_n}}{\partial \bar{x}^{k_n}},$$

showing that the tensor transforms correctly. Finally, we assert that $\det\,(\underline{D}_{k_1}\ \underline{D}_{k_2}\ \dots\ \underline{D}_{k_n})$ is a smooth function of the point m. This depends on the change-of-coordinate matrices to the inertial coordinates. But we saw that we could construct inertial frames by setting

$$\left(\frac{\partial x^i}{\partial \bar{x}^j}\right)_m = V(j)^i,$$

where the $V(j)$ were an orthogonal base of the tangent space at m. Since we can vary the coordinates of this base smoothly, the smoothness follows. ❄

Example In E_3, the Levi-Civita tensor coincides with the totally antisymmetric third-order tensor ε_{ijk} in Exercise Set 5. In the Exercises, we see how to use it to generalize the cross-product.

Exercise Set 9

1. Recall that we can define the arc length of a smooth non-null curve by

$$s(t) = \int_a^t \sqrt{\pm g_{ij} \frac{dx^i dx^j}{du\,du}}\ du.$$

Assuming that this function is invertible (so that we can express x^i as a function of s) show that

$$\left\| \frac{dx^i}{ds} \right\|^2 = 1.$$

2. Derive the equations for a geodesic with respect to the parameter t.

3. Obtain an analogue of Corollary 8.3 for the covariant partial derivatives of type (2, 0) tensors.

4. Use inertial frames argument to prove that $g_{ab|c} = g^{ab}{}_{|c} = 0$. (Also see Exercise Set 4 #1.)
5. Show that, if the columns of a matrix D are orthonormal, then det $D = \pm 1$.
6. Prove that, if ε is the Levi-Civita tensor, then, in any frame, $\varepsilon_{i_1 i_2 \ldots i_n} = 0$ whenever two of the indices are equal. Thus, the only non-zero coordinates occur when all the indices differ.
7. Use the Levi-Civita tensor to show that, if x is any inertial frame at m, and if $X(1), \ldots,$ $X(n)$ are any n contravariant vectors at m, then

$$\det \langle X(1) | \ldots | X(n) \rangle$$

is a scalar.
8. The Volume 1-Form (A Generalization of the Cross Product) If we are given $n-1$ vector fields $X(2), X(3), \ldots, X(n)$ on the n-manifold M, define a covariant vector field by

$$(X(2) \wedge X(3) \wedge \ldots \wedge X(n))_j = \varepsilon_{j\, i_2 \ldots i_n} X(2)^{i_2} X(3)^{i_3} \ldots X(n)^{i_n},$$

where ε is the Levi-Civita tensor. Show that, in any inertial frame at a point m on a Riemannian 4-manifold, $\|X(2) \wedge X(3) \wedge X(4)\|^2$ evaluated at the point m, coincides, up to sign, with the square of the usual volume of the three-dimensional parallelepiped spanned by these vectors by justifying the following facts.
(a) Restricting your attention to Riemannian 4-manifolds, let A, B, and C be vectors at m, and suppose—as you may—that you have chosen an inertial frame at m with the property that $A^1 = B^1 = C^1 = 0$. (Think about why you can you do this.) Show that, in this frame, $A \wedge B \wedge C$ has only one nonzero coordinate: the first.
(b) Show that, if we consider A, B and C as 3-vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ respectively by ignoring their first (zero) coordinate, then

$$(A \wedge B \wedge C)_1 = \boldsymbol{a} . (\boldsymbol{b} \times \boldsymbol{c}),$$

which we know to be ± the volume of the parallelepiped spanned by $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$.
(c) Defining $\|C\|^2 = C_i C_j g^{ij}$ (recall that g^{ij} is the inverse of g_{kl}), deduce that the scalar $\|A \wedge B \wedge C\|^2$ is numerically equal to square of the volume of the parallelepiped spanned by the vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$. (Note also that $\|A \wedge B \wedge C\|^2$, being a scalar, does not depend on the choice of coordinate system—we *always* get the same answer, no matter what coordinate system we choose.)
9. Define the **Levi-Civita tensor of type (n, 0)**, and show that

$$\varepsilon_{i_1 i_2 \ldots i_n} \varepsilon^{j_1 j_2 \ldots j_n} = \begin{cases} 1 & \text{if } (i_1, \ldots, i_n) \text{ is an even permutation of } (j_1, \ldots, j_n) \\ -1 & \text{if } (i_1, \ldots, i_n) \text{ is an odd permutation of } (j_1, \ldots, j_n) \end{cases} .$$

10. The Riemann Curvature Tensor

First, we need to know how to translate a vector along a curve C. Let X_j be a vector field. We have seen that a parallel vector field of constant length on M must satisfy

$$\frac{DX^j}{dt} = 0 \qquad \ldots\ldots\ldots \qquad \text{(I)}$$

for any path C in M.

Definition 10.1 The vector field X^j is **parallel along the curve *C*** if it satisfies

$$\frac{DX^j}{dt} = \frac{dX^j}{dt} + \Gamma^j_{i\,h} X^i \frac{dx^h}{dt} = 0,$$

for the specific curve C, where we are writing the Christoffel symbols $\left\{ {j \atop ih} \right\}$ as $\Gamma^j_{i\,h}$.

If X^j is parallel along C, which has parametrization with domain $[a, b]$ and corresponding points α and β on M, then, since

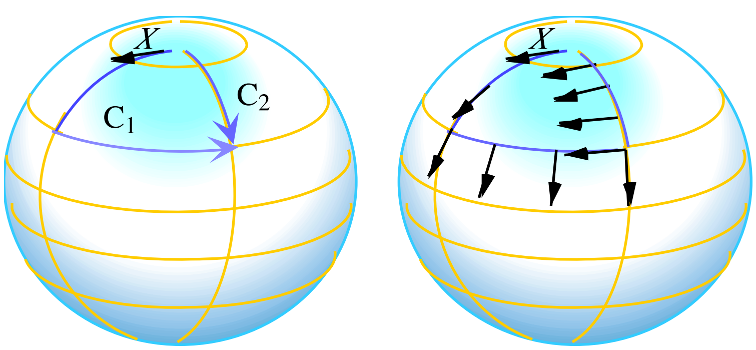
$$\frac{dX^j}{dt} = -\Gamma^j_{i\,h} X^i \frac{dx^h}{dt} \qquad \ldots\ldots\ldots \qquad \text{(I)}$$

we can integrate to obtain

$$X^j(\beta) = X^j(\alpha) - \int_a^b \Gamma^j_{i\,h} X^i \frac{dx^h}{dt}\, dt \qquad \ldots\ldots\ldots \qquad \text{(II)}$$

Question Given a fixed vector $X^j(\alpha)$ at the point $\alpha \in M$, and a curve C originating at α, it is possible to define a vector field along C by transporting the vector along C in a parallel fashion?

Answer Yes. Notice that the formula (II) is no good for this, since the integral already requires X^j to be defined along the curve before we start. But we can go back to (I), which is a system of first order linear differential equations. Such a system always has a unique solution with given initial conditions specified by $X^j(\alpha)$. Note however that it gives X^j as a function of the parameter t, and not necessarily as a well-defined function of position on M. In other words, the parallel transport of X at $p \in M$ depends on the path to p. (See the figure.) If it does not, then we have a parallelizable manifold.



Definition 10.2 If $X^j(\alpha)$ is any vector at the point $\alpha \in M$, and if C is any path from α to β in M, then the **parallel transport of $X^j(\alpha)$ along C** is the vector $X^j(\beta)$ given by the solution to the system (I) with initial conditions given by $X^j(\alpha)$.

Examples 9.3

(a) If C is a geodesic in M given by $x^i = x^i(s)$, where we are using arc-length s as the parameter (see Exercise Set 8 #1) then the vector field dx^i/ds is parallel along C. (Note that this field is only defined along C, but (I) still makes sense.) Why? because

$$\frac{D(dx^j/ds)}{Ds} = \frac{d^2x^j}{ds^2} + \Gamma^j_{i\,h}\frac{dx^i}{ds}\frac{dx^h}{ds},$$

which must be zero for a geodesic.

(b) Proper Coordinates in Relativity Along Geodesics

According to relativity, we live in a Riemannian 4-manifold M, but not the flat Minkowski space. Further, the metric in M has signature $(1, 1, 1, -1)$. Suppose C is a geodesic in M given by $x^i = x^i(t)$, satisfying the property

$$\langle \frac{dx^i}{dt}, \frac{dx^i}{dt} \rangle < 0.$$

Recall that we refer to such a geodesic as **timelike**. Looking at the discussion before Definition 7.1, we see that this corresponds, in Minkowski space, to a particle traveling at sub-light speed. It follows that we can choose an orthonormal basis of vectors $\{V(1), V(2), V(3), V(4)\}$ of the tangent space at m with the property given in the proof of 9.2, with $V(4) = dx^i/ds$ (actually, it is $dx^i/d\tau$ instead if our units have $c \neq 1$). We think of $V(4)$ as the unit vector in the direction of time, and $V(1)$, $V(2)$ and $V(3)$ as the spatial basis vectors. Using parallel translation, we obtain a similar set of vectors at each point along the path. (The fact that the curve is a geodesic guarantees that parallel translation of the time axis will remain parallel to the curve.) Finally, we can use the construction in 8.2 to flesh these frames out to full coordinate systems defined along the path. (Just having a set of orthogonal vectors in a manifold does not give a unique coordinate system, so we choose the unique local inertial one there, because in the eyes of the observer, spacetime should be flat.)

Question Does parallel transport preserve the relationship of these vectors to the curve. That is, does the vector $V(4)$ remain parallel, and do the vectors $\{V(1), V(2), V(3), V(4)\}$ remain orthogonal in the sense of 8.2?

Answer If X and Y are vector fields, then

$$\frac{d}{dt}\langle X, Y\rangle = \langle \frac{DX}{dt}, Y\rangle + \langle X, \frac{DY}{dt} \rangle,$$

where the big D's denote covariant differentiation. (Exercise Set 8 #9). But, since the terms on the right vanish for fields that have been parallel transported, we see that $\langle X, Y\rangle$ is independent of t, which means that orthogonal vectors remain orthogonal and that all the directions and magnitudes are preserved, as claimed.

Note At each point on the curve, we have a different coordinate system! All this means is that we have a huge collection of charts in our atlas; one corresponding to each point on the path. This (moving) coordinate system is called the **momentary comoving frame of reference** and corresponds to the "real life" coordinate systems.

(c) Proper Coordinates in Relativity Along Non-Geodesics

If the curve is not a geodesic, then parallel transport of a tangent vector need no longer be tangent. Thus, we cannot simply parallel translate the coordinate axes along the world line to obtain new ones, since the resulting frame may not be Lorentz. We shall see in Section 11 how to correct for that when we construct our comoving reference frames.

Question Under what conditions is parallel transport independent of the path? If this were the case, then we could use formula (I) to create a whole parallel vector field of constant length on M, since then $DX^j/dt = 0$.

Answer To answer this question, let us experiment a little with a fixed vector $V = X^j(a)$ by parallel translating it around a little rectangle consisting of four little paths. To simplify notation, let the first two coordinates of the starting point of the path (in some coordinates) be given by

$$x^1(a) = r,\ x^2(a) = s.$$

Then, choose δr and δs so small that the following paths are within the coordinate neighborhood in question:

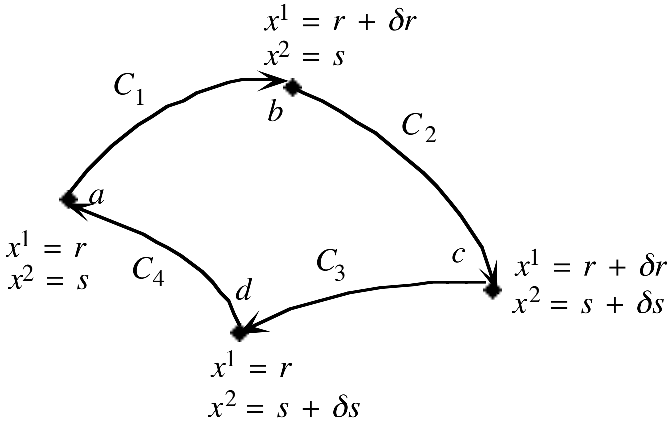
$$C_1\colon x^j(t) = \begin{cases} x^i(a) & \text{if } i \neq 1 \text{ or } 2 \\ r+t\delta r & \text{if } i = 1 \\ s & \text{if } i = 2 \end{cases}$$

$$C_2\colon x^j(t) = \begin{cases} x^i(a) & \text{if } i \neq 1 \text{ or } 2 \\ r+\delta r & \text{if } i = 1 \\ s+t\delta s & \text{if } i = 2 \end{cases}$$

$$C_3\colon x^j(t) = \begin{cases} x^i(a) & \text{if } i \neq 1 \text{ or } 2 \\ r+(1-t)\delta r & \text{if } i = 1 \\ s+\delta s & \text{if } i = 2 \end{cases}$$

$$C_4:\ x^j(t) = \begin{cases} x^i(a) & \text{if } i \neq 1 \text{ or } 2 \\ r & \text{if } i = 1 \\ s+(1-t)\delta s & \text{if } i = 2 \end{cases} \quad .$$

These paths are shown in the following diagram.



Now, if we parallel transport $X^j(\alpha)$ along C_1, we must have, by (II),

$$X^j(b) = X^j(a) - \int_0^1 \Gamma^j_{i\,h} X^i \frac{dx^h}{dt}\ dt \qquad \text{(since } t \text{ goes from 0 to 1 in } C_1\text{)}$$

$$= X^j(a) - \int_0^1 \Gamma^j_{i\,1} X^i\ \delta r\ dt\ . \qquad \text{(using the definition of } C_1 \text{ above)}$$

Warning: The integrand term $\Gamma^j_{i\,1} X^i$ is not constant, and must be evaluated as a function of t using the path C_1. However, if the path is a small one, then the integrand is approximately equal to its value at the midpoint of the path segment:

$$X^j(b) \approx X^j(a) - \Gamma^j_{i\,1} X^i(\text{midpoint of } C_1)\ \delta r$$

$$\approx X^j(a) - \left[\Gamma^j_{i\,1} X^i(a) + 0.5\frac{\partial}{\partial x^1}(\Gamma^j_{i\,1} X^i)\,\delta r \right] \delta r$$

where the partial derivative is evaluated at the point α. Similarly,

$$X^j(c) = X^j(b) - \int_0^1 \Gamma^j_{i\,2} X^i\ \delta s\ dt$$

$$\approx X^j(b) - \Gamma^j_{i\,2} X^i(\text{midpoint of } C_2)\delta s$$

$$\approx X^j(b) - \left[\Gamma^j_{i\,2} X^i(a) + \frac{\partial}{\partial x^1}(\Gamma^j_{i\,2} X^i)\,\delta r + 0.5\frac{\partial}{\partial x^2}(\Gamma^j_{i\,2} X^i)\,\delta s \right] \delta s$$

where all partial derivatives are evaluated at the point a. (This makes sense because the field is defined where we need it.)

$$X^j(d) = X^j(c) + \int_0^1 \Gamma^j_{i\,1} X^i \,\delta r\, dt$$
$$\approx X^j(c) + \Gamma^j_{i\,1} X^i(\text{midpoint of } C_3)\delta r$$
$$\approx X^j(c) + \left[\Gamma^j_{i\,1} X^i(a) + 0.5 \frac{\partial}{\partial x^1}(\Gamma^j_{i\,1} X^i)\,\delta r + \frac{\partial}{\partial x^2}(\Gamma^j_{i\,1} X^i)\,\delta s\right]\delta r$$

and the vector arrives back at the point a according to

$$X^{*j}(a) = X^j(d) + \int_0^1 \Gamma^j_{i\,2} X^i \,\delta s\, dt$$
$$\approx X^j(d) + \Gamma^j_{i\,2} X^i(\text{midpoint of } C_4)\,\delta s$$
$$\approx X^j(d) + \left[\Gamma^j_{i\,2} X^i(a) + 0.5 \frac{\partial}{\partial x^2}(\Gamma^j_{i\,2} X^i)\,\delta s\right]\delta s$$

To get the total change in the vector, you substitute back a few times and cancel lots of terms (including the ones with 0.5 in front), being left with

$$\delta X^j = X^{*j}(a) - X^j(a) \approx \left[\frac{\partial}{\partial x^2}(\Gamma^j_{i\,1} X^i) - \frac{\partial}{\partial x^1}(\Gamma^j_{i\,2} X^i)\right]\delta r \delta s$$

To analyze the partial derivatives in there, we first use the product rule, getting

$$\delta X^j \approx \left[X^i \frac{\partial}{\partial x^2}\Gamma^j_{i\,1} + \Gamma^j_{i\,1}\frac{\partial}{\partial x^2} X^i - X^i \frac{\partial}{\partial x^1}\Gamma^j_{i\,2} - \Gamma^j_{i\,2}\frac{\partial}{\partial x^1} X^i\right]\delta r \delta s \ldots\ldots\ldots \qquad \text{(III)}$$

Next, we recall the "chain rule" formula

$$\frac{DX^j}{dt} = X^j{}_{|h}\frac{dx^h}{dt}$$

in the homework. Since the term on the right must be zero *along each of the path segments* we see that (I) is equivalent to saying that the partial derivatives

$$X^j{}_{|h} = 0$$

for every index p and k (and along the relevant path segment)* since the terms dx^h/dt are non-zero. By definition of the partial derivatives, this means that

* Notice that we are taking partial derivatives in the direction of the path, so that they do make sense for this curious field that is only defined along the square path!

$$\frac{\partial X^j}{\partial x^h} + \Gamma^j_{i\,h} X^i = 0,$$

so that

$$\frac{\partial X^j}{\partial x^h} = -\,\Gamma^j_{i\,h} X^i.$$

We now substitute these expressions in (III) to obtain

$$\delta X^j \approx \left[X^i \frac{\partial}{\partial x^2}\Gamma^j_{i\,1} - \Gamma^j_{i\,1}\Gamma^i_{p\,2}X^p - X^i \frac{\partial}{\partial x^1}\Gamma^j_{i\,2} + \Gamma^j_{i\,2}\Gamma^i_{p\,1}X^p \right] \delta r \delta s$$

where everything in the brackets is evaluated at a. Now change the dummy indices in the first and third terms and obtain

$$\delta X^j \approx \left[\frac{\partial}{\partial x^2}\Gamma^j_{p\,1} - \Gamma^j_{i\,1}\Gamma^i_{p\,2} - \frac{\partial}{\partial x^1}\Gamma^j_{p\,2} + \Gamma^j_{i\,2}\Gamma^i_{p\,1} \right] X^p \; \delta r \delta s$$

This formula has the form

$$\delta X^j \approx R_{p\;12}^{\;j} X^p \; \delta r \delta s \qquad \ldots\ldots\ldots\ldots \qquad \text{(IV)}$$

(indices borrowed from the Christoffel symbol in the first term, with the extra index from the x in the denominator) where the quantity $R_{p\;12}^{\;j}$ is known as the **curvature tensor**.

Curvature Tensor

$$R_{b\;cd}^{\;a} = \left[\Gamma^i_{b\,c}\Gamma^a_{i\,d} - \Gamma^i_{b\,d}\Gamma^a_{i\,c} + \frac{\partial \Gamma^a_{b\,c}}{\partial x^d} - \frac{\partial \Gamma^a_{b\,d}}{\partial x^c} \right]$$

The terms are rearranged (and the Christoffel symbols switched) so you can see the index pattern, and also that the curvature is antisymmetric in the last two covariant indices.

$$R_{b\;cd}^{\;a} = -\,R_{b\;dc}^{\;a}$$

The fact that it is a tensor follows from the homework.

It now follows from a grid argument, that if C is any (possibly) large planar closed path within a coordinate neighborhood, then, if X is parallel transported around the loop, it arrives back to the starting point with change given by a sum of contributions of the form

(IV). If the loop is too large for a single coordinate chart, then we can break it into a grid so that each piece falls within a coordinate neighborhood. Thus we see the following.

Proposition 10.4 (Curvature and Parallel Transport)
Assume M is simply connected. A necessary and sufficient condition that parallel transport be independent of the path is that the curvature tensor vanishes.

Definition 10.5 A manifold with zero curvature is called **flat.**

Properties of the Curvature Tensor We first obtain a more explicit description of $R_b{}^a{}_{cd}$ in terms of the partial derivatives of the g_{ij}. First, introduce the notation

$$g_{ij,k} = \frac{\partial g_{ij}}{\partial x^k}$$

for partial derivatives, and remember that these are *not* tensors. Then, the Christoffel symbols and curvature tensor are given in the convenient form

$$\Gamma_b{}^a{}_c = \frac{1}{2} g^{ak}(g_{ck,b} + g_{kb,c} - g_{bc,k})$$

$$R_b{}^a{}_{cd} = [\Gamma_b{}^i{}_c \Gamma_i{}^a{}_d - \Gamma_b{}^i{}_d \Gamma_i{}^a{}_c + \Gamma_b{}^a{}_{c,d} - \Gamma_b{}^a{}_{d,c}].$$

We can lower the index by defining

$$R_{abcd} = g_{bi} R_a{}^i{}_{cd}$$

Substituting the first of the above (boxed) formulas into the second, and using symmetry of the second derivatives and the metric tensor, we find (exercise set)

Covariant Curvature Tensor in Terms of the Metric Tensor

$$R_{abcd} = \frac{1}{2}(g_{bc,ad} - g_{bd,ac} + g_{ad,bc} - g_{ac,bd}) + \Gamma_a{}^j{}_d \Gamma_{bjc} - \Gamma_a{}^j{}_c \Gamma_{bjd}$$

(We can remember this by breaking the indices a, b, c, d into pairs other than ab, cd (we can do this two ways) the pairs with a and d together are positive, the others negative.)

Notes

1. We have new kinds of Christoffel symbols Γ_{ijk} given by

$$\Gamma_{ijk} = g_{pj}\, \Gamma_i{}^p{}_k.$$

2. Some symmetry properties: $R_{abcd} = -R_{abdc} = -R_{bacd}$ and $R_{abcd} = R_{cdab}$ (see the exercise set)
3. We can raise the index again by noting that

$$g^{bi}R_{aicd} = g^{bi}g_{ij}R_a{}^j{}_{cd} = \delta^b_j R_a{}^j{}_{cd} = R_a{}^b{}_{cb}.$$

Now, let us evaluate some partial derivatives in an inertial frame (so that we can ignore the Christoffel symbols) cyclically permuting the last three indices as we go:

$$\begin{aligned} &R_{abcd,e} + R_{abec,d} + R_{abde,c} \\ &= \frac{1}{2}(g_{ad,bce} - g_{ac,bde} + g_{bc,ade} - g_{bd,ace} \\ &\qquad + g_{ac,bed} - g_{ae,bcd} + g_{be,acd} - g_{bc,aed} \\ &\qquad\qquad + g_{ae,bdc} - g_{ad,bec} + g_{bd,aec} - g_{be,adc}) \\ &= 0 \end{aligned}$$

Now, I claim this is also true for the *covariant* partial derivatives:

Bianchi Identities

$$R_{abcd|e} + R_{abec|d} + R_{abde|c} = 0$$

Indeed, let us evaluate the left-hand side at any point $m \in M$. Choose an inertial frame at m. Then the left-hand side coincides with $R_{abcd,e} + R_{abec,d} + R_{abde,c}$, which we have shown to be zero. Now, since a tensor which is zero is sone frame is zero in all frames, we get the result!

Definitions 10.6 The **Ricci tensor** is defined by

$$R_{ab} = R_a{}^i{}_{bi} = g^{ij}R_{ajbi}$$

we can raise the indices of any tensor in the usual way, getting

$$R^{ab} = g^{ai}g^{bj}R_{ij}.$$

In the exercise set, you will show that it is symmetric, and also (up to sign) is the only non-zero contraction of the curvature tensor.

We also define the **Ricci scalar** by

$$R = g^{ab}R_{ab} = g^{ab}g^{cd}R_{acbd}$$

The last thing we will do in this section is play around with the Bianchi identities. Multiplying them by g^{bc}:

$$g^{bc}[R_{abcd|e} + R_{abec|d} + R_{abde|c}] = 0$$

Since $g^{ij}{}_{|k} = 0$ (see Exercise Set 8), we can slip the g^{bc} into the derivative, getting

$$-R_{ad|e} + R_{ae|d} + R_a{}^c{}_{de|c} = 0.$$

Contracting again gives

$$g^{ad}[-R_{ad|e} + R_{ae|d} + R_a{}^c{}_{de|c}] = 0,$$

or

$$-R_{|e} + R^d{}_{e|d} + R^{dc}{}_{de|c} = 0,$$

or

$$-R_{|e} + R^d{}_{e|d} + R^c{}_{e|c} = 0.$$

Combining terms and switching the order now gives

$$R^b{}_{e|b} - \tfrac{1}{2}R_{|e} = 0,$$

or

$$R^b{}_{e|b} - \tfrac{1}{2}\delta^b_e R_{|b} = 0$$

Multiplying this by g^{ae}, we now get

$$R^{ab}{}_{|b} - \tfrac{1}{2}g^{ab}R_{|b} = 0, \qquad (R^{ab} \text{ is symmetric})$$

or

$$G^{ab}{}_{|b} = 0,$$

where we make the following definition:

Einstein Tensor

$$G^{ab} = R^{ab} - \tfrac{1}{2}g^{ab}R$$

Einstein's field equation for a vacuum states that

$$G^{ab} = 0$$

(as we shall see later…).

Example 10.7
Take the 2-sphere of radius r with polar coordinates, where we saw that

$$g_{**} = \begin{bmatrix} r^2\sin^2\phi & 0 \\ 0 & r^2 \end{bmatrix}.$$

The coordinates of the covariant curvature tensor are given by

$$R_{abcd} = \tfrac{1}{2}(g_{bc,ad} - g_{bd,ac} + g_{ad,bc} - g_{ac,bd}) + \Gamma_{a\,c}^{\,j}\Gamma_{jbd} - \Gamma_{a\,d}^{\,j}\Gamma_{jbc}.$$

Let us calculate $R_{\theta\phi\theta\phi}$. (Note: when we use Greek letters, we are referring to specific terms, so there is no summation when the indices repeat!) So, $a = c = \theta$, and $b = d = \phi$. (Incidentally, this is the same as $R_{\phi\theta\phi\theta}$ by the last exercise below.)

The only non-vanishing second derivative of g_{**} is

$$g_{\theta\theta,\phi\phi} = 2r^2(\cos^2\phi - \sin^2\phi),$$

giving

$$\tfrac{1}{2}(g_{\phi\theta,\theta\phi} - g_{\phi\phi,\theta\theta} + g_{\theta\phi,\phi\theta} - g_{\theta\theta,\phi\phi}) = r^2(\sin^2\phi - \cos^2\phi).$$

The only non-vanishing *first* derivative of g_{**} is

$$g_{\theta\theta,\phi} = 2r^2\sin\phi\cos\phi,$$

giving

$$\Gamma_{a\,c}^{\,j}\Gamma_{jbd} = \Gamma_{\theta\,\theta}^{\,j}\Gamma_{j\phi\phi} = 0,$$

since $b = d = \phi$ eliminates the second term (two of these indices need to be θ in order for the term not to vanish.)

$$\Gamma_{a\,d}^{\,j}\Gamma_{jbc} = \Gamma_{\theta\,\phi}^{\,j}\Gamma_{j\phi\theta} = \tfrac{1}{4}\left(\frac{2\cos\phi}{\sin\phi}\right)(-2r^2\sin\phi\cos\phi) = -r^2\cos^2\phi$$

Combining all these terms gives

$$\begin{aligned} R_{\theta\phi\theta\phi} &= r^2(\sin^2\phi - \cos^2\phi) + r^2\cos^2\phi \\ &= r^2\sin^2\phi. \end{aligned}$$

We now calculate

$$\begin{aligned} R_{ab} &= g^{cd}R_{acbd} \\ R_{\theta\theta} &= g^{\phi\phi}R_{\theta\phi\theta\phi} = \sin^2\phi \end{aligned}$$

and

$$R_{\phi\phi} = g^{\theta\theta}R_{\phi\theta\phi\theta} = \frac{\sin^2\phi}{\sin^2\phi} = 1.$$

All other terms vanish, since g is diagonal and R_{****} is antisymmetric. This gives

$$R = g^{ab}R_{ab} = g^{\theta\theta}R_{\theta\theta} + g^{\phi\phi}R_{\phi\phi} = \frac{1}{r^2\sin^2\phi}(\sin^2\phi) + \frac{1}{r^2} = \frac{2}{r^2}.$$

Summary of Some Properties of Curvature Etc.

$\Gamma_{a\ c}^{\ b} = \Gamma_{c\ a}^{\ b}$ $\Gamma_{abc} = \Gamma_{cba}$

$R_{a\ cd}^{\ b} = R_{a\ dc}^{\ b}$

$R_{abcd} = -R_{bacd}$ $R_{abcd} = -R_{abdc}$

$R_{abcd} = R_{cdab}$ Note that a,b and c,d always go together

$R_{ab} = R_{a\ bi}^{\ i} = g^{ij}R_{aibi}$

$R_{ab} = R_{ba}$

$R = g^{ab}R_{ab} = g^{ac}g^{bd}R_{abcd}$

$R^a_{\ b} = g^{ai}R_{ib}$

$R^{ab} = g^{ai}g^{bj}R_{ij}$

$G^{ab} = R^{ab} - \frac{1}{2}g^{ab}R$

Exercise Set 10

1. Derive the formula for the curvature tensor in terms of the g_{ij}.

2. (a) Show that the curvature tensor is antisymmetric in the last pair of variables:

$$R_{b\ cd}^{\ a} = -R_{b\ dc}^{\ a}$$

(b) Use part (a) to show that the Ricci tensor is, up to sign, the only non-zero contraction of the curvature tensor.

(c) Prove that the Ricci tensor is symmetric.

3. (cf. Rund, pp. 82-83)

(a) Show that

$$X^j_{\ |h|k} = \frac{\partial}{\partial x^k}(X^j_{\ |h}) + \Gamma_{m\ k}^{\ j}(X^m_{\ |h}) - \Gamma_{h\ k}^{\ l}(X^j_{\ |l})$$

$$= \frac{\partial^2 X^j}{\partial x^h \partial x^k} + X^l \frac{\partial}{\partial x^k}\Gamma_{l\,h}^{j} + \Gamma_{l\,h}^{j}\frac{\partial}{\partial x^k}X^l + \Gamma_{m\,k}^{\ j}\frac{\partial}{\partial x^h}X^m + \Gamma_{m\,k}^{\ j}\Gamma_{l\ h}^{\ m}X^l - \Gamma_{h\,k}^{\ l}(X^j_{\ |\,l})$$

(b) Deduce that

$$X^j_{\ |h|k} - X^j_{\ |k|h} = R^j_{l\,hk}X^l - S_{h\,k}^{\ l}X^j_{\ |\,l} = R^j_{l\,hk}X^l$$

where

$$S_{h\,k}^{\ l} = \Gamma_{h\,k}^{\ l} - \Gamma_{k\,h}^{\ l} = 0.$$

(c) Now deduce that the curvature tensor is indeed a type (1, 3) tensor.

4. Show that R_{abcd} is antisymmetric on the pairs (a, b) and (c, d).

5. Show that $R_{abcd} = R_{cdab}$ by first checking the identity in an inertial frame.

---

11. A Little More Relativity: Comoving Frames and Proper Time

Definition 11.1 A **Minkowskian 4-manifold** is a 4-manifold in which the metric has signature $(1, 1, 1, -1)$ (eg., the world according to Einstein).

By Proposition 9.2, if M is Minkowskian and $m \in M$, then one can find a locally inertial frame at m such that the metric at m has the form $\mathrm{diag}(1, 1, 1, -1)$. We actually have some flexibility: we can, if we like, adjust the scaling of the x^4-coordinate to make the metric look like $\mathrm{diag}(1, 1, 1, -c^2)$. In that case, the last coordinate is the **local time** coordinate. Later, we shall convert to units of time to make $c = 1$, but for now, let us use this latter kind of inertial frame.

Note If M is Minkowski space E_4, then inertial frames are nothing more than Lorentz frames. (We saw in Theorem 7.3 that Lorentz frames were characterized by the fact that the metric had the form $\mathrm{diag}(1, 1, 1, -c^2)$ at every point, so they are automatically inertial everywhere.)

Now let C be a timelike curve in the Minkowskian 4-manifold M.

Definition 11.2 A **momentary comoving reference frame for *C* (MCRF)** associates to each point $m \in C$ a locally inertial frame whose last basis vector is parallel to the curve and in the direction of increasing parameter s. Further, we require the frame coordinates to vary smoothly with the parameter of the curve.

Proposition 11.3 (Existence of MCRF's)
If C is any timelike curve in the Minkowskian 4-manifold M, then there exists an MCRF for C.

Proof
Fix $p_0 \in C$ and a Lorentz frame $W(1), W(2), W(3), W(4)$ of M_{p_0} (so that $g_{**} = \mathrm{diag}(1, 1, 1, -c^2)$.) We want to change this set to a new Lorentz frame $V(1), V(2), V(3), V(4)$ with

$$V(4) = \frac{dx^i}{d\tau} \qquad \text{Recall that } \tau = s/c$$

So let us take $V(4)$ as above. Then it is tangent to C at p_0. Further,

$$\|V(4)\|^2 = \left|\left|\frac{dx^i}{d\tau}\right|\right|^2 = \left|\left|\frac{dx^i}{ds}\right|\right|^2 \left|\frac{ds}{d\tau}\right|^2 = (-1)c^2 = -c^2.$$

Intuitively, $V(4)$ is the time axis for the observer at p_0: it points in the direction of increasing proper time τ. We can now flesh out this orthonormal set to obtain an inertial frame at p_a. For the other vectors, take

$$V(i) = W(i) + \frac{2}{c^2}\langle W(i), V(4)\rangle V(4)$$

for $i = 1, 2, 3$. Then

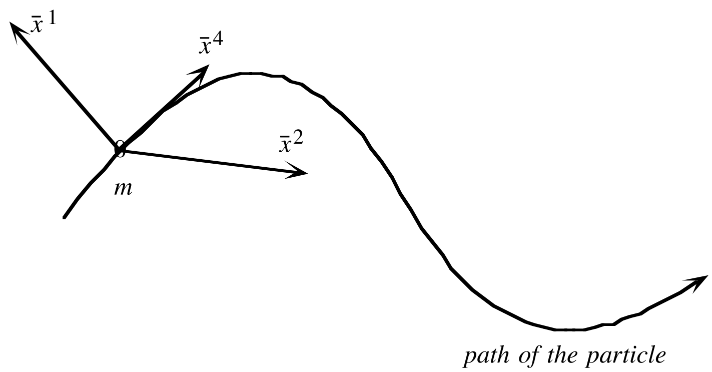
$$\langle V(i),V(j)\rangle = \langle W(i),W(j)\rangle + \frac{4}{c^2}\langle W(i),W(4)\rangle \langle W(j),W(4)\rangle + \frac{4}{c^4}\langle W(i),V(4)\rangle\langle W(j),V(4)\rangle \|V(4)\|^2$$
$$= 0$$

by orthogonality of the W's and the calculation of $\|V(4)\|^2$ above. Also,

$$\langle V(i),V(i)\rangle = \langle W(i),W(i)\rangle + \frac{4}{c^2}\langle W(i),W(4)\rangle^2 + \frac{4}{c^4}\langle W(i),V(4)\rangle^2 \|V(4)\|^2$$
$$= \|W(i)\|^{\varphi} = 1$$

so there is no need to adjust the lengths of the other axes. Call this adjustment a **time shear**. Since we now have our inertial frame at p_0, we can use 9.2 to flesh this out to an inertial frame there.

At another point p points along the curve, proceed as follows. For $V(4)$, again use $dx^i/d\tau$ (evaluated at p). For the other axes, start by talking $W(1)$, $W(2)$, and $W(3)$ to be the parallel translates of the $V(i)$ along C. These may not be orthogonal to $V(4)$, although they are orthogonal to each other (since parallel translation preserves orthogonality). To fix this, use the same time shearing trick as above to obtain the $V(i)$ at p. Note that the spatial coordinates have not changed in passing from $W(i)$ to $V(i)$—all that is changed are the time-coordinates. Now again use 9.2 to flesh this out to an inertial frame.



By construction, the frame varies smoothly with the point on the curve, so we have a smooth set of coordinates.

Proposition 11.4 (Proper Time is Time in a MCRF)
In a MCRF $\bar{x}$, the x^4-coordinate (time) is proper time τ.

"Proof"
We are assuming starting with some coordinate system x, and then switching to the MCRF $\bar{x}$. Notice that, at the point m,

$$\begin{aligned}\frac{d\bar{x}^4}{d\tau} &= \frac{\partial\bar{x}^4}{\partial x^i}\frac{dx^i}{d\tau} \\ &= \frac{\partial\bar{x}^4}{\partial x^i}V(4)^i \qquad \text{(by definition of } V(4)) \\ &= \overline{V(4)}^4 = 1. \quad \text{(since } V(4) \text{ has coordinates (0,0,0,1) in the barred system)}\end{aligned}$$

In other words, the time coordinate $\bar{x}^4$ is moving at a rate of one unit per unit of proper time τ. Therefore, they must agree.

A particular (and interesting) case of this is the following, for special relativity.

Proposition 11.5 (In SR, Proper Time = Time in the Moving Frame)
In SR, the proper time of a particle moving with a constant velocity v is the t-coordinate of the Lorentz frame moving with the particle.

Proof

$$\tau = \frac{s}{c} = \frac{1}{c}\int\sqrt{-g_{ij}\frac{dx^i dx^j}{dt\ dt}}\ dt.$$

The curve C has parametrization $(vt, 0, 0, t)$ (we are assuming here movement in the x^1-direction), and $g_{**} = \text{diag}\ (1, 1, 1, -c^2)$. Therefore, the above integral boils down to

$$\tau = \frac{1}{c}\int\sqrt{-(v^2-c^2)}\ dt$$
$$= \frac{1}{c}\int c\sqrt{1-v^2/c^2}\ dt$$
$$= t\sqrt{1-v^2/c^2}\ .$$

But, by the (inverse)Lorentz transformations:

$$t = \frac{\bar{t} + v\bar{x}/c^2}{\sqrt{1-v^2/c^2}} = \frac{\bar{t}}{\sqrt{1-v^2/c^2}}\text{, since } \bar{x} = 0 \text{ for the particle.}$$

Thus,

$$\bar{t} = t\sqrt{1-v^2/c^2} = \tau,$$

as required.

Definition 11.6 Let C be the world line of a particle in a Minkowskian manifold M. Its **four velocity** is defined by

$$u^i = \frac{dx^i}{d\tau}\ .$$

Note By the proof of Proposition 10.3, we have

$$\langle u,\ u\rangle = \left|\left|\frac{dx^i}{d\tau}\right|\right|^2 = -c^2.$$

In other words, *four-velocity is timelike and of constant magnitude.*

Example 11.7 Four Velocity in SR
Let us calculate the four-velocity of a particle moving with uniform velocity v with respect to some (Lorentz) coordinate system in Minkowski space $M = E_4$. Thus, x^i are the coordinates of the particle at proper time τ. We need to calculate the partial derivatives $dx^i/d\tau$, and we use the chain rule:

$$\frac{dx^i}{d\tau} = \frac{dx^i}{dx^4}\frac{dx^4}{d\tau}$$
$$= v^i\frac{dx^4}{d\tau} \qquad \text{for } i = 1, 2, 3$$

since x^4 is time in the unbarred system. Thus, we need to know $dx^4/d\tau$. (In the barred system, this is just 1, but this is the unbarred system...) Since $\bar{x}^4 = \tau$, we use the (inverse) Lorentz transformation:

$$x^4 = \frac{\bar{x}^4 + v\bar{x}^1/c^2}{\sqrt{1 - v^2/c^2}} ,$$

assuming for the moment that $v = (v, 0, 0)$. However, in the frame of the particle, $\bar{x}^1 = 0$, and $\bar{x}^4 = \tau$, giving

$$x^4 = \frac{\tau}{\sqrt{1 - v^2/c^2}} ,$$

and hence

$$\frac{dx^4}{d\tau} = \frac{1}{\sqrt{1 - v^2/c^2}} .$$

Now, using the more general boost transformations, we can show that this is true regardless of the direction of v if we replace v^2 in the formula by $(v^1)^2 + (v^2)^2 + (v^3)^2$ (the square magnitude of v). Thus we find

$$u^i = \frac{dx^i}{d\tau} = v^i \frac{dx^4}{d\tau} = \frac{v^i}{\sqrt{1 - v^2/c^2}} \qquad (i = 1, 2, 3)$$

and

$$u^4 = \frac{dx^4}{d\tau} = \frac{1}{\sqrt{1 - v^2/c^2}} .$$

Hence the coordinates of four velocity in the unbarred system are given by

Four Velocity in SR

$$u^* = (v^1, v^2, v^3, 1)/\sqrt{1-v^2/c^2}$$

We can now calculate $\langle u, u \rangle$ directly as

$$\langle u, u \rangle = u^* \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -c^2 \end{bmatrix} u^T$$

$$= \frac{v^2-c^2}{\sqrt{1-v^2/c^2}} = -c^2.$$

Special Relativistic Dynamics
If a contravariant "force" field $\boldsymbol{F}$ (such as an electromagnetic force) acts on a particle, then its motion behaves in accordance with

$$m_0\frac{d\boldsymbol{u}}{d\tau} = \boldsymbol{F},$$

where m_0 is a scalar the **rest mass** of the particle; its mass as measured in its rest (that is, comoving) frame.

We use the four velocity to get **four momentum,** defined by

$$p^i = m_0 u^i.$$

Its **energy** is given by the fourth coordinate, and is defined as

$$E = c^2 p^4 = \frac{m_0 c^2}{\sqrt{1-v^2/c^2}}\ .$$

Note that, for small v,

$$E = m_0(1-v^2/c^2)^{-1/2} \approx m_0 c^2 + \tfrac{1}{2} m_0 v^2.$$

In the eyes of a the comoving frame, $v = 0$, so that

$$E = m_0 c^2.$$

This is called the **rest** energy of the particle, since it is the energy in a comoving frame.

Definitions 11.7 If M is any locally Minkowskian 4-manifold and C is a timelike path or spacelike (thought of as the world line of a particle), we can define its **four momentum** as its four velocity times its rest mass, where the rest mass is the mass as measured in any MCRF.

Exercise Set 11
1. What are the coordinates of four velocity in a comoving frame? Use the result to check that $\langle u, u\rangle = -c^2$ directly in an MCRF.
2. What can you say about $\langle \boldsymbol{p}, \boldsymbol{p}\rangle$, where $\boldsymbol{p}$ is the 4-momentum?
3. Is energy a scalar? Explain
4. Look up and obtain the classical Lorentz transformations for velocity. (We have kind of done it already.)
5. Look up and obtain the classical Lorentz transformations for mass.

---

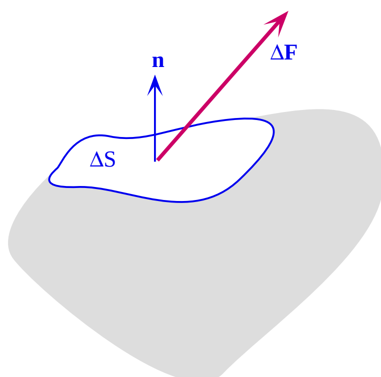
12. The Stress Tensor and the Relativistic Stress-Energy Tensor

Classical Stress Tensor

The classical stress tensor measures the internal forces that parts of a medium—such as a fluid or the interior of a star—exert on other parts (even though there may be zero net force at each point, as in the case of a fluid at equilibrium).

This is how you measure it: if ΔS is an element of surface in the medium, then the material on each side of this interface is exerting a force on the other side. (In equilibrium, these forces will cancel out.) To measure it physically, pretend that all the material on one side is suddenly removed. Then the force that would be experienced is the force we are talking about. (It can go in either direction: for a liquid under pressure, it will push out, whereas for a stretched medium, it will tend to contract in.)

To make this more precise, we need to distinguish one side of the surface ΔS from the other, and for this we replace ΔS by a vector $\boldsymbol{\Delta S} = \boldsymbol{n}\Delta S$ whose magnitude is ΔS and whose direction is normal to the surface element ($\boldsymbol{n}$ is a unit normal). Then associated to that surface element there is a vector $\boldsymbol{\Delta F}$ representing the force exerted by the fluid *behind* the surface (on the side opposite the direction of the vector $\boldsymbol{\Delta S}$) on the fluid on the other side of the interface.



Since we this force is clearly effected by the magnitude ΔS, we use instead the force per unit area (the *pressure)* given by

$$\boldsymbol{T}(\boldsymbol{n}) = \lim_{\Delta S \to 0} \frac{\boldsymbol{\Delta F}}{\Delta S} .$$

Note that $\boldsymbol{T}$ is a function only of the direction $\boldsymbol{n}$ (as well as being a function of the point in space at which we are doing the slicing of the medium); specifying $\boldsymbol{n}$ at some point in turn specifies an interface (the surface normal to $\boldsymbol{n}$ at that point) and hence we can define $\boldsymbol{T}$.

One last adjustment: why insist that $\boldsymbol{n}$ be a *unit* vector? If we replace $\boldsymbol{n}$ by an arbitrary vector $\boldsymbol{v}$, still normal to $\boldsymbol{\Delta S}$, we can still define $\boldsymbol{T}(\boldsymbol{v})$ by multiplying $\boldsymbol{T}(\boldsymbol{v}/|\boldsymbol{v}|)$ by $|\boldsymbol{v}|$. Thus, for general $\boldsymbol{v}$ normal to $\boldsymbol{\Delta S}$,

$$\boldsymbol{T}(\boldsymbol{v}) = \lim_{\Delta S \to 0} \frac{\boldsymbol{\Delta F}}{\Delta S} .|\boldsymbol{v}|.$$

We now find that $\boldsymbol{T}$ has this rather interesting algebraic property: $\boldsymbol{T}$ operates on vector fields to give new vector fields. If is were a linear operator, it would therefore be a tensor, and we could define its coordinates by

$$T^{ab} = \boldsymbol{T}(\boldsymbol{e}_b)^a,$$

the a-component of stress on the b-interface. In fact, we have

Proposition 12.1 (Linearity and Symmetry)
$\boldsymbol{T}$ is a symmetric tensor, called the **stress tensor**

Sketch of Proof To show it's a tensor, we need to establish linearity. By definition, we already have

$$\boldsymbol{T}(\lambda\boldsymbol{v}) = \lambda\boldsymbol{T}(\boldsymbol{v})$$

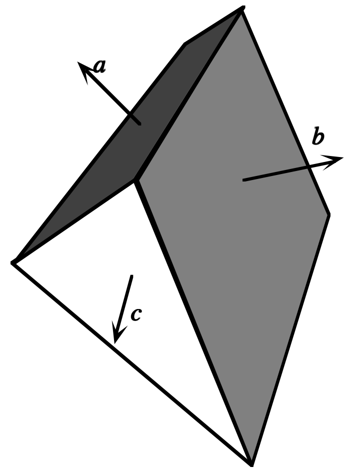
for any constant λ. Thus, all we need show is that if $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are three vectors whose sum is zero, that

$$\boldsymbol{T}(\boldsymbol{a}) + \boldsymbol{T}(\boldsymbol{b}) + \boldsymbol{T}(\boldsymbol{c}) = 0.$$

Further, we can assume that the first two vectors are at right angles.[1] Since all three vectors are coplanar, we can think of the three forces above as stresses on the faces of a prism as shown in the figure. (Note that the vector $\boldsymbol{c}$ in the figure is meant to be at right angles to the bottom face, pointing downwards, and coplanar with $\boldsymbol{a}$ and $\boldsymbol{b}$.)

[1] If we have proved the additive property for vectors at right-angles, then we have it for all pairs:

$$\begin{aligned}\boldsymbol{P}(\boldsymbol{a}+\boldsymbol{b}) &= \boldsymbol{P}(\boldsymbol{a}^{perp}+k\boldsymbol{b} + \boldsymbol{b}) \text{ for some constnat } k, \text{ where } \boldsymbol{a}^{perp} \text{ is orthogonal to } \boldsymbol{a}\\ &= \boldsymbol{P}(\boldsymbol{a}^{perp} + (k+1)\boldsymbol{b})\\ &= \boldsymbol{P}(\boldsymbol{a}^{perp}) + (k+1)\boldsymbol{P}(\boldsymbol{b}) \quad \text{by hypothesized linearlity}\\ &= \boldsymbol{P}(\boldsymbol{a}) + \boldsymbol{P}(\boldsymbol{b})\end{aligned}$$



If we take a prism that is much longer that it is thick, we can ignore the forces on the ends. It now follows from Pythagoras' theorem that the areas in this prism are proportional to the three vectors. Therefore, multiplying through by a constant reduces the equation to one about actual forces on the faces of the prism, with $\boldsymbol{T}(\boldsymbol{a}) + \boldsymbol{T}(\boldsymbol{b}) + \boldsymbol{T}(\boldsymbol{c})$ the resultant force (since the lengths of the vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are equal to the respective areas). If this force was not zero, then there would be a resultant force $\boldsymbol{F}$ on the prism, and hence an acceleration of its material. The trouble is, if we cut all the areas in half by scaling all linear dimensions down by a factor α, then the areas scale down by a factor of α^2, whereas the volume (and hence mass) scales down by a factor α^3. In other words,

$$\boldsymbol{T}(\alpha^2\boldsymbol{a}) + \boldsymbol{T}(\alpha^2\boldsymbol{b}) + \boldsymbol{T}(\alpha^2\boldsymbol{c}) = \alpha^2\boldsymbol{F}$$

is the resultant force on the scaled version of the prism, whereas its mass is proportional to α^3. Thus its acceleration is proportional to $1/\alpha$ (using Newton's law). This means that, as α becomes small (and hence the prism shrinks**) the acceleration becomes infinite—hardly a likely proposition.

The argument that the resulting tensor is symmetric follows by a similar argument applied to a square prism; the asymmetry results in a rotational force on the prism, and its angular acceleration would become infinite if this were not zero. ❇

The Relativistic Stress-Energy Tensor

Now we would like to generalize the stress tensor to 4-dimensional space. First we set the scenario for our discussion:

We now work in a 4-manifold M whose metric has signature $(1, 1, 1, -1)$.

We have already call such a manifold a **locally Minkowskian** 4-manifold. (All this means is that we are using different units for time in our MCRFs.)

** Honey, I shrank the prism.

Example 12.2
Let M be Minkowski space, where one unit of time is defined to be the time it takes light to travel one spacial unit. (For example, if units are measured in meters, then a unit of time would be approximately 0.000 000 003 3 seconds.) In these units, $c = 1$, so the metric does have this form.

The use of MCRFs allows us to define new physical scalar fields as follows: If we are, say, in the interior of a star (which we think of as a continuous fluid) we can measure the pressure at a point by hitching a ride on a small solid object moving with the fluid. Since this should be a smooth function, we consider the pressure, so measured, to be a scalar field. Mathematically, we are defining the field by specifying its value on MCRFs. Note that there is a question here about ambiguity: MCRFs are not unique except for the time direction: once we have specified the time direction, the other axes might be “spinning” about the path—it is hard to prescribe directions for the remaining axes in a convoluted twisting path. However, since we are using a small solid object, we can choose directions for the other axes at proper time 0, and then the "solid-ness" hypothesis guarantees (by definition of solid-ness!) that the other axes remain at right angles; that is, that we continue to have an MCRF aftger allplying a time-shear as in Lecture 11.

Now, we would like to measure a 4-space analogue of the force exerted across a plane, except this time, the only way we can divide 4-space is by using a *hyperplane*; the span of three vectors in some frame of reference. Thus, we seek a 4-dimensional analogue of the quantity $\boldsymbol{n}\Delta S$. By coincidence, we just happen to have such a gizmo lying around: the Levi-Civita tensor. Namely, if $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ are any three vectors in 4-space, then we can define an analogue of $\boldsymbol{n}\Delta S$ to be $\varepsilon_{ijkl}a^i b^k c^l$, where ε is the Levi-Civita tensor. (See the exercises.)

Next, we want to measure stress by generalizing the classical formula

$$\text{stress} = \boldsymbol{T}(\boldsymbol{n}) = \frac{\Delta \boldsymbol{F}}{\Delta S}$$

for such a surface element. Hopefully, the space-coordinates of the stress will continue to measure force. The first step is to get rid of all mention of unit vectors—they just dont arise in Minkowski space (recall that vectors can be time-like, space-like, or null...). We first rewrite the formula as

$$\boldsymbol{T}(\boldsymbol{n}\Delta S) = \Delta \boldsymbol{F},$$

the total force across the area element ΔS. Now multiply both sides by a time coordinate increment:

$$\boldsymbol{T}(\boldsymbol{n}\Delta S \Delta x^4) = \Delta \boldsymbol{F} \Delta x^4 = \Delta \boldsymbol{p},$$

where $\boldsymbol{p}$ is the 3-momentum.[#] This is fine for three of the dimensions. In other words,

$$\boldsymbol{T}(\boldsymbol{n}\Delta V) = \Delta\boldsymbol{p}, \qquad \text{or} \qquad \boldsymbol{T}(\boldsymbol{n}) = \frac{\Delta\boldsymbol{p}}{\Delta V} \qquad \ldots \qquad \text{(I)}$$

where V is volume in Euclidean 4-space, and where we take the limit as $\Delta V \longrightarrow 0$.

But now, generalizing to 4-space is forced on us: first replace momentum by the 4-momentum $\boldsymbol{P}$, and then, noting that $\boldsymbol{n}\Delta S\Delta x^4$ is a 3-volume element in 4-space (because it is a product of three coordinate invrements), replace it by the correct analogue for Minkowski space,

$$(\Delta V)_i = \varepsilon_{ijkl}\Delta x^j \Delta y^k \Delta z^l,$$

getting

$$\boldsymbol{T}(\Delta\boldsymbol{V}) = \Delta\boldsymbol{P},$$

where $\Delta\boldsymbol{P}$ is 4-momentum exerted on the positive side of the 3-volume ΔV by the opposite side. But, there is a catch: the quantity $\Delta\boldsymbol{V}$ has to be really small (in terms of coordinates) for this formula to be accurate. Thus, we rewrite the above formula in differential form:

$$\boldsymbol{T}(d\boldsymbol{V}) = \boldsymbol{T}(\boldsymbol{n}dV) = d\boldsymbol{P}$$

This describes $\boldsymbol{T}$ as a function which converts the covariant vector $d\boldsymbol{V}$ into a contravariant field ($\boldsymbol{P}$), and thus suggests a type (2, 0) tensor. To get an honest tensor, we must define $\boldsymbol{T}$ on *arbitrary* covariant vectors (not just those of the form $\Delta\boldsymbol{V}$). However, every covariant vector Y_* defines a 3-volume as follows.

Recall that a one-form at a point p is a linear real-valued function on the tangent space T_p at that point. If it is non-zero, then its kernel, which consists of all vectors which map to zero, is a three-dimensional subspace of T_p. This describes (locally) a (hyper-)surface. (In the special case that the one-form is the gradient of a scalar field ϕ, that surface coincides with the level surface of ϕ passing through p.) If we choose a basis $\{v, w, u\}$ for this subspace of T_p, then we can recover the one-form at p (up to constant multiples) by forming $\varepsilon_{ijkl}v^j w^k u^l$.[†] This gives us the following formal definition of the tensor $\boldsymbol{T}$ at a point:

Definition 12.3 (The Stress Energy Tensor) For an arbitrary covariant vector $\boldsymbol{Y}$ at p, we choose a basis $\{v, w, u\}$ for its kernel, scaled so that $Y_i = \varepsilon_{ijkl}v^j w^k u^l$, and define $\boldsymbol{T}(\boldsymbol{Y})$ as follows: Form the parallelepiped $\Delta V = \{r_1 v + r_r w + r_3 u \mid 0 \le r_i \le 1\}$ in the tangent space, and compute the total 4-momentum $\boldsymbol{P}$ exerted on the positive side of the volume

---

[#] Classically, force is the time rate of change of momentum.

[†] Indeed, all you have to check is that the covariant vector $\varepsilon_{ijkl}v^j w^k u^l$ has u, w, and v in its kernel. But that is immediate from the anti-symmetric properties of the Levi-Civita tensor.

element ΔV on the positive side[2] of this volume element by the negative side. Call this quantity $\boldsymbol{P}(1)$. More generally, define

$$\boldsymbol{P}(\varepsilon) = \text{total 4-momentum } \boldsymbol{P} \text{ exerted on the positive side of the (scaled) volume element } \varepsilon^3 \Delta V \text{ on the positive side of this volume element by the negative side.}$$

Then define

$$\boldsymbol{T}(\boldsymbol{Y}) = \lim_{\varepsilon \longrightarrow 0} \frac{\boldsymbol{P}(\varepsilon)}{\varepsilon^3} .$$

Note Of course, physical reality intervenes here: how do you measure momentum across volume elements in the tangent space? Well, you do all your measurements in a locally intertial frame. Proposition 8.5 then guarnatees that you get the same physical measurements near the origin regardless of the inertial frame you use (we are, after all, letting ε approach zero).

To evaluate its coordinates on an orthonormal (Lorentz) frame, we define

$$T^{ab} = \boldsymbol{T}(\boldsymbol{e_b})^a,$$

so that we can take u, w, and v to be the other three basis vectors. This permits us to use the simpler formula (I) to obtain the coordinates. Of interest to us is a more usable form—in terms of quantities that can be measured. For this, we need to move into an MCRF, and look at an example.

Note It can be shown, by an argument similar to the one we used at the beginning of this section, that T is a symmetric tensor.

Definition 11.4 Classically, a fluid has **no viscosity** if its stress tensor is diagonal in an MCFR (viscosity is a force parallel to the interfaces).

Thus, for a viscosity-free fluid, the top 3×3 portion of matrix should be diagonal in all MCRFs (independent of spacial axes). This forces it to be a constant multiple of the identity (since every vector is an eigenvector implies that all the eigenvalues are equal…). This single eigenvector measures the force at right-angles to the interface, and is called the **pressure**, p.

Question Why the pressure?
Answer Let us calculate T^{11} (in an MCRF). It is given by

$$T^{11} = \boldsymbol{T}(\boldsymbol{e}_1)^1 = \frac{\Delta \boldsymbol{P}^1}{\Delta V} ,$$

---

[2] "positive" being given by the direction of $\boldsymbol{Y}$

where the 4-momentum is obtained physically by suddenly removing all material on the positive side of the x^1-axis, and then measuring 1-component of the 4-momentum at the origin. Since we are in an MCRF, we can use the SR 4-velocity formula:

$$\boldsymbol{P} = m_0(v^1, v^2, v^3, 1)/\sqrt{1-v^2/c^2} .$$

At the instant the material is removed, the velocity is zero in the MCRF, so

$$\boldsymbol{P}(t=0) = m_0(0, 0, 0, 1).$$

After an interval Δt in this frame, the 4-momentum changes to

$$\boldsymbol{P}(t=1) = m_0(\Delta v, 0, 0, 1)/\sqrt{1-(\Delta v)^2/c^2} ,$$

since there is no viscosity (we must take $\Delta v^2 = \Delta v^3 = 0$ or else we will get off-diagonal spatial terms in the stress tensor). Thus,

$$\Delta\boldsymbol{P} = m_0(\Delta v, 0, 0, 1)/\sqrt{1-(\Delta v)^2/c^2} .$$

This gives

$$(\Delta\boldsymbol{P})^1 = \frac{m_0\Delta v}{\sqrt{1-(\Delta v)^2/c^2}} = m\Delta v \qquad (m \text{ is the apparent mass})$$
$$= \Delta(mv)$$
$$= \text{Change of measured momentum}$$

Thus,

$$\frac{\Delta\boldsymbol{P}^1}{\Delta V} = \frac{\Delta(mv)}{\Delta y\Delta z\Delta t} = \frac{\Delta F}{\Delta y\Delta z} \qquad \text{(force = rate of change of momentum)}$$

and we interpret force per unit area as pressure.

What about the fourth coordinate? The 4th coordinate of the 4-momentum is the energy. A component of the form $T^{4,1}$ measures energy-flow per unit time, per unit area, in the direction of the x^1-axis. In a **perfect fluid**, we insist that, in addition to zero viscosity, we also have zero heat conduction. This forces all these off-diagonal terms to be zero as well. Finally, T^{44} measures energy per unit volume in the direction of the time-axis. This is the **total energy density**, ρ. Think of is as the “energy being transferred from the past to the future.”

This gives the stress-energy tensor in a comoving frame of the particle as

$$T = \begin{bmatrix} p & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & \rho \end{bmatrix}.$$

What about other frames? To do this, all we need do is express T as a tensor whose coordinates in a the comoving frame happen to be as above. To help us, we recall from above that the coordinates of the 4-velocity in the particle's frame are

$u = [0 \quad 0 \;\; 0 \;\; 1]$ (just set $\mathbf{v} = 0$ in the 4-velocity).

(It follows that

$$u^a u^b = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

in this frame.) We can use that, together with the metric tensor

$$g = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

to express T as

$$T^{ab} = (\rho + p)u^a u^b + p g^{ab}.$$

Stress-Energy Tensor for Perfect Fluid
The stress-energy tensor of a perfect fluid (no viscosity and no heat conduction) is given at a point $m \in M$ by

$$T^{ab} = (\rho + p)u^a u^b + p g^{ab},$$

where:
ρ is the mass energy density of the fluid
p is the pressure
u^i is its 4-velocity

Note that the scalars in this definition are their physical magnitudes as measured in a MCRF.

Conservation Laws

Let us now go back to the general formulation of T (not necessarily in a perfect fluid), work in an MCRF, and calculate some covariant derivatives of T. Consider a little cube with each side of length Δl, oriented along the axes (in the MCRF). We saw above that T^{41} measures energy-flow per unit time, per unit area, in the direction of the x^1-axis. Thus, the quantity

$$T^{41}{}_{,1}\Delta l$$

is the approximate increase of that quantity (per unit area per unit time). Thus, the increase of *outflowing* energy per unit time in the little cube is

$$T^{41}{}_{,1}(\Delta l)^3$$

due to energy flow in the x^1-direction. Adding the corresponding quantities for the other directions gives

$$-\frac{\partial E}{\partial t} = T^{41}{}_{,1}(\Delta l)^3 + T^{42}{}_{,2}(\Delta l)^3 + T^{43}{}_{,3}(\Delta l)^3,$$

which is an expression of the law of *conservation of energy*. Since E is given by $T^{44}(\Delta l)^3$, and $t = x^4$, we therefore get

$$-T^{44}{}_{,4}(\Delta l)^3 = (T^{41}{}_{,1} + T^{42}{}_{,2} + T^{43}{}_{,3})(\Delta l)^3,$$

giving

$$T^{41}{}_{,1} + T^{42}{}_{,2} + T^{43}{}_{,3} + T^{44}{}_{,4} = 0$$

A similar argument using each of the three components of momentum instead of energy now gives us the law of conservation of momentum (3 coordinates):

$$T^{a1}{}_{,1} + T^{a2}{}_{,2} + T^{a3}{}_{,3} + T^{a4}{}_{,4} = 0$$

for $a = 1, 2, 3$. Combining all of these and reverting to an arbitrary frame now gives us:

Einstein's Conservation Law

$$\nabla.T = 0$$

where $\nabla.T$ is the contravariant vector given by $(\nabla.T)^j = T^{jk}{}_{|k}$.

This law combines both energy conservation and momentum conservation into a single elegant law.

Exercise Set 12
1. If $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ are any three vector fields in locally Minkowskain 4-manifold, show that the field $\varepsilon_{ijkl}a^i b^k c^l$ is orthogonal to $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$.(ε is the Levi-Civita tensor.)

---

13. Three Basic Premises of General Relativity

Spacetime

General relativity postulates that spacetime (the set of all events) is a smooth 4-dimensional Riemannian manifold M, where points are called **events**, with the properties A1-A3 listed below.

A1. Locally, M is Minkowski spacetime (so that special relativity holds locally).

This means that, if we diagonalize the scalar product on the tangent space at any point, we obtain the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

The metric is measurable by clocks and rods.

Before stating the next axiom, we recall some definitions.

Definitions 13.1 Let M satisfy axiom A1. If V^i is a contravariant vector at a point in M, define

$$\|V^i\|^2 = \langle V^i, V^j \rangle = V^i V^j g_{ij}.$$

(Note that we are not defining $\|V^i\|$ here.) We say the vector V^i is

timelike if $\|V^i\|^2 < 0$,

lightlike if $\|V^i\|^2 = 0$,

and **spacelike** if $\|V^i\|^2 > 0$,

Examples 13.2
(a) If a particle moves with constant velocity $\boldsymbol{v}$ in some Lorentz frame, then at time $t = x^4$ its position is

$$\boldsymbol{x} = \boldsymbol{a} + \boldsymbol{v}x^4.$$

Using the local coordinate x^4 as a parameter, we obtain a path in M given by

$$x^i(x^4) = \begin{cases} a^i + v^i x^4 & \text{if } i = 1, 2, 3 \\ x^4 & \text{if } i = 4 \end{cases}$$

so that the tangent vector (velocity) dx^i/dx^4 has coordinates $(v^1, v^2, v^3, 1)$ and hence square magnitude

$$\|(v^1, v^2, v^3, 1)\|^2 = |\boldsymbol{v}|^2 - c^2.$$

It is timelike at sub-light speeds, lightlike at light speed, and spacelike at faster-than-light speeds.
(b) If $\boldsymbol{u}$ is the proper velocity of some particle in locally Minkowskian spacetime, then we saw (normal condition in Section 10) that $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = -c^2 = -1$ in our units.

A2. Freely falling particles move on timelike geodesics of M.

Here, a freely falling particle is one that is effected only by gravity, and recall that a **timelike** geodesic is a geodesic $x^i(t)$ with the property that $\|dx^i/dt\|^2 < 0$ in any paramaterization. (This property is independent of the parameterization—see the exercise set.)

A3 (Strong Equivalence Principle) All physical laws that hold in flat Minkowski space (ie. "special relativity") are expressible in terms of vectors and tensors, and are meaningful in the manifold M, continue to hold in *every* frame (provided we replace derivatives by covariant derivatives).

Note Here are some consequences:
1. No physical laws can use the term "straight line," since that concept has no meaning in M; what's straight in the eyes of one chart is curved in the eyes of another. "Geodesic," on the other hand, does make sense, since it is independent of the choice of coordinates.
2. If we can write down physical laws, such as Maxwell's equations, that work in Minkowski space, then those same laws must work in curved space-time, without the addition of any new terms, such as the curvature tensor. In other words, there can be no form of Maxwell's equations for general curved spacetime that involve the curvature tensor.

An example of such a law is the conservation law, $\nabla.T = 0$, which is thus postulated to hold in all frames.

A Consequence of the Axioms: Forces in Almost Flat Space

Suppose now that the metric in our frame is almost Lorentz, with a slight, not necessarily constant, deviation ϕ from the Minkowski metric, as follows.

$$g_{**} = \begin{bmatrix} 1+2\phi & 0 & 0 & 0 \\ 0 & 1+2\phi & 0 & 0 \\ 0 & 0 & 1+2\phi & 0 \\ 0 & 0 & 0 & -1+2\phi \end{bmatrix} \qquad \ldots\ldots\ldots\ldots$$

(I)

or

$$ds^2 = (1+2\phi)(dx^2 + dy^2 + dz^2) - (1-2\phi)dt^2.$$

Notes

1. We are *not* in an inertial frame (modulo scaling) since ϕ need not be constant, but we are in a frame that is *almost* inertial.

2. The metric g_{**} is obtained from the Minkowski g by adding a small multiple of the identity matrix. We shall see that such a metric does arise, to first order of approximation, as a consequence of Einstein's field equations.

Now, we would like to examine the behavior of a particle falling freely under the influence of this metric. What do the timelike geodesics look like? Let us assume we have a particle falling freely, with 4-momentum $P = m_0U$, where U is its 4-velocity, $dx^i/d\tau$. The paramaterized path $x^i(\tau)$ must satisfy the geodesic equation, by A2. Definition 8.1 gives this as

$$\frac{d^2x^i}{d\tau^2} + \Gamma^{\,i}_{r\,s}\frac{dx^r}{d\tau}\frac{dx^s}{d\tau} = 0.$$

Multiplying both sides by $m_0{}^2$ gives

$$m_0\frac{d^2(m_0x^i)}{d\tau^2} + \Gamma^{\,i}_{r\,s}\frac{d(m_0x^r)}{d\tau}\frac{d(m_0x^s)}{d\tau} = 0,$$

or

$$m_0\frac{dP^i}{d\tau} + \Gamma^{\,i}_{r\,s}P^rP^s = 0 \qquad\qquad \text{(since } P^i = d(m_0x^i/d\tau))$$

where, by the (ordinary) chain rule (note that we are not taking covariant derivatives here... that is, $dP^i/d\tau$ is *not* a vector—see Section 7 on covariant differentiation),

$$\frac{dP^i}{d\tau} = P^i{}_{,k}\frac{dx^k}{d\tau}$$

so that

$$P^i{}_{,k}\frac{dm_0x^k}{d\tau} + \Gamma^{\,i}_{r\,s}P^rP^s = 0,$$

or

$$P^i{}_{,k}P^k + \Gamma_{r\,s}^{\,i}P^rP^s = 0 \quad \ldots\ldots\ldots\ldots \qquad \text{(I)}$$

Now let us do some estimation for *slowly-moving* particles $v << 1$ (the speed of light), where we work in a frame where g has the given form.[‡] First, since the frame is almost inertial (Lorentz), we are close to being in SR, so that

$$P^* \approx m_0U^* = m_0[v^1, v^2, v^3, 1] \qquad \text{(we are taking } c = 1 \text{here)}$$
$$\approx [0, 0, 0, m_0] \qquad \text{(since } v << 1\text{)}$$

(in other words, the frame is almost comoving) Thus (I) reduces to

$$P^i{}_{,4}m_0 + \Gamma_{4\,4}^{\,i}\,m_0{}^2 = 0 \qquad \ldots\ldots\ldots\ldots \qquad \text{(II)}$$

Let us now look at the spatial coordinates, $i = 1, 2, 3$. By definition,

$$\Gamma_{4\,4}^{\,i} = \frac{1}{2}g^{ij}(g_{4j,4} + g_{j4,4} - g_{44,j}).$$

We now evaluate this at a specific coordinate $i = 1, 2$ or 3, where we use the definition of the metric g, recalling that $g^{**} = (g_{**})^{-1}$, and obtain

$$\frac{1}{2}(1+2\phi)^{-1}(0 + 0 - 2\phi_{,i}) \approx \frac{1}{2}(1-2\phi)(-2\phi_{,i}) \approx -\phi_{,i}.$$

[‡] Why don't we work in an inertial frame (the frame of the particle)? Well, in an inertial frame, we adjust the coordinates to make g = diag[1, 1, 1, −1] at the origin of our coordinate system. The first requirement of an inertail frame is that, $\phi(0, 0, 0, 0) = 0$. This you can certainly do, if you like; it doesn't effect the ensuing calculation at all. The next requirement is more serious: that the partial derivatives of the g_{ij} vanish. This would force the geodesics to be uninteresting (straight) at the origin, since the Christoffel symbols vanish, and (II) becomes

$$P^i{}_{|4}P^4 = 0,$$

that is, since $P^4 \approx m_0$ and $P^i{}_{|4} = \frac{d}{dx^4}(m_0v^i)$ = rate of change of momentum, that

rate of change of momentum = 0,

so that the particle is experiencing no force (even though it's in a gravitational field).

Question But what does this mean? What is going on here?

Answer All this is telling us is that an inertial frame in a gravitational field is one in which a particle experiences no force. That is, it is a “freely falling” frame. To experience one, try bungee jumping off the top of a tall building. As you fall, you experience no gravitational force—as though you were in outer space with no gravity present.

This is not, however, the situation we are studying here. We want to be in a frame where the metric is **not** locally constant. so it would defeat the purpose to choose an inertial frame.

(Here and in what follows, we are ignoring terms of order $O(\phi^2)$.) Substituting this information in (II), and using the fact that

$$P^i{}_{,4} = \frac{\partial}{\partial x^4}(m_o v^i),$$

the time-rate of change of momentum, or the "force" as measured in that frame (see the exercise set), we can rewrite (II) as

$$m_0 \frac{\partial}{\partial x^4}(m_o v^i) - m_0{}^2 \phi_{,i} = 0,$$

or

$$\frac{\partial}{\partial x^4}(m_o v^i) - m_0 \phi_{,i} = 0.$$

Thinking of x^4 as time t, and adopting vector notation for three-dimensional objects, we have, in old fashioned 3-vector notation,

$$\frac{\partial}{\partial t}(m_o \boldsymbol{v}) = m_0 \boldsymbol{\nabla}\phi,$$

that is

$$\boldsymbol{F} = m\boldsymbol{\nabla}\phi.$$

This is the Newtonian force experienced by a particle in a force field potential of ϕ. (See the exercise set.) In other words, we have found that we can duplicate, to a good approximation, the physical effects of Newton-like gravitational force from a simple distortion of the metric. In other words—and this is what Einstein realized—gravity is nothing more than the geometry of spacetime; it is not a mysterious "force" at all.

Exercise Set 13

1. Show that, if $x^i = x^i(t)$ has the property that $\|dx^i/dt\|^2 < 0$ for some parameter t, then $\|dx^i/dts\|^2 < 0$ for any other parameter s such that $ds/dt \neq 0$ along the curve. In other words, the property of being timelike does not depend on the choice of paramaterization.

2. What is wrong with the following (slickly worded) argument based on the Strong Equivalence Principle?

I claim that there can be no physical law of the form $A = R$ in curved spacetime, where A is some physical quantity and R is any quantity derived from the curvature tensor. (Since we shall see that Einstein's Field Equations have this form, it would follow from this argument that he was wrong!) Indeed, if the postulated law $A = R$ was true, then in flat spacetime it would reduce to $A = 0$. But then we have a physical law in SR, which must, by the Strong Equivalence Principle, generalize to $A = 0$ in curved spacetime as well. Hence the original law $A = R$ was wrong.

3. Gravity and Antigravity Newton's law of gravity says that a particle of mass M exerts a force on another particle of mass m according to the formula

$$\boldsymbol{F} = -\frac{GMm\boldsymbol{r}}{r^3},$$

where $\boldsymbol{r} = \langle x, y, z\rangle$, $r = |\boldsymbol{r}|$, and G is a constant that depends on the units; if the masses M and m are given in kilograms, then $G \approx 6.67 \times 10^{-11}$, and the resulting force is measured in newtons.[î] (Note that the magnitude of $\boldsymbol{F}$ is proportional to the inverse square of the distance $\boldsymbol{r}$. The negative sign makes the force an attractive one.) Show by direct calculation that

$$\boldsymbol{F} = m\nabla\phi,$$

where

$$\phi = \frac{GM}{r}.$$

Hence write down a metric tensor that would result in an inverse square repelling force ("antigravity").

---

14. The Einstein Field Equations and Derivation of Newton's Law

Einstein's field equations show how the sources of gravitational fields alter the metric. They can actually be motivated by Newton's law for gravitational potential ϕ, with which we begin this discussion.

First, Newton's law postulates the existence of a certain scalar field ϕ, called *gravitational potential* which exerts a force on a unit mass given by

$$\boldsymbol{F} = \nabla\phi \qquad \text{(classical gravitational field)}$$

Further, ϕ satisfies

$$\nabla^2\phi = \nabla.(\nabla\phi) = 4\pi G\rho \qquad \ldots \qquad \text{(I)}$$

Div(gravitational field) = constant × mass density

where ρ is the mass density and G is a constant. (The divergence theorem then gives the more familiar $\boldsymbol{F} = \nabla\phi = GM/r^2$ for a spherical source of mass M—see the exercise set.) In relativity, we need an invariant analogue of (I). First, we generalize the mass density to energy density (recall that energy and mass are interchangeable according to relativity),

[î] A Newton is the force that will cause a 1-kilogram mass to accelerate at 1 m/sec^2.

which in turn is only one of the components of the stress-energy tensor T. Thus we had better use the whole of T.

Question What about the mysterious gravitational potential ϕ?
Answer That is a more subtle issue. Since the second principle of general relativity tells us that particles move along geodesics, we should interpret the gravitational potential as somehow effecting the geodesics. But the most fundamental determinant of geodesics is the underlying metric g. Thus we will generalize ϕ to g. In other words, Einstein replaced a mysterious "force" by a purely geometric quantity. Put another way, gravity is nothing but a distortion of the local geometry in space-time. But we are getting ahead of ourselves...

Finally, we generalize the (second order differential) operator ∇ to some yet-to-be-determined second order differential operator Δ. This allows us to generalize (I) to

$$\Delta(g^{**}) = kT^{**},$$

where k is some constant. In an MCRF, $\Delta(g)$ is some linear combination of $g^{ab}{}_{,ij}$, $g^{ab}{}_{,i}$ and g^{ab}, and must also be symmetric (since T is). Examples of such a tensors are the Ricci tensors R^{ab}, $g^{ab}R$, as well as g^{ab}. Let us take a linear combination as our candidate:

$$R^{ab} + \mu g^{ab}R + \Lambda g^{ab} = kT^{ab} \qquad \ldots \qquad \text{(II)}$$

We now apply the conservation laws $T^{ab}{}_{|b} = 0$, giving

$$(R^{ab} + \mu g^{ab}R)_{|b} = 0 \qquad \ldots \qquad \text{(a)}$$

since $g^{ab}{}_{|b} = 0$ already (Exercise Set 8 #4). But in §9 we also saw that

$$(R^{ab} - \tfrac{1}{2} g^{ab}R)_{|b} = 0, \qquad \ldots \qquad \text{(b)}$$

where the term in parentheses is the Einstein tensor G^{ab}. Calculating (a) $-$ (b), using the product rule for differentiation and the fact that $g^{ab}{}_{|b} = 0$, we find

$$(\mu+\tfrac{1}{2})g^{ab}R_{|b} = 0$$

giving (upon multiplication by g^{**})

$$(\mu+\tfrac{1}{2})R_{|j} = 0$$

which surely implies, in general, that μ must equal $-\frac{1}{2}$. Thus, (II) becomes

$$G^{ab} + \Lambda g^{ab} = kT^{ab}.$$

Finally, the requirement that these equations reduce to Newton's for $v << 1$ tells us that $k = 8\pi$ (discussed below) so that we have

> **Einstein's Field Equations**
>
> $$G^{ab} + \Lambda g^{ab} = 8\pi T^{ab}$$

The constant Λ is called the **cosmological constant**. Einstein at first put $\Lambda = 0$, but later changed his mind when looking at the large scale behavior of the universe. Later still, he changed his mind again, and expressed regret that he had ever come up with it in the first place. The cosmological constant remains a problem child to this day.† We shall set it equal to zero.

Solution of Einstein's Equations for Static Spherically Symmetric Stars

In the case of spherical symmetry, we use polar coordinates (r, θ, ϕ, t) with origin thought of as at the center of the star as our coordinate system (note it is singular there, so in fact this coordinate system does not include the origin) and restrict attention to g of the form

$$g_{**} = \begin{bmatrix} g_{rr} & 0 & 0 & g_{rt} \\ 0 & r^2 & 0 & 0 \\ 0 & 0 & r^2\sin^2\theta & 0 \\ g_{rt} & 0 & 0 & -g_{tt} \end{bmatrix},$$

or

$$ds^2 = 2g_{rt}\, dr\, dt + g_{rr}\, dr^2 + r^2\, d\theta^2 + r^2 \sin\theta\, d\phi^2 - g_t\, dt^2,$$

where each of the coordinates is a function of r and t only. In other words, at any fixed time t, the surfaces $\theta = \text{const}$, $\phi = \text{const}$ and $r = \text{const}$ are all orthogonal. (This causes the zeros to be in the positions shown.)

---

† The requirement that Newton's laws be the limit of general relativity for small v forces lambda to be very small. Setting it equal to zero gives all the correct predictions for the motions of planets to within measurable accuracy. Put another way, if $\Lambda \neq 0$, then experimental data shows that it must be very small indeed. Also, we could take that term over to the right-hand side of the equation and incorporate it into the stress-energy tensor, thus regarding $-\Lambda g^{ab}/8\pi$ as the stress-energy tensor of empty space.
Following is an excerpt from an article in Scientific American (September, 1996, p. 22):

> … Yet the cosmological constant itself is a source of much puzzlement. Indeed, Christopher T. Hill of Fermilab calls it "the biggest problem in all of physics." Current big bang models require that lambda is small or zero, and various observations support that assumption. Hill points out, however, that current particle physics theory predicts a cosmological constant much, much greater—by a factor of at least 10^{52}, large enough to have crunched the universe back down to nothing immediately after the big bang. "Something is happening to suppress this vacuum density," says Alan Guth of MIT, one of the developers of the inflationary theory. Nobody knows, however, what that something is …

Question Explain why the non-zeros terms have the above form.
Answer For motivation, let us first look at the standard metric on a 2-sphere of radius r: (see Example 5.2(d))

$$g_{**} = \begin{bmatrix} r^2 & 0 \\ 0 & r^2\sin^2\theta \end{bmatrix}.$$

If we throw r in as the third coordinate, we could calculate

$$g_{**} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{bmatrix}.$$

Moving into Minkowski space, we have

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 - dt^2 \\ &= dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2) - dt^2, \end{aligned}$$

giving us the metric

$$g_{**} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & r^2 & 0 & 0 \\ 0 & 0 & r^2\sin^2\theta & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

(Minkowski space metrtic in polar coords.)

For the general spherically symmetric stellar medium, we can still define the radial coordinate to make $g_{\theta\theta} = r^2$ (through adjustment by scaling if necessary). Further, we take as the *definition* of spherical symmetry, that the geometry of the surfaces $r = t =$ const. are spherical, thus foring us to have the central 2×2 block.

For **static** spherical symmetry, we also require, among other things, (a) that the geometry be unchanged under time-reversal, and (b) that g be independent of time t. For (a), if we change coordinates using

$$(r, \theta, \phi, t) \longrightarrow (r, \theta, \phi, -t),$$

then the metric remains unchanged; that is, $\bar{g} = g$. But changing coordinates in this way amounts to multiplying on the left and right (we have an order 2 tensor here) by the change-of-coordinates matrix diag $(1, 1, 1, -1)$, giving

$$\bar{g}_{**} = \begin{bmatrix} g_{rr} & 0 & 0 & -g_{rt} \\ 0 & r^2 & 0 & 0 \\ 0 & 0 & r^2\sin^2\theta & 0 \\ -g_{rt} & 0 & 0 & -g_{tt} \end{bmatrix}.$$

Setting $\bar{g} = g$ gives $g_{rt} = 0$. Combining this with (b) results in g of the form

$$g_{**} = \begin{bmatrix} e^{2\Lambda} & 0 & 0 & 0 \\ 0 & r^2 & 0 & 0 \\ 0 & 0 & r^2\sin^2\theta & 0 \\ 0 & 0 & 0 & -e^{2\Phi} \end{bmatrix},$$

where we have introduced the exponentials to fix the signs, and where $\Lambda = \Lambda(r)$, and $\Phi = \Phi(r)$. Using this version of g, we can calculate the Einstein tensor to be (see the exercise set!)

$$G^{**} = \begin{bmatrix} \frac{2}{r}\Phi' e^{-4\Lambda} - \frac{1}{r^2}e^{2\Lambda}(1-e^{-2\Lambda}) & 0 & 0 & 0 \\ 0 & e^{-2\Lambda}[\Phi'' + (\Phi')^2 + \frac{\Phi'}{r} - \Phi'\Lambda' - \frac{\Lambda'}{r}] & 0 & 0 \\ 0 & 0 & \frac{G^{\theta\theta}}{\sin^2\theta} & 0 \\ 0 & 0 & 0 & \frac{1}{r^2}\,e^{-2\Phi}\frac{d}{dr}[r(1-e^{-2\Lambda})] \end{bmatrix}$$

We also need to calculate the stress energy tensor,

$$T^{ab} = (\rho + p)u^a u^b + pg^{ab}.$$

In the static case, there is assumed to be no flow of star material in our frame, so that $u^1 = u^2 = u^3 = 0$. Further, the normal condition for four velocity, $\langle u, u\rangle = -1$, gives

$$[0,\ 0,\ 0,\ u^4]\begin{bmatrix} e^{2\Lambda} & 0 & 0 & 0 \\ 0 & r^2 & 0 & 0 \\ 0 & 0 & r^2\sin^2\theta & 0 \\ 0 & 0 & 0 & -e^{2\Phi} \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 0 \\ u^4 \end{bmatrix} = -1$$

whence

$$u^4 = e^{-\Phi},$$

so that $T^{44} = (\rho+p)e^{-2\Phi} + p(-e^{-2\Phi})$ (note that we are using g^{**} here). Hence,

$$T^{**} = (\rho + p)u^* u^* + pg^{**}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (\rho+p)e^{-2\Phi} \end{bmatrix} + p\begin{bmatrix} e^{2\Lambda} & 0 & 0 & 0 \\ 0 & r^2 & 0 & 0 \\ 0 & 0 & r^2\sin^2\theta & 0 \\ 0 & 0 & 0 & -e^{2\Phi} \end{bmatrix}$$

$$= \begin{bmatrix} pe^{-2\Lambda} & 0 & 0 & 0 \\ 0 & \frac{p}{r^2} & 0 & 0 \\ 0 & 0 & \frac{p}{r^2\sin^2\theta} & 0 \\ 0 & 0 & 0 & \rho e^{-2\Phi} \end{bmatrix}.$$

(a) Equations of Motion $T^{ab}{}_{|b} = 0$

To solve these, we first notice that we are not in an inertial frame (the metric g is not nice at the origin; in fact, nothing is even defined there!) so we need the Christoffel symbols, and use

$$T^{ab}{}_{|b} = \frac{\partial T^{ab}}{\partial x^b} + \Gamma_k{}^a{}_b T^{kb} + \Gamma_b{}^b{}_k T^{ak},$$

where

$$\Gamma_h{}^p{}_k = \tfrac{1}{2} g^{lp}\left(\frac{\partial g_{kl}}{\partial x^h} + \frac{\partial g_{lh}}{\partial x^k} - \frac{\partial g_{hk}}{\partial x^l}\right).$$

Now, lots of the terms in $T^{ab}{}_{|b}$ vanish by symmetry, and the restricted nature of the functions. We shall focus on $a = 1$, the r-coordinate. We have:

$$T^{1b}{}_{|b} = T^{11}{}_{|1} + T^{12}{}_{|2} + T^{13}{}_{|3} + T^{14}{}_{|4},$$

and we calculate these terms one-at-a-time.

$a = 1, b = 1$: $$T^{11}{}_{|1} = \frac{\partial T^{11}}{\partial x^1} + \Gamma_1{}^1{}_1 T^{11} + \Gamma_1{}^1{}_1 T^{11}.$$

To evaluate this, first look at the term $\Gamma_1{}^1{}_1$:

$$\begin{aligned} \Gamma_1{}^1{}_1 &= \tfrac{1}{2} g^{l1}(g_{1l,1} + g_{l1,1} - g_{11,l}) \\ &= \tfrac{1}{2} g^{11}(g_{11,1} + g_{11,1} - g_{11,1}) \qquad \text{(because } g \text{ is diagonal, whence } l = 1) \\ &= \tfrac{1}{2} g^{11}(g_{11,1}) \\ &= \tfrac{1}{2} e^{-2\Lambda} e^{2\Lambda}.2\Lambda'(r) = \Lambda'(r). \end{aligned}$$

Hence,

$$T^{11}{}_{|1} = \frac{dp}{dr} e^{-2\Lambda} + (-\ 2p\Lambda'(r)e^{-2\Lambda}) + 2\Lambda'(r)pe^{-2\Lambda} = \frac{dp}{dr} e^{-2\Lambda}.$$

Now for the next term:

$$\begin{aligned} a = 1,\ b = 2: \quad T^{12}{}_{|2} &= \frac{\partial T^{12}}{\partial x^2} + \Gamma_2{}^1{}_2 T^{22} + \Gamma_2{}^2{}_1 T^{11} \\ &= 0 + \tfrac{1}{2} g^{l1}(g_{2l,2}+g_{l2,2}-g_{22,l})T^{22} + \tfrac{1}{2} g^{l2}(g_{1l,2}+g_{l2,1}-g_{21,l})T^{11} \\ &= \tfrac{1}{2} g^{11}(-g_{22,1})\ T^{22} + \tfrac{1}{2} g^{22}(g_{22,1})T^{11} \\ &= \tfrac{1}{2} e^{-2\Lambda}(-2r)\frac{p}{r^2} + \tfrac{1}{2}\frac{1}{r^2\sin\theta} 2r\sin\theta pe^{-2\Lambda} \\ &= 0. \end{aligned}$$

Similarly (exercise set)

$$T^{13}{}_{|3} = 0.$$

Finally,

$$\begin{aligned} a = 1,\ b = 4: \quad T^{14}{}_{|4} &= \frac{\partial T^{14}}{\partial x^4} + \Gamma_4{}^1{}_4 T^{44} + \Gamma_4{}^4{}_1 T^{11} \\ &= \tfrac{1}{2} g^{11}(-g_{44,1})\ T^{44} + \tfrac{1}{2} g^{44}(g_{44,1})T^{11} \\ &= \tfrac{1}{2} e^{-2\Lambda}(2\Phi'(r)e^{2\Phi})\rho e^{-2\Phi} + \tfrac{1}{2}(-e^{-2\Phi})(-2\Phi'(r)e^{2\Phi})pe^{-2\Lambda} \\ &= e^{-2\Lambda}\Phi'(r)[\rho + p]. \end{aligned}$$

Hence, the conservation equation becomes

$$\begin{aligned} & T^{1a}{}_{|a} = 0 \\ \Leftrightarrow\ & \left(\frac{dp}{dr} + \frac{d\Phi}{dr}(\rho+p)\right) e^{-2\Lambda} = 0 \\ \Leftrightarrow\ & \boxed{\frac{dp}{dr} = -(\rho+p)\frac{d\Phi}{dr}}\ . \end{aligned}$$

This gives the pressure gradient required to keep the plasma static in a star.

Note In classical mechanics, the term on the right has ρ rather than $\rho+p$. Thus, the pressure gradient is larger in relativistic theory than in classical theory. This increased pressure gradient corresponds to greater values for p, and hence bigger values for all the components of T. By Einstein's field equations, this now leads to even greater values of Φ (manifested as gravitational force) thereby causing even larger values of the pressure gradient. If p is large to begin with (big stars) this vicious cycle diverges, ending in the gravitational collapse of a star, leading to neutron stars or, in extreme cases, black holes.

(b) Einstein Field Equations $G^{ab} = 8\pi T^{ab}$

Looking at the (4,4) component first, and substituting from the expressions for G and T, we find

$$\frac{1}{r^2} e^{-2\Phi} \frac{d}{dr} [r(1-e^{-2\Lambda})] = 8\pi\rho e^{-2\Phi}.$$

If we define

$$\tfrac{1}{2} r(1-e^{-2\Lambda}) = m(r),$$

then the equation becomes

$$\frac{1}{r^2} e^{-2\Phi} \frac{dm(r)}{dr} = 4\pi\rho e^{-2\Phi},$$

or

$$\boxed{\frac{dm(r)}{dr} = 4\pi r^2 \rho} \qquad \ldots\ldots \quad \text{(I)}$$

This looks like an equation for classical mass, since classically,

$$M(R) = \iint_0^R 4\pi r^2 \rho(r)\ dr$$

where the integrand is the mass of a shell whose thickness is dr. Thus,

$$\frac{dM(R)}{dr} = 4\pi^2 \rho(r).$$

Here, ρ is energy density, and by our choice of units, energy is equal to rest mass, so we interpret $m(r)$ as the total mass of the star enclosed by a sphere of radius r.

Now look at the (1, 1) component:

$$\frac{2}{r} \Phi' e^{-4\Lambda} - \frac{1}{r^2} (1-e^{-2\Lambda}) = 8\pi p e^{-2\Lambda}$$

$$\Rightarrow \quad \frac{2}{r} \Phi' - \frac{e^{2\Lambda}}{r^2} (1-e^{-2\Lambda}) = 8\pi p e^{2\Lambda}$$

$$\Rightarrow \quad 2r\Phi' - e^{2\Lambda}(1-e^{-2\Lambda}) = 8\pi r^2 p e^{2\Lambda}$$

$$\Rightarrow \quad \Phi' = e^{2\Lambda}\frac{(1-e^{-2\Lambda}) + 8\pi r^2 p}{2r} .$$

In the expression for m, solve for $e^{2\Lambda}$ to get

$$e^{2\Lambda} = \frac{1}{1-2m/r} ,$$

giving

$$\frac{d\Phi}{dr} = \frac{8\pi r^2 p + 2m/r}{2r(1-2m/r)} ,$$

or

$$\boxed{\frac{d\Phi}{dr} = \frac{4\pi r^3 p + m}{r(r-2m)}} \qquad \cdots\cdots\cdots \qquad \text{(II)}$$

It can be checked using the Bianchi identities that we in fact get no additional information from the (2,2) and (3,3) components, so we ignore them.

Consequences of the Field Equations: Outside the Star

Outside the star we take $p = \rho \approx 0$, and $m(r) = M$, the total stellar mass, getting

$$\text{(I):} \frac{dm}{dr} = 0 \qquad \text{(nothing new, since } m = M = \text{constant)}$$

$$\text{(II):} \frac{d\Phi}{dr} = \frac{M}{r(r-2M)} ,$$

which is a separable first order differential equation with solution

$$e^{2\Phi} = 1 - \frac{2M}{r} .$$

if we impose the boundary condition $\Phi \longrightarrow 0$ as $r \longrightarrow +\infty$. (See the exercise Set).

Recalling from the definition of m that

$$e^{2\Lambda} = \frac{1}{1-2M/r} ,$$

we can now express the metric outside a star as follows:

Schwarzschild Metric

$$g_{**} = \begin{bmatrix} \frac{1}{1-2M/r} & 0 & 0 & 0 \\ 0 & r^2 & 0 & 0 \\ 0 & 0 & r^2\sin^2\theta & 0 \\ 0 & 0 & 0 & -(1-2M/r) \end{bmatrix}$$

In the exercise set, you will see how this leads to Newton's Law of Gravity.

Exercise Set 14

1. Use $\nabla^2\phi = 4\pi G\rho$ and the divergence theorem to deduce Newton's law $\mathbf{\nabla}\phi = GM/r^2$ for a spherical mass of uniform density ρ.

2. Calculate the Einstein tensor for the metric $g = \text{diag}(e^{2\Lambda}, r^2, r^2\sin\theta, -e^{2\Phi})$, and verify that it agrees with that in the notes.

3. Referring to the notes above, show that $T^{13}{}_{|3} = 0$.

4. Show that $T^{i4}{}_{|4} = 0$ for $i = 2, 3, 4$.

5. If we impose the condition that, far from the star, spacetime is flat, show that this is equivalent to saying that $\lim_{r\to+\infty} \Phi(r) = \lim_{r\to+\infty} \Lambda(r) = 0$. Hence obtain the formula

$$e^{2\Phi} = 1 - \frac{2M}{r} .$$

6. A Derivation of Newton's Law of Gravity

(a) Show that, at a large distance R from a static stable star, the Schwarzschild metric can be approximated as

$$g_{**} \approx \begin{bmatrix} 1+2M/R & 0 & 0 & 0 \\ 0 & R^2 & 0 & 0 \\ 0 & 0 & R^2\sin^2\theta & 0 \\ 0 & 0 & 0 & -(1-2M/R) \end{bmatrix}.$$

(b) (Schutz, p. 272 #9) Define a new coordinate $\bar{R}$ by $R = \bar{R}(1+M/\bar{R})^2$, and deduce that, in terms of the new coordinates (ignoring terms of order $1/R^2$)

$$g_{**} \approx \begin{bmatrix} 1+2M/\bar{R} & 0 & 0 & 0 \\ 0 & \bar{R}^2(1+2M/\bar{R})^2 & 0 & 0 \\ 0 & 0 & \bar{R}^2(1+2M/\bar{R})^2\sin^2\theta & 0 \\ 0 & 0 & 0 & -(1-2M/\bar{R}) \end{bmatrix}.$$

(c) Now convert to Cartesian coordinates, (x, y, z, t) to obtain

$$g_{**} \approx \begin{bmatrix} 1+2M/\bar{R} & 0 & 0 & 0 \\ 0 & 1+2M/\bar{R} & 0 & 0 \\ 0 & 0 & 1+2M/R & 0 \\ 0 & 0 & 0 & -(1-2M/\bar{R}) \end{bmatrix}.$$

(d) Now refer to the last formula in Section 10, and obtain Newton's Law of Gravity. To how many kilograms does one unit of M correspond?

---

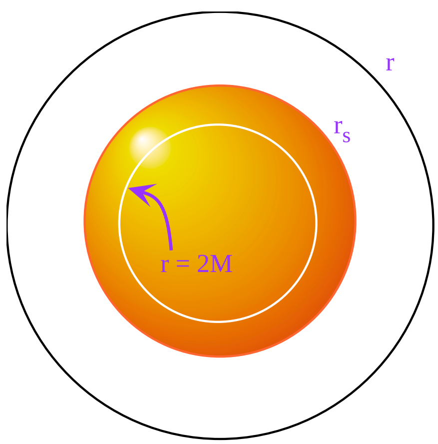
15. The Schwarzschild Metric and Event Horizons

We saw that the metric outside a spherically symmetric static stable star (Schwarzschild metric) is given by

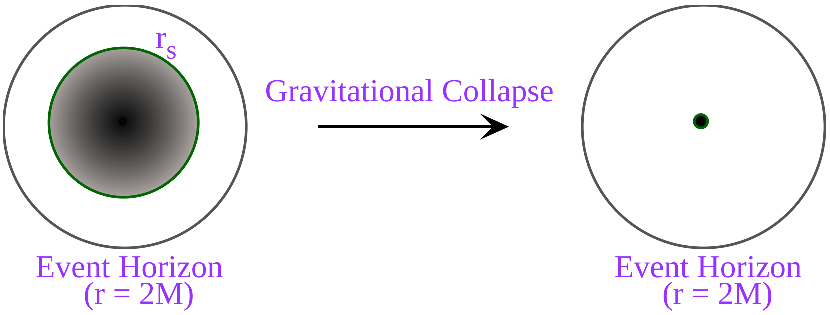
$$ds^2 = \frac{1}{1-2M/r}dr^2 + r^2 d\Omega^2 - (1-2M/r)dt^2,$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta\, d\phi^2$. We see immediately that something strange happens when $2M = r$, and we look at two cases.

Case 1 (Not-So-Dense Stars) Radius of the star, $r_s > 2M$.
If we recall that the Schwarzschild metric is only valid for outside a star; that is, $r > r_s$, we find that $r > 2M$ as well, and so $1-2M/r$ is positive, and never zero. (If $r \le 2M$, we are inside the star, and the Schwarzschild metric no longer applies.)



Case 2 (Extremely Dense Stars) Radius of the star, $r_s < 2M$.
Here, two things happen: First, as a consequence of the equations of motion, it can be shown that in fact the pressure inside the star is unable to hold up against the gravitational forces, and the star collapses (see the next section) overwhelming even the quantum mechanical forces. In fact, it collapses to a singularity, a point with infinite density and no physical dimension, a **black hole**. For such objects, we have two distinct regions, defined by $r > 2M$ and $r < 2M$, separated by the **event horizon**, $r = 2M$, where the metric goes infinite.



Particles Falling Inwards
Suppose a particle is falling radially inwards. Let us see how long, on the particle's clock (proper time), it takes to reach the event horizon. Out approach will be as follows:

(1) Use the principle that the path is a geodesic in space time.
(2) Deduce information about $dr/d\tau$.
(3) Integrate $d\tau$ to see how long it takes.

Recall first the geodesic equation for such a particle,

$$P^i{}_{|k}P^k + \Gamma_{r\,s}^{\,i} P^r P^s = 0.$$

We saw in the derivation (look back) that it came from the equation

$$m_0 \frac{dP^i}{d\tau} + \Gamma_{r\,s}^{\,i} P^r P^s = 0 \qquad \ldots\ldots\ldots \qquad \text{(I)}$$

There is a covariant version of this:

$$m_0 \frac{dP_s}{d\tau} - \Gamma_{r\,s}^{\,i} P^r P_i = 0.$$

Derivation This is obtained as follows:

Multiplying both sides of (I) by g_{ia} gives

$$m_0 \frac{dP^i}{d\tau} g_{ia} + \Gamma_{r\,s}^{\,i} P^r P^s g_{ia} = 0,$$

or

$$m_0 \frac{d(g_{ia}P^i)}{d\tau} - m_0 \frac{dg_{ia}}{d\tau} P^i + \Gamma_{r\,s}^{\,i} P^r P^s g_{ia} = 0$$

$$m_0 \frac{d(P_a)}{d\tau} - m_0 \frac{dg_{ia}}{d\tau} P^i + \Gamma_{r\,s}^{\,i} P^r P^s g_{ia} = 0$$

$$m_0\frac{d(P_a)}{d\tau} - m_0P^i(\frac{Dg_{ia}}{D\tau} + \Gamma_{r\,i}^{\,k}g_{ka}\frac{dx^r}{d\tau} + \Gamma_{a\,r}^{\,k}g_{ik}\frac{dx^r}{d\tau}) + \Gamma_{r\,s}^{\,i}P^rP^sg_{ia} = 0$$

$\|$ (by definition of $Dg_{ia}/D\tau$)

0

$$m_0\frac{d(P_a)}{d\tau} - \Gamma_{r\,i}^{\,k}P^iP^rg_{ka} - \Gamma_{a\,r}^{\,k}P^iP^rg_{ik} + \Gamma_{r\,s}^{\,i}P^rP^sg_{ia} = 0,$$

leaving

$$m_0\frac{d(P_a)}{d\tau} \qquad - \Gamma_{a\,r}^{\,k}P^iP^rg_{ik} \qquad = 0,$$

or

$$m_0\frac{d(P_a)}{d\tau} - \Gamma_{a\,r}^{\,k}P_kP^r = 0,$$

which is the claimed covariant version.

Now take this covariant version and write out the Christoffel symbols:

$$m_0\frac{dP_s}{d\tau} = \Gamma_{r\,s}^{\,i}P^rP_i$$

$$m_0\frac{dP_s}{d\tau} = \frac{1}{2}g^{ik}(g_{rk,s} + g_{ks,r} - g_{sr,k})\,P^rP_i$$

$$m_0\frac{dP_s}{d\tau} = \frac{1}{2}(g_{rk,s} + g_{ks,r} - g_{sr,k})\,P^rP^k$$

But the sum of the second and third terms in parentheses is skew-symmetric in r and k, whereas the term outside is symmetric in them. This results in them canceling when we sum over repeated indices. Thus, we are left with

$$m_0\frac{dP_s}{d\tau} = \frac{1}{2}g_{rk,s}P^rP^k \quad \ldots\ldots\ldots \qquad \text{(II)}$$

But by spherical symmetry, g is independent of x^i if $i = 2, 3, 4$. Therefore $g_{rk,s} = 0$ unless $s = 1$. This means that P_2, P_3 and P_4 are constant along the trajectory. Since P_4 is constant, we define

$$E = -P_4/m_0,$$

another constant.

Question What is the meaning of E?
Answer Recall that the fourth coordinate of four momentum is the energy. Suppose the particle starts at rest at $r = \infty$ and then falls inward. Since space is flat there, and the particle is at rest, we have

$$P^* = [0, 0, 0, m_0] \qquad \text{(fourth coordinate is rest energy} = m_0\text{)}$$

(which corresponds to $P_* = [0, 0, 0, -m_0]$, since $P^* = P_* g^{**}$). Thus, $E = -P_4/m_0 = 1$, the rest energy per unit mass.

As the particle moves radially inwards, $P^2 = P^3 = 0$. What about P^1? Now we know the first coordinate of the contravariant momentum is given by

$$P^1 = m_0 \frac{dr}{d\tau} \qquad \text{(by definition, } P^i = m_0 \frac{dx^i}{d\tau}\text{, and } x^1 = r\text{)}$$

Thus, using the metric to get the fourth contravariant coordinate,

$$P^* = (m_0 \frac{dr}{d\tau}, 0, 0, m_0 E(1-2M/r)^{-1})$$

we now invoke the normalization condition $\{u, u\rangle = -1$, whence $\langle P, P\rangle = -m_0{}^2$, so that

$$-m_0{}^2 = m_0{}^2 \left(\frac{dr}{d\tau}\right)^2 (1-2M/r)^{-1} - m_0{}^2 E^2 (1-2M/r)^{-1},$$

giving

$$\left(\frac{dr}{d\tau}\right)^2 = E^2 - 1 + 2M/r,$$

which is the next step in our quest:

$$d\tau = -\frac{dr}{\sqrt{E^2 - 1 + 2M/r}},$$

where we have introduced the negative sign since r is a decreasing function of τ. Therefore, the total time elapsed is

$$\mathcal{T} = \int_R^{2M} -\frac{dr}{\sqrt{E^2 - 1 + 2M/r}},$$

which, though improper, is finite.* This is the time it takes, on the hapless victim's clock, to reach the event horizon.

---

* See Schutz, p. 289.

Now let's recalculate this from the point of view of an observer who is stationary with respect to the star. That is, let us use the coordinate x^4 as time t. How is it related to proper time? Well, the four velocity tells how:

$$V^4 =_{\text{defn}} \frac{dx^4}{d\tau} = \frac{dt}{d\tau} .$$

We can get V^4 from the formula for P^* (and divide by m_0) so that

$$dt = V^4 d\tau = E(1-2M/r)^{-1}\ d\tau$$

giving a total time of

$$T = \int_R^{2M} - \frac{dr}{E(1-2M/r)\sqrt{E^2-1+2M/r}} .$$

This integral *diverges!* So, in the eyes of an outside observer, it takes that particle infinitely long to get there!

Inside the Event Horizon—A Dialogue

Tortoise: I seem to recall that the metric for a stationary observer (situated inside the event horizon) is still given by the Schwarzschild metric

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

Achilles: Indeed, but notice that now the coefficient of dr^2 is negative, while that of dt^2 is positive. What could that signify (if anything)?

Tortoise: Let us do a little thought experiment. If we are unfortunate(?) enough to be there watching a particle follow either a null or timelike world line, then, with respect to any parameter (such as τ) we must have $dr/d\tau \neq 0$. In other words, r must always change with the parameter!

Achilles: So you mean nothing can sit still. Why so?

Tortoise: Simple. First: for any world line, the vector $dx^i/d\tau$ is non-zero, (or else it would not be a path at all!) so *some* coordinate must be non-zero. But now if we calculate $\|dx^i/d\tau\|^2$ using the signature $(-, +, +, +)$ we get

$$- \text{something} \times \left(\frac{dr}{d\tau}\right)^2 + \text{something} \times \text{the others},$$

so the only way the answer can come out zero or negative is if the *first* coordinate ($dr/d\tau$) is non-zero.

Achilles: I think I see your reasoning... we *could* get a null path if all the coordinates were zero, but that just can't happen in a path! So you mean to tell me that this is true even of light beams. Mmm.... So you're telling me that r must change along the world line of any particle or photon! But that begs a question, since r is always changing with τ, does it increase or decrease with proper time τ?

Tortoise: To tell you the truth, I looked in the Green Book, and all it said was the “obviously” r must decrease with τ, but I couldn't see anything obvious about that.

Achilles: Well, let *me* try a thought experiment for a change. If you accept for the moment the claim that a particle fired toward the black hole will move so as to decrease r, then there is at least one direction for which $dr/d\tau < 0$. Now imagine a particle being fired in any direction. Since $dr/d\tau$ will be a continuous function of the angle in which the particle is fired, we conclude that it must *always* be negative.

Tortoise: Nice try, my friend, but you are being too hasty (as usual). That argument can work against you: suppose that a particle fired *away* from the black hole will move (initially at least) so as to *increase* r, then your argument proves that r *increases* no matter what direction the particle is fired. Back to the drawing board.

Achilles: I see your point...

Tortoise (interrupting)**:** Not only that. You might recall from Lecture 38 (or thereabouts) that the 4-velocity of as radially moving particle in free-fall is given by

$$V^* = (\frac{dr}{d\tau}, 0, 0, E(1-2M/r)^{-1}),$$

so that the fourth coordinate, $dt/d\tau = E(1-2M/r)^{-1}$, is negative inside the horizon. Therefore, proper time moves in the *opposite* direction to coordinate time!

Achilles: Now I'm really confused. Does this mean that for r to decrease with coordinate time, it has to *increase* with proper time?

Tortoise: Yes. So you were (as usual) totally wrong in your reason for asserting that $dr/d\tau$ is negative for an inward falling particle.

Achilles: OK. So now the burden of proof is on you! You have to explain what the hell is going on.

Tortoise: That's easy. You might dimly recall the equation

$$\left(\frac{dr}{d\tau}\right)^2 = E^2 - 1 + 2M/r$$

on p. 112 of those excellent differential geometry notes, wherein we saw that we can take $E = 1$ for a particle starting at rest far from the black hole. In other words,

$$\left(\frac{dr}{d\tau}\right)^2 = 2M/r.$$

Notice that this is constant and never zero, so that $dr/d\tau$ can never change sign during the trajectory of the particle, even as (in its comoving frame) it passes through the event horizon. Therefore, since r was initially decreasing with τ (outside, in "normal" space-time), it must continue to do so *throughout its world line.* In other words, *photons that originate outside the horizon can never escape in their comoving frame.* Now (and here's the catch), since there are *some* particles whose world-lines have the property that the arc-length parameter (proper time) decreases with increasing r, and since r is the unique coordinate in the stationary frame that plays the formal role of time, and further since, in any frame, all world lines must move in the same direction with respect to the local time coordinate (meaning r) as their parameter increases, it follows that *all* world lines must decrease r with increasing proper time. Ergo, Achilles, r must always decrease with increasing proper time τ.

Of course, a consequence of all of this is that no light, communication, or any physical object, can escape from within the event horizon. They are all doomed to fall into the singularity.

Achilles: But what about the stationary observer?

Tortoise: Interesting point...the quantity $dt/d\tau = E(1-2M/r)^{-1}$ is negative, meaning proper time goes in the opposite direction to coordinate time and also becomes large as it approaches the horizon, so it would seem to the stationary observer inside the event horizon that things do move out toward the horizon, but take infinitely long to get there. There is a catch, however, there can be no "stationary observer" according to the above analysis...

Achilles: Oh.

Exercise Set 15

1. Verify that the integral for the infalling particle diverges the case $E = 1$.

2. Mini-Black Holes How heavy is a black hole with event horizon of radius one meter? [Hint: Recall that the "M" corresponds to $G\times$total mass.]

3. Calculate the Riemann coordinates of curvature tensor R_{abcd} at the event horizon. $r = 2M$.

16. White Dwarfs, Neutron Stars and Black Holes—Guest Lecture by Gregory C. Levine

I Introduction

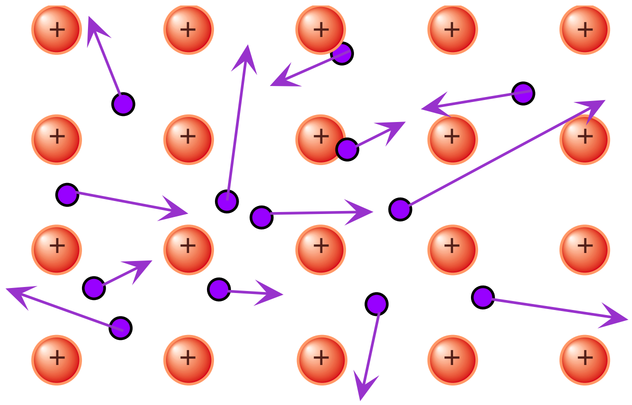
In this section we will look at the physical mechanisms responsible for the formation compact stellar objects. Compact objects such as white dwarf stars, neutron stars, and ultimately black holes, represent the final state of a star's evolution. Stars are born in gaseous nebulae in which clouds of hydrogen coalesce becoming highly compressed and heated through the gravitational interaction. At a temperature of about 10^7 K, a nuclear reaction begins converting hydrogen into the next heavier element, helium, and releasing a large quantity of electromagnetic energy (light). The helium accumulates at the center of the star and eventually becomes compressed and heated enough (10^8 K) to initiate nuclear fusion of helium into heavier elements.

So far, the star is held in "near-equilibrium" by the countervailing forces of gravity, which compresses the star, and pressure from the vast electromagnetic energy produced during nuclear fusion, which tends to make it expand. However, as the star burns hotter and ignites heavier elements which accumulate in the core, electromagnetic pressure becomes less and less effective against gravitational collapse. In most stars, this becomes a serious problem when the core has reached the carbon rich phase but the temperature is still insufficient to fuse carbon into iron. Even if a star *has* reached sufficient temperature to create iron, no other nuclear fusion reactions producing heavier elements are exothermic and the star has exhausted its nuclear fuel. Without electromagnetic energy to hold the core up, one would think that the core would become unstable and begin to collapse---but another mechanism intervenes.

II The Electron Gas

But there is another "force" that holds the core up; now we will turn to a study of this force and how the balance between this force and gravity lead to the various stellar compact objects: white dwarfs, neutron stars and black holes.

The stabilizing force that keeps the stellar core from collapsing operates at terrestrial scales as well. All solid matter resists compression and we will trace the origin of this behavior in a material that turns out to most resemble a stellar compact object: ordinary metal. Although metal is "hard" by human standards, it is to some degree elastic---capable of stretching and compression. Metals all have a similar atomic structure. Positively charged metal ion cores form a regular crystalline lattice and negatively charged valence electrons form a kind of gas that uniformly permeates the lattice.



Suprisingly, the bulk properties of the metal such as heat capacity, compressibility, and thermal conductivity are almost exclusively properties of the electron gas and not the underlying framework of the metal ion cores. We will begin by studying the properties of an electron gas alone and then see if it is possible to justify such a simple model for a metal (or a star).

To proceed, two very important principles from Quantum Mechanics need to be introduced:

Pauli Exclusion Principle: Electrons cannot be in the same quantum state. For our purposes, this will effectively mean that electrons cannot be at the same point in space.
Heisenberg Uncertainty Principle: A quantum particle has no precise position, x, or momentum, p. However, the uncertainties in the outcome of experiment aimed at simultaneously determining both quantities is constrained in the following way. Upon repeated measurements, the "spread" in momentum, Δp, of a particle absolutely confined to a region in space of size Δx, is constrained by

$$\Delta x \Delta p \geq \frac{\hbar}{2}$$

where $\hbar \approx 6.6\times 10^{-34}$ Joule-sec is a fundamental constant of nature (the Planck constant).

Here is how these two laws act together to give one of the familiar properties of metals. The Pauli Exclusion Principle tends to make electrons stay as far apart as possible. Each of N electrons confined in a box of volume R^3 will typically have R^3/N space of its own. Therefore, the average interparticle spacing is $a_0 = R/N^{1/3}$. (The situation is actually a bit more complicated than this *link*). Since the electrons are spatially confined within a region of linear size a_0, the uncertainty in momentum is $\Delta p \approx \hbar/a_0$. The precise meaning of Δp^2 is the *variance* of a large set of measurements of momentum. Denoting average by angle brackets,

$$\Delta p^2 = \langle (p - \langle p \rangle)^2 \rangle = \langle p^2 \rangle - \langle p \rangle^2.$$

Therefore, the average value of p^2 must be greater than or equal to Δp^2.

Based on these results, let us calculate how the energy of an electron gas depends upon the size of the box containing it. The kinetic energy of a particle of mass m and speed v is

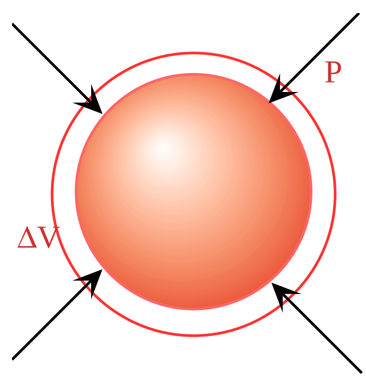
$$\varepsilon = \frac{1}{2}mv^2 = \frac{p^2}{2m}$$

Now, taking the minimum value of momentum, $p^2 \approx \Delta p^2 \approx \hbar^2/a_0^2$, we arrive at the energy, $\varepsilon = \hbar^2/m_e a_0^2$, for a single electron of mass m_e. The total kinetic energy of N electrons is then $E_e = N\varepsilon$. Finally, putting in the dependence of a_0 on N and the system size, R, we get for E_e,

$$E_e \approx \frac{\hbar^2}{m_e R^2} N^{5/3}.$$

As the system size R is reduced, the energy increases. Even though the electrons do not interact with one another, there is an effective repulsive force resisting compression. The origin of this force is the uncertainty principle! (neglecting e-e interactions and neglecting temperature.)

Let us test out this model by calculating the compressibility of metal. Consider a metal block that undergoes a small change in volume, ΔV, due to an applied pressure P.



The **bulk modulus**, B, is defined as the constant of proportionality between the applied pressure and the fractional volume change.

$$P = B\frac{\Delta V}{V}.$$

The outward pressure (towards positive R) exerted by the electron gas is defined in the usual way in terms of a derivative of the total energy of the system:

$$P = \frac{F}{A} = -\frac{1}{A}\frac{\partial E_e}{\partial R}$$

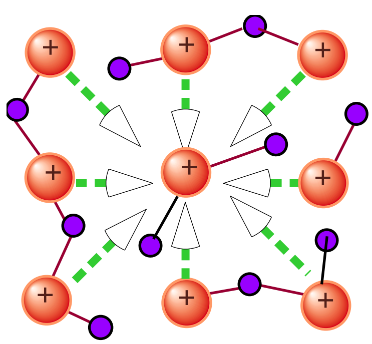
The bulk modulus is then defined as

$$B = V\frac{\partial P}{\partial V} = -\frac{5}{9}\frac{\hbar^2}{m_e}\left(\frac{N}{V}\right)^{5/3} \approx 10^{-10} - 10^{-11}\ \text{N/m}^2.$$

(We've taken the volume per electron to be 1 nm^3.) The values of B for Steel and Aluminum are $B_{\text{steel}} \approx 6\times10^{-10}$ N/m^2 and $B_{\text{Al}} \approx 2\times10^{-10}$ N/m^2. It is hard to imagine that this excellent agreement in magnitude is wholly fortuitous (it is not). Having seen that the Heisenberg uncertainty principle is the underlying physics behind the rigidity of metal, we will now see that it is also physical mechanism that keeps stars from collapsing under their own weight.

III Compact Objects

A star can only be in a condition of static equilibrium if there is some force to counteract the compressive force of gravity. In large stars this countervailing force is the radiation pressure from thermally excited atoms emitting light. But in a white dwarf star, the force counteracting gravity has its origin in the uncertainty principle, as it did in a metal. The elements making up the star (mostly iron) exist in a completely ionized state because of the high temperatures. One can think of the star as a gas of positive charge atomic nuclei and negative charge electrons. Each metal nucleus is a few thousand times heavier than the set of electrons that were attached to it, so the nuclei (and not the electrons) are responsible for the sizable gravitational force holding the star together. The electrons are strongly *electrostatically* bound to core of the star and therefore coexist in the same volume as the nuclear core---gravity pulling the nuclei together and the uncertainty principle effectively pushing the electrons apart.



We will proceed in the same way as in the calculation of the bulk modulus by finding an expression for the total energy and taking its derivative with respect to R to find the effective force.

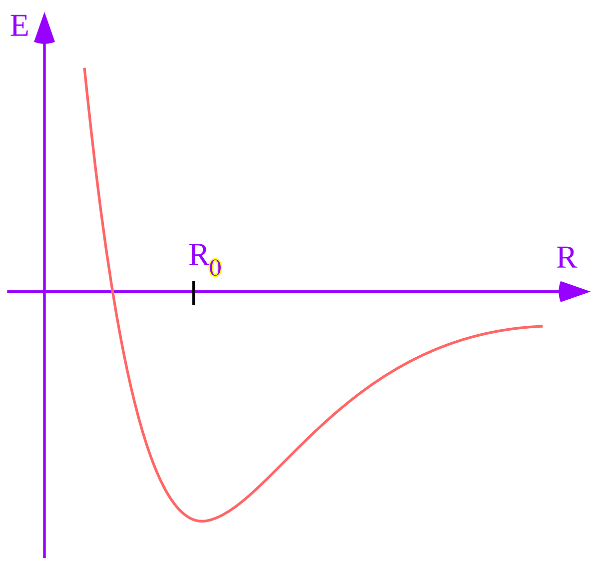
The gravitational potential energy of sphere of mass M and radius R is approximately

$$E_g \approx -\frac{GM^2}{R}$$

where G ≈ 7 × 10^{-11} Nm^2/kg^2 is the gravitational constant. (The exact result has a coeffient of order unity in front; we are doing only "order-of-magnitude" calculations and ignoring such factors.) The negative sign means that the force of gravity is *attractive*---energy decreases with decreasing R. We would like to express E_g in terms of N, like E_e---this will make the resulting expressions easier to adapt to neutron stars later on. The mass M of the star is the collective mass of the nucleons, to an excellent approximation. As you may know from chemisty, the number of nucleons (protons and neutrons) is roughly double the number of electrons, for light elements. If μ is the average number of nucleons per electron, for the heavier elements making up the star, The mass of the star is expressed as $M{=}\mu m_n N$. Putting the expressions for the electron kinetic energy and the gravitational potential energy together, we get the total energy E:

$$E = E_e + E_g \approx \frac{\hbar^2 N^{5/3}}{m_e R^2} - \frac{G\mu^2 m_n{}^2 N^2}{R} .$$

The graph of the function $E(R)$



reveals that there is a radius at which the energy is minimum---that is to say, a radius R_0 where the force $F = -\partial E/\partial R$ is zero and the star is in mechanical equilibrium. A rough calculation of R_0 gives:

$$R_0 = \frac{\hbar^2 N^{-1/3}}{Gm_e\mu^2 m_n{}^2} \approx 10^7 \text{ m} = 10{,}000 \text{ km}.$$

where we have used $N \approx 10^{57}$, a reasonable value for a star such as our sun. R_0 corresponds to a star that is a little bigger than earth---a reasonable estimate for a white dwarf star! The mass density ρ may also be calculated assuming the radius R_0: $\rho \approx 10^9$ kg/m^3 = $10^5 \times$ density of steel. On the average, the electrons are much closer to the nuclei in the white dwarf than they are in ordinary matter.

Under some circumstances, the star can collapse to an object even more compact than a white dwarf---a neutron star. The Special Theory of Relatvity plays an important role in this further collapse. If we calculate the kinetic energy of the most energetic electrons in the white dwarf, we get:

$$\varepsilon \approx \frac{\hbar^2}{m_e a_0{}^2} = \frac{\hbar^2}{m_e}\left(\frac{\mathrm{N}}{\mathrm{R}_0}\right)^{2/3} \approx 100^{-14} \text{ Joules.}$$

This energy is actually quite close to the rest mass energy of the electron itself, $m_e c^2 = 10^{-13}$ Joules. Recall that the expression for the kinetic energy, $\varepsilon = p^2/2m$, is only a nonrelativistic approximation. Rest mass energy is a scalar formed from the product

$$p^\mu p_\mu = \frac{\varepsilon^2}{c^2} - p^2 = (mc)^2.$$

The exact expression for the energy ε of a relativistic particle is then:

$$\varepsilon = \sqrt{(pc)^2 + (mc^2)^2} = mc^2 + \frac{p^2}{2m} + \text{terms of order} \geq \left(\frac{p}{mc}\right)^4 .$$

When $p \approx mc$ (or, equivalently, when $p^2/m \approx mc^2$ as above) the higher order terms cannot be neglected.

Since the full expression for ε is unwieldy for our simple approximation schemes, we will look at the *extreme* relativistic limit, $p >> mc$. In this case, $\varepsilon \approx pc$. This limit is effectively the limit for extremely massive stars, where the huge compressive force of gravity will force the electrons to have compensatingly high kinetic energies and enter the extreme relativistic regime.

The different form for the energy of the electrons (now *linear* rather than quadratic in p) will have dramatic consequences for the stability equation for the radius R_0 derived earlier. The calculation proceeds as before; according to the uncertainty principle the estimate for the momentum of an electron within the star is

$$p = \frac{\hbar}{a_0} = \frac{\hbar N^{1/3}}{R} .$$

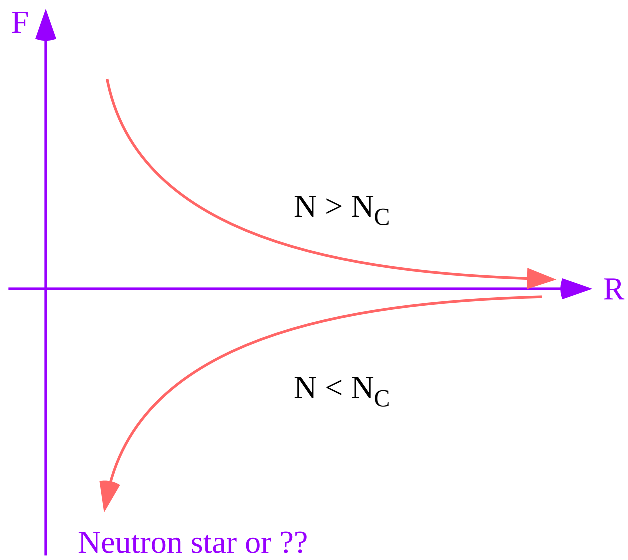
Therefore, the total electron energy is given by

$$E_e \approx N\varepsilon \approx Npc \approx \frac{\hbar c N^{4/3}}{R}$$

The same expression as before for E_g results in the following expression for the total energy:

$$E + E_e + E_g \approx \frac{\hbar c N^{4/3}}{R} - \frac{G\mu^2 m_n{}^2 N^2}{R} .$$

The energy E(R) has a completely different behavior than in the nonrelativistic case. If we look at the force $F = -\partial E/\partial R$ it is just equal to E/R. If the total energy is positive, the force always induces expansion; if the total energy is negative, the force always induces compression. Thus, if the total energy E is negative, the star will continue to collapse (with an ever increasing inward force) unless some other force intervenes. These behaviors are suggested in the figure below.



The expression for total energy tells us that the critical value of N (denoted by N_C) for which the energy crosses over to negative value is

$$N_C = \frac{1}{\mu^3}\left(\frac{\hbar c}{G m_n{}^2}\right)^{3/2} .$$

This is conventionally written in terms of a critical mass for a star, M_C, that separates the two behaviors: expansion or collapse. The critical mass is

$$M_C = \mu N_c m_n = \frac{1}{\mu^2 m_n{}^2}\left(\frac{\hbar c}{G}\right)^{3/2} .$$

If $M > M_C$, the star will continue to collapse and its electrons will be pushed closer and closer to the nuclei. At some point, a *nuclear* reaction begins to occur in which electrons and protons combine to form neutrons (and neutrinos which are nearly massless and noninteracting). A sufficiently dense star is unstable against such an interaction and all electrons and protons are converted to neutrons leaving behind a chargeless and nonluminous star: a neutron star.

You may be wondering: what holds the neutron star up? Neutrons are chargeless and the nuclear force between neutrons (and protons) is only attractive, so what keeps the neutron star from further collapse? Just as with electrons, neutrons obey the Pauli Exclusion Principle. Consequently, they avoid one another when they are confined and have a sizable kinetic energy due to the uncertainty principle. If the neutrons are nonrelativistic, the previous calculation for the radius of the white dwarf star will work just the same, with the replacement $m_e \longrightarrow m_n$. This change reduces the radius R_0 of the neutron star by a factor of ≈2000 (the ratio of m_n to m_e) and $R_0 \approx 10$ km. One of these would comfortably fit on Long Island but would produce somewhat disruptive effects.

Finally, if the neutron star is massive enough to make its neutrons relativistic, continued collapse is possible if the total energy is negative, as before in the white dwarf case. The expression for the critical mass M_C is easily adapted to neutrons by setting $\mu = 1$. Since $\mu \approx 2$ for a white dwarf, we would expect that a star about four times more massive than a white dwarf is susceptible to unlimited collapse. No known laws of physics are capable of interrupting the collapse of a neutron star. In a sense, the laws of physics leave the door open for the formation of stellar black holes.

References and Suggested Further Reading

(Listed in the rough order reflecting the degree to which they were used.)

Bernard F. Schutz, *A First Course in General Relativity* (Cambridge University Press, 1986

David Lovelock and Hanno Rund, *Tensors, Differential Forms, and Variational Principles* (Dover, 1989)

Charles E. Weatherburn, *An Introduction to Riemannian Geometry and the Tensor Calculus* (Cambridge University Press, 1963)

Charles W. Misner, Kip S. Thorne and John A. Wheeler, *Gravitation* (W.H. Freeman, 1973)

Keith R. Symon, *Mechanics* (3rd. Ed. Addison Wesley)