Introduction to Differential Geometry

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 600

Subsection 1

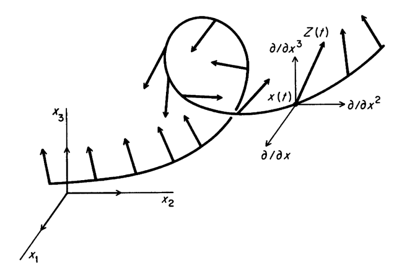
Differentiation of Vector Fields along Curves in $\mathbb{R}^n$

Vector Fields Along Curves in $\mathbb{R}^n$

- Let C be a curve in $\mathbb{R}^n$ given by

$$x(t) = (x^1(t), \ldots, x^n(t)), \quad a < t < b.$$

- Suppose that $Z(t) = Z_{x(t)}$ is a vector field defined along C.

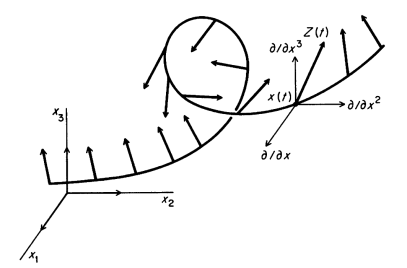


- Thus, to each $t \in (a, b)$, is assigned a vector

$$Z(t) = \sum a^i(t) \left(\frac{\partial}{\partial x^i} \right)_{x(t)} \in T_{x(t)}(\mathbb{R}^n).$$

Vector Fields Along Curves in $\mathbb{R}^n$ (Cont'd)

- We will suppose Z to be of class C^1 at least.



- This means that the components $a^i(t)$ are continuously differentiable functions of t on the interval (a, b).
- The velocity vector of the (parametrized) curve itself is an example.
- In this case, we have

$$a^i(t) = \dot{x}^i(t).$$

Difference of Two Vectors

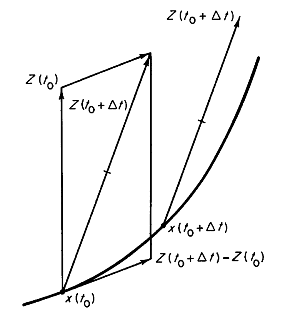
- We define a derivative, or rate of change, of $Z(t)$ with respect to t.
- It will be denoted $\dot{Z}(t)$ or $\frac{dZ}{dt}$.
- It will again be a vector field along the curve.
- In general, neither $Z(t)$ nor its derivative are tangent to the curve.
- In $\mathbb{R}^n$ we have a natural parallelism (or natural isomorphism) of $T_p(\mathbb{R}^n)$ and $T_q(\mathbb{R}^n)$, for any distinct $p, q \in \mathbb{R}^n$.
- So we are able to give meaning to

$$Z(t_0 + \Delta t) - Z(t_0),$$

the difference of a vector in $T_{x(t_0+\Delta t)}(\mathbb{R}^n)$ and a vector in $T_{x(t_0)}(\mathbb{R}^n)$.

Difference of Two Vectors (Cont'd)

- The difference $Z(t_0 + \Delta t) - Z(t_0)$ of a vector in $T_{x(t_0+\Delta t)}(\mathbb{R}^n)$ and a vector in $T_{x(t_0)}(\mathbb{R}^n)$.



- For definiteness we suppose $Z(t_0 + \Delta t)$ moved to or identified with the corresponding vector in $T_{x(t_0)}(\mathbb{R}^n)$.
- Further, we suppose that the subtraction is performed there.

Definition of the Derivative

- This identification allows us to define the differential quotient

$$\frac{1}{\Delta t}[Z(t_0+\Delta t)-Z(t_0)]=\sum_{i=1}^{n}\frac{a^i(t_0+\Delta t)-a^i(t_0)}{\Delta t}\left(\frac{\partial}{\partial x^i}\right)_{x(t_0)}.$$

- We have to justify this equality.
- Suppose we write vectors in terms of the basis

$$\frac{\partial}{\partial x^1},\ldots,\frac{\partial}{\partial x^n}.$$

- This is a field of *parallel* frames on $\mathbb{R}^n$.
- Thus, vectors at distinct points, say $Z(t_0+\Delta t)$ and $Z(t_0)$, are parallel if and only if they have the same components.

Definition of the Derivative (Cont'd)

- Passing to the limit as $\Delta t \to 0$ gives the definition

$$\dot{Z}(t_0) = \left(\frac{dZ}{dt}\right)_{t_0} = \sum \dot{a}^i(t_0)\left(\frac{\partial}{\partial x^i}\right)_{x(t_0)} \in T_{x(t_0)}(\mathbb{R}^n).$$

Remark: We look at a useful consequence of this formula.

Suppose we introduce a new parameter on the curve, say s, by

$$t = f(s), \text{ with } t_0 = f(s_0).$$

Then

$$\left(\frac{dZ}{ds}\right)_{s_0} = \left(\frac{dt}{ds}\right)_{s_0}\left(\frac{dZ}{dt}\right)_{t_0}.$$

Here $(\frac{dt}{ds})_{s_0}$ is a scalar, whereas the other terms are vectors.

Example

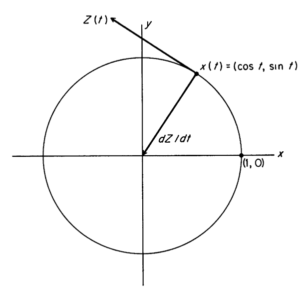
- Consider the curve $x(t) = (\cos t, \sin t)$, a unit circle in $\mathbb{R}^2$.
- Suppose

$$Z(t) = -\sin t \left(\frac{\partial}{\partial x}\right) + \cos t \left(\frac{\partial}{\partial y}\right).$$

- This is the velocity vector of the point which traces out the circle.
- Then

$$\frac{dZ}{dt} = -\cos t \left(\frac{\partial}{\partial x}\right) - \sin t \left(\frac{\partial}{\partial y}\right).$$

- This is a vector at $x(t) = (\cos t, \sin t)$ which has constant length $+1$ and points toward the origin.



Constant or Parallel Vector Fields

Definition

A vector field $Z(t)$ is **constant** or **parallel** along the curve $x(t)$ if and only if $\frac{dZ}{dt} = 0$ for all t.

- Suppose that $Z_1(t)$ and $Z_2(t)$ are vector fields of the above type defined along the same curve C.
- Let $f(t)$ be a differentiable function of t on $a < t < b$.
- Then $f(t)Z(t)$ and $Z_1(t) + Z_2(t)$ are vector fields along C.

Properties of Derivatives

- We have the following easy consequences of the definition.
- For sums
$$\frac{d}{dt}(Z_1(t)+Z_2(t))=\frac{dZ_1}{dt}+\frac{dZ_2}{dt}.$$
- For products by a differentiable function,
$$\frac{d}{dt}(f(t)Z(t))=\frac{df}{dt}Z(t)+f(t)\frac{dZ}{dt}.$$
- For inner products,
$$\frac{d}{dt}(Z_1(t),Z_2(t))=\left(\frac{dZ_1}{dt},Z_2(t)\right)+\left(Z_1(t),\frac{dZ_2}{dt}\right),$$
where (Z_1,Z_2) is the standard inner product in $\mathbb{R}^n$.

Using Other Field Frames

- We sometimes find it convenient to use a field of frames, other than the natural one, say
$$F_1(t), \ldots, F_n(t),$$
defined and of class C^1 at least along $x(t)$.
- Then $Z(t)$ has a unique expression as a linear combination of these vectors at each $x(t)$,
$$Z(t) = b^1(t)F_1(t) + \cdots + b^n(t)F_n(t).$$
- Differentiating this expression we obtain
$$\frac{dZ}{dt} = \sum_{j=1}^{n} \left(\frac{db^j}{dt} F_j(t) + b^j(t) \frac{dF_j}{dt} \right).$$

Using Other Field Frames (Cont'd)

- Now $\frac{dF_j}{dt}$ are vectors along $x(t)$.
- So they too are linear combinations of $F_k(t)$,

$$\frac{dF_j}{dt} = \sum_{k=1}^{n} a_j^k(t) F_k(t).$$

- This gives the formula

$$\frac{dZ}{dt} = \sum_k \left(\frac{db^k}{dt} + \sum_j b^j(t) a_j^k(t) \right) F_k(t).$$

- Note that, when the frames $F_1(t), \ldots, F_n(t)$ are parallel, $a_j^k(t) \equiv 0$.
- So the last formula includes the original formula as a special case.

Parametrization by Arc Length

- The length of the curve from a fixed point $x_0 = x(t_0)$ is given by
$$s = \int_{t_0}^{t} (\dot{x}(t), \dot{x}(t))^{1/2} dt.$$
- So
$$\frac{ds}{dt} = (\dot{x}(t), \dot{x}(t)).$$
- If s is used as parameter, then $\frac{ds}{dt} \equiv \frac{ds}{ds} \equiv 1$.
- So $\dot{x}(s)$ is a unit vector tangent to the curve.
- Let
$$T(s) = \dot{x}(s)$$
denote this unit tangent vector.

Invariance of the Parametrization by Arc Length

- Arc length, the parameter s (to within an additive constant), and $T(s)$ are determined by the (induced) Riemannian metric on $x(s)$.
- They do not depend on the particular rectangular Cartesian coordinates or origin used.
- So they and the derivatives of $T(s)$ are geometric invariants of the curve.
- This means that they are the same at corresponding points for congruent curves.

Curvature

- We have
$$(T(s), T(s)) = (\dot{x}(s), \dot{x}(s)) = 1.$$
- So, differentiating, we get
$$\frac{d}{ds}(T(s), T(s)) = 0.$$
- Using the inner product rule, we obtain
$$2\left(T(s), \frac{dT}{ds}\right) \equiv 0.$$
- Therefore, one of the following holds:
    - $\frac{dT}{ds}$ is zero;
    - $\frac{dT}{ds}$ is a nonzero vector orthogonal to $T(s)$ at each point of the curve.
- We define the **curvature** $k(s)$ by
$$k(s) = \left\|\frac{dT}{ds}\right\|.$$

Curvature (Cont'd)

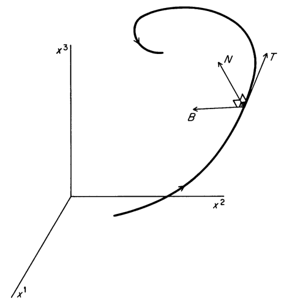
- Suppose $k(s) \neq 0$.
- Then, we let $N(s)$ be the unique unit vector defined by

$$\frac{dT}{ds} = k(s)N(s).$$

- We also let $B(s)$ be the uniquely determined unit vector, such that

$$T(s),\ N(s),\ B(s)$$

  define an orthonormal frame with the orientation of $\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}$.



Non-Zero Curvature

- We assume that $k(s) \neq 0$ at all points of a curve under consideration.
- This assumption is justified, since it is the generic or typical situation for a space curve.

Theorem

If $k(s) \equiv 0$ on the interval of definition, then $x(s)$ is a straight line segment on that interval.
Conversely, for a straight line $x(s)$, $k(s) \equiv 0$.

- Suppose the curve $x(s)$ is a straight line.

  Then it is given, in terms of arclength, by

$$x^i(s) = a^i + b^i s, \quad i = 1, 2, 3,$$

  where $\sum_{i=1}^{3} (b^i)^2 = 1$.

Non-Zero Curvature (Cont'd)

- Thus
$$T = \sum_{i=1}^{3} b^i \frac{\partial}{\partial x^i}.$$
So $\frac{dT}{ds} \equiv 0$.
Conversely, suppose $k(s) \equiv 0$.
Then
$$\frac{dT}{ds} \equiv 0.$$
But
$$T = \sum \frac{dx^i}{ds} \frac{\partial}{\partial x^i},$$
where s is arclength.

Non-Zero Curvature (Cont'd)

- This implies
$$\frac{d^2x^i}{ds^2} = 0, \quad i = 1, 2, 3.$$
Thus
$$x^i(s) = a^i + b^i s, \quad i = 1, 2, 3,$$
with a^i and b^i constants.
So the curve is a straight line.
- Note that $T(s)$ and $k(s)$ are defined for a curve in $\mathbb{R}^n$, for any n (not just $n = 3$), and the proposition just proved is still valid.

The Matrix of the Derivation

- For convenience of notation, we let $F_1(s)$, $F_2(s)$, $F_3(s)$ denote $T(s)$, $N(s)$, $B(s)$, respectively.
- Since this is a field of orthonormal frames, we have
$$(F_i(s), F_j(s)) \equiv \delta_{ij}.$$
- Differentiation of these equations gives the relations
$$\left(\frac{dF_i}{ds}, F_j(s)\right) + \left(F_i(s), \frac{dF_j}{ds}\right) \equiv 0, \quad i,j = 1,2,3.$$
- As we pointed out in the derivation, $\frac{dF_j}{ds}$ must be a linear combination of the $F_k(s)$, for every s.
- So we may write
$$\frac{dF_j}{ds} = \sum_k a_j^k F_k(s), \quad j = 1,2,3.$$

The Matrix of the Derivation (Cont'd)

- Combining, we get
$$\left(\sum_k a_i^k F_k, F_j\right) + \left(F_i, \sum_k a_j^k F_k\right) \equiv 0.$$
- Equivalently
$$a_i^j(s) + a_j^i(s) \equiv 0, \quad 1 \leq i, j \leq 3.$$
- This means that the matrix $(a_j^i(s))$ is skew-symmetric.
- By definition $\frac{dT}{ds} = k(s)N$.
- This gives
$$a_1^2(s) = k(s).$$
- So $a_3^1(s) \equiv 0 \equiv a_1^3(s)$.
- Finally, we use the notation
$$a_2^3(s) = \tau(s).$$

The Matrix of the Derivation (Cont'd)

- Rewriting in terms of T, N, B, we have the **Frenet-Serret formulas**

$$\begin{aligned}
\frac{dT}{ds} &= k(s)N, \\
\frac{dN}{ds} &= -k(s)T + \tau(s)B, \\
\frac{dB}{ds} &= -\tau(s)N.
\end{aligned}$$

- They, express the derivatives with respect to s of T, N and B, which are called the **tangent**, **normal** and **binormal vectors**, respectively, of $x(s)$, in terms of these vectors themselves.

Definition

$k(s)$ is called the **curvature** and $\tau(s)$ the **torsion** of the curve C at $x(s)$.

- Curvature measures deviation of C from a straight line.
- Torsion measures "twisting" or deviation from being a plane curve.

Characterization of Plane Curves

Theorem

A curve in $\boldsymbol{E}^3$ lies in a plane if and only if $\tau(s) \equiv 0$.

- Suppose the curve lies in a plane.

  By the definition of $T(s)$ and $\frac{dT}{ds}$, we see that these vectors lie in the plane of the curve for each point $x(s)$ of the curve.

  Thus, $B(s)$ has a fixed direction, orthogonal to the plane.

  So it is always parallel to a fixed unit vector, orthogonal to the plane.

  Therefore, $\frac{dB}{ds} \equiv 0$.

  This gives $\tau(s) \equiv 0$.

Characterization of Plane Curves (Converse)

- Suppose that $\tau(s) \equiv 0$.
  Then $\frac{dB}{ds} \equiv 0$ and B is a constant vector along the curve.
  We choose the coordinate axes so that:
  - The curve passes through the origin 0 at $s = 0$;
  - $B(s)$ is parallel to $\frac{\partial}{\partial x^3}$, the unit vector in the direction of the x^3-axis.

  Then $x(s) = (x^1(s), x^2(s), x^3(s))$ determines the vector $\boldsymbol{x}(s)$ from the origin 0 to the point $x(s)$ on the curve.

  Differentiating $(\boldsymbol{x}(s), B(s))$, we have

$$\frac{d}{ds}(\boldsymbol{x}(s), B(s)) \equiv (T(s), B(s)) + \left(\boldsymbol{x}(s), \frac{dB}{ds}\right) = (T(s), B(s)) \equiv 0.$$

  So $(\boldsymbol{x}(s), B(s))$ is constant.

  Now $x(s_0) = 0$, that is, $\boldsymbol{x}(s_0) = 0$.

  So the vector $\boldsymbol{x}(s)$ [or line $\overline{0x(s)}$] is always perpendicular to $B = \frac{\partial}{\partial x^3}$.

  Thus, the curve lies in the x^1x^2-plane.

Dynamics of a Moving Particle

- We consider briefly the dynamics of a moving particle in space.
- Suppose its position $p(t)$ is given as a function of time t.
- Let $s(t)$ be the length of path traversed from time $t = 0$ to time t,

$$s(t) = \int_0^t \left(\left(\frac{dp}{dt}, \frac{dp}{dt} \right) \right)^{1/2} dt.$$

- Then the **speed** with which the particle moves along the curve is

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

Dynamics of a Moving Particle (Cont'd)

- Its **velocity vector** is given by

$$\boldsymbol{v}(t) = \frac{dp}{dt} = \frac{dp}{ds}\frac{ds}{dt} = T\frac{ds}{dt},$$

  where T is the unit tangent vector.
- Differentiating, we get the **acceleration**

$$\boldsymbol{a}(t) = \frac{d^2p}{dt^2} = \frac{dT}{ds}\left(\frac{ds}{dt}\right)^2 + T\frac{d^2s}{dt^2}.$$

- We have $\frac{dT}{ds} = kN$.
- So we obtain

$$\boldsymbol{a}(t) = \frac{d^2s}{dt^2}T + k\left(\frac{ds}{dt}\right)^2 N.$$

Dynamics of a Moving Particle (Cont'd)

- The acceleration decomposes into the sum of two vectors:
  - One in the direction of the curve, whose magnitude is the time rate of change of the speed $\frac{d^2s}{dt^2}$;
  - The other normal to the curve and directly proportional to both the square of the speed and to the curvature.

    The curvature depends only on the curve.
- If the motion is a straight line motion, then $k = 0$.

  In this case, $\boldsymbol{a}$ has the direction of the line.
- If the particle moves at constant speed, then $\frac{d^2s}{dt^2} = 0$.

  In that case the acceleration depends only on the shape of the path.
- The same remarks also apply to the force F acting on the particle, which by Newton's Second Law, $F = m\boldsymbol{a}$, is proportional to $\boldsymbol{a}$ with the mass m as constant of proportionality.

Curvature of Plane Curves

- We consider the case of a curve C lying on an oriented plane.
- Suppose a curve, parametrized by arclength, is given by

$$s \to (x(s), y(s)).$$

- Then the unit tangent vector is

$$T = \dot{x}(s)\frac{\partial}{\partial x} + \dot{y}(s)\frac{\partial}{\partial y}.$$

- If $\frac{dT}{ds} \neq 0$, then we may as before define

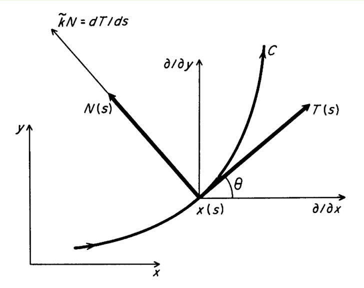
$$k(s) = \left\| \frac{dT}{ds} \right\|.$$

- That is:
  - We consider the curve as a space curve $(x(s), y(s), 0)$, whose z-coordinate $z(s) = 0$;
  - We use the same definitions as before.

Curvature of Plane Curves (Cont'd)

- However, for plane curves a more refined definition of curvature is possible.

- At each point of C choose N so that
$$T,\ N$$
have the same orientation as $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$ (this uniquely determines T, N).



- Then define the curvature $\widetilde{k}(s)$ so that
$$\widetilde{k}(s)N = \frac{dT}{ds}.$$

- This allows $\widetilde{k}(s)$ to be negative, zero or positive.

Curvature of Plane Curves (Cont'd)

- The curvature thus defined for a plane curve has the previously defined curvature of C (considered as a space curve) as its absolute value, $k(s) = |\widetilde{k}(s)|$.
- To carry our interpretation somewhat further, let $\theta(s)$ be the angle of T with the positive x-axis
- Then we have

$$T(s) = \cos\theta(s)\frac{\partial}{\partial x} + \sin\theta(s)\frac{\partial}{\partial y}.$$

- Differentiating with respect to s,

$$\frac{dT}{ds} = -\dot{\theta}(s)\sin\theta(s)\frac{\partial}{\partial x} + \dot{\theta}(s)\cos\theta(s)\frac{\partial}{\partial y}.$$

Curvature of Plane Curves (Cont'd)

- The unit vector $N(s)$ chosen so that $T(s), N(s)$ is an oriented orthonormal basis is

$$N(s) = -\sin\theta \frac{\partial}{\partial x} + \cos\theta \frac{\partial}{\partial y}.$$

- This is because the determinant of the coefficients of T, N as combinations of $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$, is

$$\det \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = +1.$$

- Thus,

$$\widetilde{k}(s) = \dot{\theta}(s) = \frac{d\theta}{ds},$$

the rate of turning of the tangent vector T with respect to arclength.

Curvature of Plane Curves (Cont'd)

- We got
$$\widetilde{k}(s) = \dot{\theta}(s).$$
- Moving along C in the direction of increasing s, the curvature is:
  - Positive, when the tangent is turning counterclockwise;
  - Negative, otherwise.
- Its sign depends on the sense of the curve (direction of increasing s) and the orientation of the plane, but not on the coordinates.

Example

- Suppose C is a circle of radius r.
- The curve parametrized by arclength is

$$s \to \left(r \cos \frac{s}{r}, r \sin \frac{s}{r} \right).$$

- So we have

$$T = -\sin\left(\frac{s}{r}\right)\frac{\partial}{\partial x} + \cos\left(\frac{s}{r}\right)\frac{\partial}{\partial y}.$$

- Then we get

$$\widetilde{k}N = \frac{dT}{ds} = -\frac{1}{r}\cos\left(\frac{s}{r}\right)\frac{\partial}{\partial x} - \frac{1}{r}\sin\left(\frac{s}{r}\right)\frac{\partial}{\partial y}.$$

Example (Cont'd)

- We got $T = -\sin\left(\frac{s}{r}\right)\frac{\partial}{\partial x} + \cos\left(\frac{s}{r}\right)\frac{\partial}{\partial y}$.
- So
$$N = -\cos\left(\frac{s}{r}\right)\frac{\partial}{\partial x} - \sin\left(\frac{s}{r}\right)\frac{\partial}{\partial y}$$
is the unique unit vector such that T, N has the orientation of $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}$.
- Thus, we have
$$\widetilde{k}(s) = \frac{1}{r}.$$
- So the curvature is a constant.
- If, as we have assumed by our parametrization, the circle is traversed in the counterclockwise sense, it is a positive number.
- In any case, its magnitude is inversely proportional to the radius.

Dynamics of Moving Particle (Revisited)

- We return momentarily to the dynamics of a moving particle.
- Suppose a particle moves on a circle in such a way that its speed is constant v_0.
- Then the force F acting on the particle is
$$F = m\boldsymbol{a} = \frac{mv_0^2}{r}N.$$
- Now N is the unit normal vector.
- So F is directed toward the center of the circle.
- Moreover, its magnitude is $\frac{mv_0^2}{r}$.
- This gives the usual formula for the centripetal force necessary to keep the particle in a circular orbit.

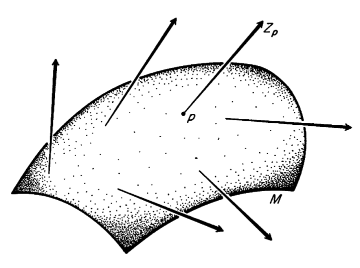
Subsection 2

Differentiation of Vector Fields on Submanifolds of $\mathbb{R}^n$

Vector Fields on Submanifolds of $\mathbb{R}^n$

- We are concerned with a vector field Z defined at each point of a manifold $M \subseteq \mathbb{R}^n$ but not necessarily tangent to M.
- That is, to each $p \in M$, we assign

$$Z_p \in T_p(\mathbb{R}^n).$$



- When Z is such that Z_p is tangent to M, $Z_p \in T_p(M) \subseteq T_p(\mathbb{R}^n)$.
- In that case, we shall say that Z is a **vector field on** M or a **tangent vector field**.
- Only in this case does Z have meaning for M as an abstract manifold, independent of any imbedding or immersion in $\mathbb{R}^n$.

Class of a Vector Field

- In any case differentiability of Z may be given meaning.
- The components of Z, relative to the canonical frames of $\mathbb{R}^n$ at points of M, will be functions on M,

$$Z_p = \sum_{\alpha=1}^{n} a^\alpha(p) \left(\frac{\partial}{\partial x^\alpha}\right)_p.$$

- By definition, we say that Z is of **class** C^r if $a^\alpha(p)$, $\alpha = 1, \ldots, n$, are of class C^r on M.
- In particular, the vector fields

$$\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$$

of $\mathbb{R}^n$, restricted to M, are C^∞-vector fields *along* M (but rarely *on* M).

Decomposition Into Tangent and Normal Space

- If $p \in M$, then $T_p(\mathbb{R}^n)$ and its subspace $T_p(M)$ carry the standard inner product of $\mathbb{R}^n$.
- So M has the induced Riemannian metric.
- This allows us to decompose any vector Z_p, $p \in M$, in a unique way into

$$Z_p = Z_p' + Z_p'',$$

  with:
  - $Z_p' \in T_p(M)$;
  - $Z_p'' \in T_p^{\perp}(M)$, the orthogonal complement of $T_p(M)$.

Decomposition and Projections (Cont'd)

- This reflects the direct sum decomposition of $T_p(\mathbb{R}^n)$ into mutually orthogonal subspaces,

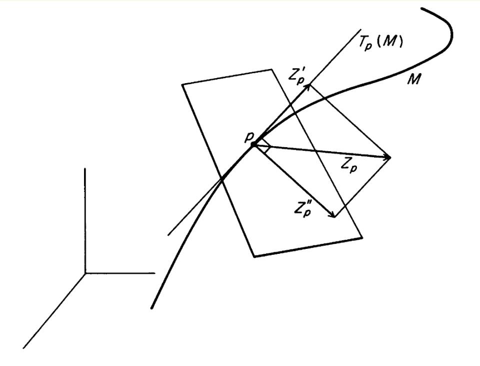
$$T_p(\mathbb{R}^n) = T_p(M) \oplus T_p^{\perp}(M),$$

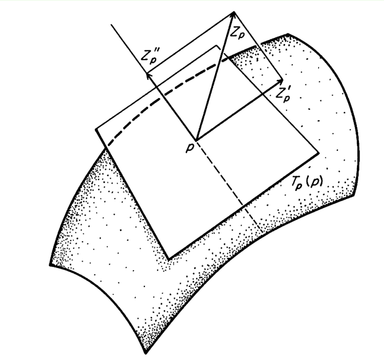
  where:
  - $T_p(M)$ is called the **tangent space** to M at p;
  - $T_p^{\perp}(M)$ is called the **normal space** to M at p.
- Let π', π'' denote the projections,

$$\pi'(Z_p) = Z'_p \quad \text{and} \quad \pi''(Z_p) = Z''_p.$$

- They are linear mappings of $T_p(\mathbb{R}^n)$ onto the subspaces tangent and normal to M.

Illustration of the Decomposition





Decomposition Lemma

- Suppose that Z is a vector field along M of class C^r.
- Then $\pi'(Z)$ and $\pi''(Z)$ are also vector fields, which are tangent and normal to M, provided that they are differentiable.

Lemma

Let Z be a vector field along M of class C^r.
Then $\pi'(Z)$ and $\pi''(Z)$ define mutually orthogonal C^r-vector fields Z', Z'' along M, such that

$$Z = Z' + Z''.$$

That is, at each $p \in M$:

- $Z'_p \in T_p(M)$;
- $(Z'_p, Z''_p) = 0$.

Decomposition Lemma (Cont'd)

Lemma (Cont'd)

If f is a function of class C^r on M, then

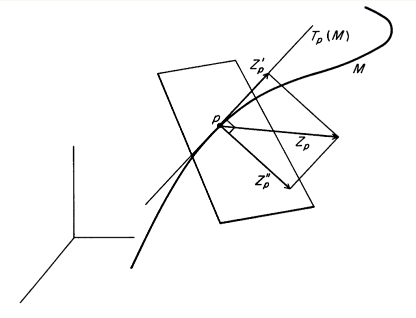
$$\pi'(fZ) = f\pi'(Z) \quad \text{and} \quad \pi''(fZ) = f\pi''(Z).$$

Further, given two such vector fields Z_1, Z_2, then:

- $\pi'(Z_1 + Z_2) = \pi'(Z_1) + \pi'(Z_2)$;
- $\pi''(Z_1 + Z_2) = \pi''(Z_1) + \pi''(Z_2)$.

Example

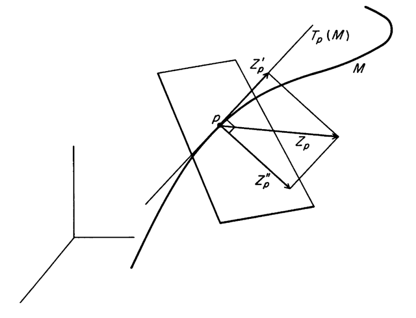
- A vector field Z along a curve decomposes uniquely into the sum of:



  - A tangent vector field $\pi'(Z) = (Z, T)T$;
  - A vector field in the normal plane $\pi''(Z) = (Z, N)N + (Z, B)B$.

Example (Cont'd)

- Consider the case of an arbitrary C^∞ imbedded manifold M.



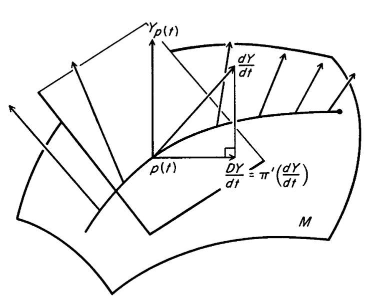
- We see that

$$\pi'\left(\frac{\partial}{\partial x^\alpha}\right), \quad \alpha = 1, \ldots, n,$$

applied at each $p \in M$, gives a C^∞ tangent vector field to M.

The Covariant Derivative

- Let Y be a tangent vector field to $M \subseteq \mathbb{R}^n$.
- That is, for each $p \in M$, $Y_p \in T_p(M)$, or, equivalently, $\pi'(Y) \equiv Y$.
- Let $p(t)$ be a curve on M of class C^1 or higher, defined on some t-interval.
- Then $Y(t) = Y_{p(t)}$ is a vector field along the curve.
- As such, we can ignore M and differentiate $Y(t)$ as a vector field along a curve in $\mathbb{R}^n$.
- In this way, we obtain $\frac{dY}{dt}$, another vector field along the curve.
- In general, of course, $\frac{dY}{dt}$ will not be tangent to M.
- At each point $p(t)$ we may project $\frac{dY}{dt}$ to a tangent vector $\pi'(\frac{dY}{dt})$.



The Covariant Derivative (Cont'd)

Definition

The projection $\pi'(\frac{dY}{dt})$ is denoted $\frac{DY}{dt}$ and is called the **covariant derivative of the tangent vector field** Y **on** M **along the curve** $p(t)$.

- It is important to note that $Y(t)$ need not be the restriction to a curve $p(t)$ of a vector field Y on M for $\frac{DY}{dt}$ to be defined.
- It suffices that $Y(t)$ be a vector field along $p(t)$, so defined that it is always tangent to M, i.e., such that $Y(t) \in T_{p(t)}(M)$.
- Then, as above, $\frac{DY}{dt} = \pi'(\frac{dY}{dt})$, where $\frac{dY}{dt}$ is the derivative of the vector field along a curve, as defined in the previous section.

Properties of the Covariant Derivative

- Suppose that we have vector fields $Y_1(t)$ and $Y_2(t)$ along $p(t)$ on M and tangent to M.

Theorem

Let $Y(t), Y_1(t), Y_2(t)$ be as above and $f(t)$ a C^1 function of t.
Then we have:

- $\frac{D}{dt}(Y_1 + Y_2) = \frac{DY_1}{dt} + \frac{DY_2}{dt}$;
- $\frac{D}{dt}(f(t)Y(t)) = \frac{df}{dt}Y(t) + f(t)\frac{DY}{dt}$;
- $\frac{d}{dt}(Y_1, Y_2) = \left(\frac{DY_1}{dt}, Y_2\right) + \left(Y_1, \frac{DY_2}{dt}\right)$.

- The last equation concerns the induced Riemannian metric on M.
- This is the inner product on $T_p(M)$, at each $p \in M$, induced by the inner product in $T_p(\mathbb{R}^n)$.

Properties of the Covariant Derivative (Cont'd)

- These properties are immediate consequences of:
  - The definitions;
  - The properties of π';
  - The corresponding statements for ordinary derivatives.
- For the fist property, start with

$$\frac{d}{dt}(Y_1(t)+Y_2(t))=\frac{dY_1}{dt}+\frac{dY_2}{dt}.$$

Apply π' to both sides to get

$$\frac{D}{dt}(Y_1(t)+Y_2(t))=\pi'\left(\frac{dZ_1}{dt}+\frac{dZ_2}{dt}\right).$$

Then use linearity to obtain

$$\frac{D}{dt}(Y_1+Y_2)=\frac{DY_1}{dt}+\frac{DY_2}{dt}.$$

Properties of the Covariant Derivative (Cont'd)

- For the second property, start with

$$\frac{d}{dt}(f(t)Y(t)) = \frac{df}{dt}Y(t) + f(t)\frac{dY}{dt}.$$

Then we get

$$\begin{aligned}\frac{D}{dt}(fY) &= \pi' \frac{d}{dt}(fY) \\ &= \pi' \left(\frac{df}{dt}Y + f\frac{dY}{dt}\right) \\ &= \frac{df}{dt}Y + f\frac{DY}{dt}.\end{aligned}$$

Properties of the Covariant Derivative (Cont'd)

- The last property follows from

$$\frac{d}{dt}(Y_1(t), Y_2(t)) = \left(\frac{dY_1}{dt}, Y_2(t)\right) + \left(Y_1(t), \frac{dY_2}{dt}\right).$$

Note that, for $i = 1, 2$,

$$\frac{dY_i}{dt} = \pi'\left(\frac{dY_i}{dt}\right) + \pi''\left(\frac{dY_i}{dt}\right) = \frac{DY_i}{dt} + \pi''\left(\frac{dY_i}{dt}\right).$$

Note, also, that $\pi''\left(\frac{dY_i}{dt}\right)$ is orthogonal to $T_{p(t)}(M)$.

So we have

$$\begin{aligned}\left(\frac{DY_1}{dt} + \pi''\left(\frac{dY_1}{dt}\right), Y_2\right) + \left(Y_1, \frac{DY_2}{dt} + \pi''\left(\frac{dY_2}{dt}\right)\right) \\ = \left(\frac{DY_1}{dt}, Y_2\right) + \left(Y_1, \frac{DY_2}{dt}\right).\end{aligned}$$

Remark

- Suppose we change to a new parameter, say s, using $t = f(s)$.
- Then, since $\frac{dt}{ds} = f'(s)$ is a scalar,

$$\frac{DY}{ds} = \frac{DY}{dt}\frac{dt}{ds}.$$

- Alternatively, we may apply π' to the relation

$$\frac{dY}{ds} = \frac{dY}{dt}\frac{dt}{ds}$$

of the previous section.

Constant or Parallel Vector Fields

Definition

Given $M \subseteq \mathbb{R}^n$ as above, let $Y_{p(t)}$ be a vector field, such that:

- $Y_{p(t)}$ is defined at each point of a curve $p(t)$ on M;
- $Y_{p(t)}$ at each point is tangent to M.

That is, $Y_{p(t)}$ is a vector field along $p(t)$ tangent to M.
Then we shall say that $Y_{p(t)}$ is a **constant** or **parallel** vector field if

$$\frac{DY}{dt} = 0.$$

More generally if Y is a tangent vector field on all of M, then it is **constant** or **parallel** if it has this property along every curve on M.

Remarks

- It is very important to note that $\frac{DY}{dt}$ may be identically zero even though $\frac{dY}{dt}$ is not.
- Thus, a vector field along a curve may be:
  - Constant considered as a vector field on a submanifold M of $\mathbb{R}^n$;
  - Non constant considered as a vector field along the same curve in $\mathbb{R}^n$.

Example

- Let $M = S^1$, the unit circle in $\mathbb{R}^2$.
- Its parametric representation is

$$t \to (\cos t, \sin t).$$

- It may be considered as defining a curve on M.
- Let $Y(t)$ be the unit tangent vector to this curve.
- As we have seen $\frac{dY}{dt}$ is orthogonal to $Y(t)$, that is, normal to M.
- Hence,

$$\frac{DY}{dt} = \pi' \left(\frac{dY}{dt} \right) = 0.$$

- On the other hand, $\frac{dY}{dt}$ is never zero.
- In fact has constant length $+1$.

Example (Cont'd)

- Now any great circle on the unit sphere $S^{n-1} \subseteq \mathbb{R}^n$ is congruent to the great circle

$$t \to p(t) = (\cos t, \sin t, 0, \ldots, 0)$$

on the intersection of S^{n-1} and the 2-plane $x^3 = \ldots = x^n = 0$ of $\mathbb{R}^n$.

- So the unit tangent vector to any great circle arc $p(t)$, parametrized by arclength, has the same property,

$$\frac{DY}{dt} = \frac{D}{dt}\left(\frac{dp}{dt}\right) \equiv 0.$$

Geodesics

- In general the derivative of a tangent vector field to M along a curve $p(t)$ in M has both normal and tangential components nonzero.
- If a curve on M is such that

$$\frac{D}{dt}\frac{dp}{dt} = 0,$$

that is, the (covariant) derivative of the unit tangent vector to the curve is zero along the curve, then we shall say the curve is a **geodesic** of M.
- So the great circles on the unit sphere in $\mathbb{R}^n$ are geodesics.

The Case of $\mathbb{R}^n$

- In the case in which M is an open subset of $\mathbb{R}^n$ or all of $\mathbb{R}^n$, then $\frac{dY}{dt} = \frac{DY}{dt}$, that is, in $\mathbb{R}^n$ itself, as might be expected, covariant differentiation is just the usual differentiation.
- In this special case, according to a previous theorem, the only curves $p(t)$ for which $\frac{D}{dt}\frac{dp}{dt} = \frac{d}{dt}\frac{dp}{dt}$ vanishes identically are straight lines parametrized by arclength - or with t proportional to arclength.
- Thus geodesics on an imbedded manifold M are those curves which in some sense generalize the concept of straight line - even though they may not look "straight" when viewed from the ambient space $\mathbb{R}^n$.

Parametrization of a Manifold

- Suppose $\dim M = m$ and that U, φ is a local coordinate system on M with $\varphi(U) = W$, an open subset of $\mathbb{R}^m$.
- We denote the local coordinates by $u^1, \ldots, u^m$.
- $\varphi^{-1} : W \to \mathbb{R}^n$ is an imbedding of W whose image is U, an open subset of M.
- We have previously referred to φ^{-1} as a *parametrization* of M.
- Let $u = (u^1, \ldots, u^m)$.
- Then
$$\varphi^{-1}(u) = (g^1(u), \ldots, g^n(u)), \quad u \in W,$$
gives φ^{-1} in terms of its coordinate mappings $g^\alpha(u)$.
- We let α, β, γ, and so on, denote indices that range from 1 to n.
- We let i, j, k, and so on, denote indices ranging from 1 to m.

The Coordinate Frames

- The coordinate frames will be denoted $F_1, \ldots, F_m$.
- They span the tangent space to M at each point.
- This tangent space $T_p(M)$ at $p \in M$ is a subspace of $T_p(\mathbb{R}^n)$.
- So these vectors are linear combinations of $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$.
- In fact, generalizing earlier formulas for $m = 2$ and $n = 3$ we have:

$$F_{ip} = \varphi_*^{-1}\left(\frac{\partial}{\partial u^i}\right) = \sum_{\alpha=1}^{n}\left(\frac{\partial g^\alpha}{\partial u^i}\right)_{\varphi(p)}\frac{\partial}{\partial x^\alpha}.$$

Covariant Derivative and Local Coordinates

- Suppose that $p(t)$ is a curve on M of class C^1.
- Let $Y(t) = Y_{p(t)}$ be a vector field along the curve which is always tangent to M.
- Then $Y(t)$ may be written as a linear combination of $F_1, \ldots, F_m$,

$$Y(t) = \sum_{k=1}^{m} b^k(t) F_k.$$

- The derivative

$$\frac{dY}{dt} = \sum \left(\frac{db^k}{dt} F_k + b^k \frac{dF_k}{dt} \right)$$

is not tangent to M in general.

Covariant Derivative and Local Coordinates (Cont'd)

- Take
$$\frac{dY}{dt} = \sum \left(\frac{db^k}{dt} F_k + b^k \frac{dF_k}{dt} \right).$$
- By projecting, we obtain
$$\frac{DY}{dt} = \pi' \left(\frac{dY}{dt} \right) = \sum_{k=1}^{m} \left(\frac{db^k}{dt} F_k + b^k \pi' \left(\frac{dF_k}{dt} \right) \right).$$
- Equivalently,
$$\frac{DY}{dt} = \sum_{k=1}^{m} \left(\frac{db^k}{dt} F_k + b^k \frac{DF_k}{dt} \right).$$
- We know that $\frac{DF_i}{dt}$, $i = 1, \ldots, m$, are vectors tangent to M.
- So they may be expressed as linear combinations of $F_1, \ldots, F_m$.

Covariant Derivative and Local Coordinates (Cont'd)

- Suppose that the curve $p(t)$ is given in local coordinates by

$$\varphi(p(t)) = (u^1(t), \ldots, u^m(t)).$$

- Then in the expression for F_{ip} the components are (composite) functions $(\frac{\partial g^\alpha}{\partial u^i})_{\varphi(p(t))}$ of t through $u^1(t), \ldots, u^m(t)$.
- Further, at each $p(t)$, by the ordinary chain rule of differentiation, and the properties of π',

$$\frac{DF_i}{dt} = \pi'\left(\frac{dF_i}{dt}\right) = \sum_{\alpha=1}^{n}\sum_{j=1}^{m}\frac{\partial^2 g^\alpha}{\partial u^i \partial u^j}\frac{du^j}{dt}\pi'\left(\frac{\partial}{\partial x^\alpha}\right).$$

- The derivatives $\frac{\partial^2 g^\alpha}{\partial u^j \partial u^i}$ are functions of $u^1, \ldots, u^m$ and are evaluated at $u(t) = (u^1(t), \ldots, u^m(t))$ in this formula.

Covariant Derivative and Local Coordinates (Cont'd)

- We assume M is imbedded in $\mathbb{R}^n$ by a C^∞ imbedding.
- We know that then $\frac{\partial}{\partial x^\alpha}$, restricted to M, is a C^∞ vector field along M.
- By a previous lemma,
$$\pi'\left(\frac{\partial}{\partial x^\alpha}\right)$$
defines a C^∞ tangent vector field on M.
- This must have a unique expression on U of the form
$$\pi'\left(\frac{\partial}{\partial x^\alpha}\right) = \sum_{k=1}^{m} a_\alpha^k(u) F_k.$$
- The $a_\alpha^k(u)$ are C^∞ functions on M which we do not compute.

Covariant Derivative and Local Coordinates (Cont'd)

- We have
$$\pi'\left(\frac{\partial}{\partial x^\alpha}\right) = \sum_{k=1}^{m} a_\alpha^k(u) F_k.$$
- Using the $a_\alpha^k(u)$ and the coordinate functions $g^\alpha(u)$ of the parametrization φ^{-1}, we define the C^∞ functions $\Gamma_{ij}^k(u)$ as
$$\Gamma_{ij}^k = \sum_\alpha \frac{\partial^2 g^\alpha}{\partial u^i \partial u^j} a_\alpha^k = \Gamma_{ji}^k, \quad 1 \le i, j, k \le m.$$
- Symmetry in i, j is due to interchangeability of the order of differentiation.

Covariant Derivative and Local Coordinates (Cont'd)

- We do not explicitly compute the Γ_{ij}^k, but we use them to write new formulas for $\frac{DF_i}{dt}$:

$$\frac{DF_i}{dt} = \sum_{j,k=1}^{m} \Gamma_{ij}^k \frac{du^j}{dt} F_k, \quad i = 1, \ldots, m$$

  at each $p = p(t)$, the Γ_{ij}^k being evaluated at $(u^1(t), \ldots, u^m(t))$.
- Consider the particular case of the curve given by

$$u^i = \begin{cases} \text{constant}, & \text{if } i \neq j \\ t, & \text{if } i = j. \end{cases}$$

- This gives the formula for the covariant derivative of the vector field F_i along the jth coordinate curve, conveniently denoted $\frac{DF_i}{\partial u^j}$,

$$\frac{DF_i}{\partial u^j} = \sum_k \Gamma_{ij}^k F_k.$$

Covariant Derivative and Local Coordinates (Cont'd)

- We get an interpretation of the meaning of $\Gamma_{ij}^k(u)$.
- It is the kth component (relative to the coordinate frames) of the covariant derivative of F_i along that curve in which only the jth coordinate is allowed to vary, that is, along a coordinate curve.
- We look again at

$$\frac{DY}{dt} = \sum_{k=1}^{m} \left(\frac{db^k}{dt} F_k + b^k \frac{DF_k}{dt} \right).$$

- Using the formulas above, we may write those as

$$\frac{DY}{dt} = \sum_{k=1}^{m} \left(\frac{db^k}{dt} + \sum_{i,j=1}^{m} \Gamma_{ij}^k(u(t)) b^i(t) \frac{du^j}{dt} \right) F_k.$$

Covariant Derivative and Local Coordinates (Cont'd)

- The formula

$$\frac{DY}{dt} = \sum_{k=1}^{m} \left(\frac{db^k}{dt} + \sum_{i,j=1}^{m} \Gamma_{ij}^k(u(t)) b^i(t) \frac{du^j}{dt} \right) F_k$$

  expresses $\frac{DY}{dt}$ in terms of the field of frames $F_1, \ldots, F_s$ on $U \subseteq M$, frames defined independently of either $p(t)$ or Y.
- The components of the covariant derivative are the terms in brackets.
- The functions $\Gamma_{ij}^k(u)$ are defined over all of U and in the formula are evaluated at points of the curve.
- Indeed for every coordinate neighborhood on M we have frames F_i, $i = 1, \ldots, m$, and functions Γ_{ij}^k which give $\frac{DF_i}{\partial u^j}$.
- From these data $\frac{DY}{dt}$ can then be computed according by ordinary differentiation of the components of Y and coordinates of $p(t)$.

Directional Derivative of a Vector Field

- Let Y be a tangent vector field on M which is defined everywhere - not just along some curve.
- On the coordinate neighborhood U we write

$$Y = \sum_{k=1}^{m} b^k(u) F_k.$$

- Let p be a point of U, such that $\varphi(p) = (u_0^1, \ldots, u_0^m)$.
- Let X_p be a tangent vector at p,

$$X_p = \sum a^j F_{jp},$$

where a^j is constant for $j = 1, \ldots, m$.

Directional Derivative of a Vector Field (Cont'd)

- Now choose any differentiable curve $p(t)$ whatsoever with:
  - $p(t_0) = p$;
  - $(\frac{dp}{dt})_{t_0} = X_p$.
- So, in local coordinates, it is defined by

$$u(t) = (u^1(t), \ldots, u^m(t)),$$

  with:
  - $u^i(t_0) = u_0^i$;
  - $(\frac{du^i}{dt})_{t_0} = a^i$.
- Then we may compute $\left(\frac{DY}{dt}\right)_{t=t_0}$ as above with a surprising result.

Directional Derivative of a Vector Field (Cont'd)

- First, we observe that $Y(t) = \sum b^k(u(t))F_k$ implies that

$$\left(\frac{db^k}{dt}\right)_{t_0} = \sum_{j=1}^{m} \left(\frac{\partial b^k}{\partial u^j}\right)_{u_0} a^j = X_p b^k.$$

- Taking this into consideration, the formula for the covariant derivative gives

$$\left(\frac{DY}{dt}\right)_{t_0} = \sum_k \left(X_p b^k + \sum_{i,j} \Gamma^k_{ij}(u_0) b^i(u_0) a^j\right) F_k.$$

- A careful examination of this formula discloses the remarkable fact that the right-hand side does not depend on $p(t)$ but only on its tangent vector X_p at p.
- We know that $(\frac{DY}{dt})_{t_0}$ is a vector in $T_p(M)$.
- So this formula defines a mapping of $T_p(M)$ to itself $X_p \to (\frac{DY}{dt})_{t_0}$.

Directional Derivative of a Vector Field (Cont'd)

- We introduce the notation $\nabla_{X_p} Y$ for the image of X_p,

$$\nabla_{X_p} Y = \left(\frac{DY}{dt}\right)_{t_0},$$

along any curve $p(t)$ with $p(t_0) = p$ and $(\frac{dp}{dt})_{t_0} = X_p$.
- We have defined previously a "directional derivative" $X_p f$ of a function f with respect to a vector X_p.
- What we have just now done is define in similar fashion a rate of change of the vector field Y at p in the direction X_p.
- It is worth commenting that, as a consequence of our notation, along the curve $p(t)$, we have at each point

$$\nabla_{\frac{dp}{dt}} Y = \frac{DY}{dt}.$$

Properties of $\nabla_{X_p} Y$

Theorem

Let $M \subseteq \mathbb{R}^n$ be a submanifold. For any tangent vector field Y of class C^r, $r > 1$, on M, we have at each point $p \in M$ a linear mapping

$$\begin{array}{rcl} T_p(M) & \to & T_p(M); \\ X_p & \mapsto & \nabla_{X_p} Y. \end{array}$$

Then $\nabla_{X_p} Y$, being defined as above, has the following properties:

(1) If X, Y are vector fields of class C^r (of class C^∞) on M, then $\nabla_X Y$, defined by

$$(\nabla_X Y)_p = \nabla_{X_p} Y,$$

is a C^{r-1} (respectively, C^∞) vector field on M.

Properties of $\nabla_{X_p} Y$ (Cont'd)

Theorem (Cont'd)

(2) The map $T_p(M) \times \mathfrak{X}(M) \to T_p(M)$ given by

$$(X_p, Y) \to \nabla_{X_p} Y$$

is $\mathbb{R}$-linear in X_p and Y.
For a function f, differentiable on a neighborhood of p,

$$\nabla_{X_p}(fY) = (X_p f) Y_p + f(p) \nabla_{X_p} Y.$$

Properties of $\nabla_{X_p} Y$ (Cont'd)

Theorem (Cont'd)

(3) If $X, Y \in \mathfrak{X}(M)$, then

$$[X, Y] = \nabla_X Y - \nabla_Y X.$$

(4) If Y_1 and Y_2 are vector fields and (Y_1, Y_2) their inner product, then

$$X_p(Y_1, Y_2) = (\nabla_{X_p} Y_1, Y_{2p}) + (Y_{1p}, \nabla_{X_p} Y_2).$$

Proof of the Properties

- Let $Y = \sum b^k F_k$ and $X = \sum a^k F_k$ in the notation just used.

  The b^k are functions of the local coordinates $(u^1, \ldots, u^m)$.

  So are the a^k when X is a vector field.

  We have
$$X_p b^k = \sum_{j=1}^{m} \frac{\partial b^k}{\partial u^j} a^j.$$
  So the definition $\nabla_{X_p} Y = \frac{DY}{dt}$ and the formula obtained previously for $(\frac{DY}{dt})_{t_0}$ imply that
$$\nabla_{X_p} Y = \sum_k \sum_j \left(\frac{\partial b^k}{\partial u^j} a^j + \sum_i \Gamma^k_{ij} b^i a^j \right) F_k.$$
  This formula, valid for each $p \in U$, yields Properties (1) and (2).

Proof of the Properties (Cont'd)

- Property (4) expresses an earlier property of $\frac{DY}{dt}$.

  To see this, note that

  $$X_p f = \frac{df}{dt},$$

  the derivative of $f(p(t))$, when we assume $X_p = \frac{dp}{dt}$.

  In particular this holds for $f = (Y_1, Y_2)$.

  Only Property (3) requires more careful verification.

  We will verify Property (3) by direct computation in a coordinate neighborhood U, φ using our previous notation.

Proof of the Properties (Cont'd)

- With X and Y given on U as above we compute $[X, Y]$,

$$[X, Y] = \sum_{k,j} \left(\frac{\partial b^k}{\partial u^j} a^j - \frac{\partial a^k}{\partial u^j} b^j \right) F_k.$$

  Using the formula for $\left(\frac{DY}{dt}\right)_{t_0}$ we compute $\nabla_{X_p} Y - \nabla_{Y_p} X$.

  We have

$$\nabla_{X_p} Y - \nabla_{Y_p} X = \sum_k \left\{ \left(\frac{\partial b^k}{\partial x^j} a^j - \sum \frac{\partial a^k}{\partial x^j} b^j \right) + \sum_{i,j} \Gamma_{ij}^k (b^i a^j - a^i b^j) \right\} F_k.$$

  Since $\Gamma_{ij}^k = \Gamma_{ji}^k$, the second sum is zero.

  So the expression reduces to the first term in the parentheses.

Directional and Covariant Derivatives

- A careful reexamination of what we have done will show that $\nabla_{X_p} Y$ depends for its definition only on the Euclidean structure of $\mathbb{R}^n$.
- That is $\nabla_{X_p} Y$ depends on $\boldsymbol{E}^n$ and on the imbedding of M in $\boldsymbol{E}^n$.
- It is independent of local coordinates, although we use them in its definition and in the proof above.
- However, $\frac{dY}{dt}$ and $\frac{DY}{dt} = \pi'\left(\frac{dY}{dt}\right)$ are geometric in nature.
- The same holds for $\nabla_{X_p} Y$.
- If $\nabla_{X_p} Y$ is axiomatized and defined first, then $\frac{DY}{dt}$ could be introduced by
$$\frac{DY}{dt} = \nabla_{\frac{dp}{dt}} Y.$$
- This would allow us to reverse our definitions and steps above.

Roles of X and Y in ∇

- Recall that the symbol $\nabla_X Y$ defines an $\mathbb{R}$-bilinear mapping

$$\begin{array}{rcl} \mathfrak{X}(M) \times \mathfrak{X}(M) & \to & \mathfrak{X}(M); \\ (X, Y) & \to & \nabla_X Y. \end{array}$$

- There is a partial duality of roles of X and Y.
- But there is also an important difference.
- Namely, $\nabla_X Y$ is C^∞-linear in the first variable but not the second.

Directional and Lie Derivatives

- Suppose X and Y are vector fields on M.
- Then the Lie derivative

$$L_X Y = [X, Y]$$

  gives a rate of change, or derivative, of Y in the direction of X.
- However, this derivative requires a vector field X, not just a vector X_p at a single point, as does $\nabla_{X_p} Y$.
- Thus, the two concepts of differentiation are essentially different.
- Property (3) gives the precise relationship between the two.

Subsection 3

Differentiation on Riemannian Manifolds

Connections

Definition

A C^∞ **connection** ∇ on a manifold M is a mapping

$$\begin{array}{rrcl} \nabla: & \mathfrak{X}(M) \times \mathfrak{X}(M) & \to & \mathfrak{X}(M); \\ & (X, Y) & \to & \nabla_X Y, \end{array}$$

which has the following linearity properties.
For all $f, g \in C^\infty(M)$ and $X, X', Y, Y' \in \mathfrak{X}(M)$:

(1) $\nabla_{fX+gX'} Y = f(\nabla_X Y) + g(\nabla_{X'} Y)$;

(2) $\nabla_X(fY + gY') = f\nabla_X Y + g\nabla_X Y' + (Xf)Y + (Xg)Y'$.

- Note the asymmetry in the roles of X and Y.
  V is $C^\infty(M)$ linear in X but not in Y.
- In the special case f is a constant function, we have $Xf = 0$.
  Then ∇ is $\mathbb{R}$-linear in both variables.

Special Properties for Imbedded Manifolds

- By a previous theorem, connections exist for M imbedded in Euclidean space.
- In addition, in this special case, we also have:
  (3) The symmetry property
  $$[X, Y] = \nabla_X Y - \nabla_Y X;$$
  (4) The inner product rule
  $$\nabla_X (Y, Y') = (\nabla_X Y, Y') + (Y, \nabla_X Y').$$

Riemannian Connections

Definition

A C^∞ connection which also has the Symmetry and Inner Product Properties (3) and (4) is called a **Riemannian connection**.

- Note that, in these definitions, it is only Property (4) that involves the Riemannian metric.
- Thus, on arbitrary differentiable manifolds, one may study:
  - C^∞ connections [Properties (1) and (2)];
  - Symmetric C^∞ connections [properties (1)-(3)].

Fundamental Theorem of Riemannian Geometry

Theorem (Fundamental Theorem of Riemannian Geometry)

Let M be a Riemannian manifold. Then there exists a uniquely determined Riemannian connection on M.

- We will prove this theorem in several steps.
- The method is somewhat similar to that of the existence proof for the operator d on $\bigwedge(M)$.

Relation With Manifolds Imbedded in $\mathbb{R}^n$

- In the discussion of differentiation on manifolds imbedded in $\mathbb{R}^n$, we defined the map

$$\begin{array}{rcl} T_p(M) & \to & T_p(M); \\ X_p & \mapsto & \nabla_{X_p} Y. \end{array}$$

- We used the vector field Y, but without any assumption that X_p was the value at p of a vector field X.
- However, given vector fields X and Y, a vector field $\nabla_X Y$ was then defined by

$$(\nabla_X Y)_p = \nabla_{X_p} Y, \quad p \in M.$$

- We thus obtained a map ∇ of pairs (X, Y) of vector fields to a vector field $\nabla_X Y$, as in our present definition.

Relation With Manifolds Imbedded in $\mathbb{R}^n$

- We have now taken this map on pairs of vector fields as the primary notion.
- We wish to see that, conversely, Y defines a linear map of

$$T_p(M) \to T_p(M),$$

for each $p \in M$.
- That is, we wish to see that $(\nabla_X Y)_p$ depends not on the vector field X but only on its value X_p at p.

Vanishing Property

Lemma

Let $X, Y \in \mathfrak{X}(M)$ and suppose that, on an open set $U \subseteq M$,

$$X = 0 \quad \text{or} \quad Y = 0.$$

If ∇ is a connection [satisfying Properties (1) and (2) of the definition], then the vector field $\nabla_X Y = 0$ on U.

- Suppose that $Y = 0$ on U and $q \in U$.
  Then there are:
  - A relatively compact neighborhood V of q, with $\overline{V} \subseteq U$;
  - A C^∞ function f, such that $f = 1$ on $\overline{V}$ and $f = 0$ outside U.

  Since $Y = 0$ on U, $fY \equiv 0$ on M.

Vanishing Property (Cont'd)

- Property (2) implies that ∇_X takes the 0-vector field to 0.

  Therefore $\nabla_X(fY) \equiv 0$ on M.

  But then, using Property (2) again, we have

  $$0 = (\nabla_X(fY))_q = (X_q f)Y_q + f(q)(\nabla_X Y)_q = (\nabla_X Y)_q.$$

  q is an arbitrary point of U.

  So this completes the proof when $Y = 0$ on U.

  A parallel proof using Property (1) applies when $X = 0$ on U.

Equivalence With Earlier Definitions

Corollary

Let p be any point of M. If $X, X' \in \mathfrak{X}(M)$ such that $X_p = X'_p$, then for every vector field Y,

$$(\nabla_X Y)_p = (\nabla_{X'} Y)_p.$$

Denote this uniquely determined vector by $\nabla_{X_p} Y$.
Then the mapping from $T_p(M) \to T_p(M)$ defined by

$$X_p \to \nabla_{X_p} Y$$

is linear.

- Let U, φ be a coordinate neighborhood of the point p.

  Let V be a relatively compact neighborhood of p, with $\overline{V} \subseteq U$.

  Let f a C^∞ function on M which is 1 on $\overline{V}$ and 0 outside U, as in the proof of the lemma.

Equivalence With Earlier Definitions (Cont'd)

- Let $X \in \mathfrak{X}(M)$.
  Then, on U, we have
  $$X = \sum_{i=1}^{n} a_i E_i,$$
  with:
  - $a_i \in C^\infty(U)$;
  - $E_1, \ldots, E_n$ the vectors of the coordinate frames.

  We define $\widetilde{X}, \widetilde{E}_1, \ldots, \widetilde{E}_n \in \mathfrak{X}(M)$ and $\widetilde{a}_1, \ldots, \widetilde{a}_m \in C^\infty(M)$, by
  $$\widetilde{X} = \begin{cases} fX, & \text{on } U, \\ 0, & \text{else}, \end{cases} \quad \widetilde{E}_i = \begin{cases} fE_i, & \text{on } U, \\ 0, & \text{else}, \end{cases} \quad \widetilde{a}_i = \begin{cases} fa_i, & \text{on } U, \\ 0, & \text{else}. \end{cases}$$
  Then we have, on all of M,
  $$\widetilde{X} = \widetilde{a}_1 \widetilde{E}_1 + \cdots + \widetilde{a}_n \widetilde{E}_n.$$
  But on $\overline{V}$, we have $\widetilde{X} = X$, $\widetilde{E}_i = E_i$ and $\widetilde{a}_i = a_i$.
  So this reduces to the equation above.

Equivalence With Earlier Definitions (Cont'd)

- Applying the preceding lemma and Property (1) of ∇ gives that on V,

$$\nabla_X Y = \nabla_{\widetilde{X}} Y = \sum_{i=1}^{n} \widetilde{a}_i \nabla_{\widetilde{E}_i} Y.$$

Hence

$$(\nabla_X Y)_p = \sum \widetilde{a}_i(p)(\nabla_{\widetilde{E}_i} Y)_p = \sum a_i(p)(\nabla_{\widetilde{E}_i} Y)_p.$$

The right side depends only on the value Y_p of the vector field X at p.

This proves the first statement.

Note $\nabla_{X_p} Y = (\nabla_X Y)_p$ depends linearly on the components $a_1(p), \ldots, a_n(p)$ of X_p relative to the basis $E_{1p}, \ldots, E_{np}$ of $T_p(M)$.

This shows that $X_p \to \nabla_{X_p} Y = (\nabla_X Y)_p$ is a linear mapping of $T_p(M)$ into itself.

Restriction of a Connection to Open Subsets

- An important consequence of the lemma is that it allows us to define (unambiguously) the **restriction** ∇^U of a connection ∇ defined on M to any open subset $U \subseteq M$.
- Let X, Y be C^∞-vector fields on U and let $p \in U$.
- We again choose a neighborhood V of p with $\overline{V} \subseteq U$.
- Take a C^∞ function f which is $+1$ on V and vanishes outside U.
- Then $\widetilde{X} = fX$ and $\widetilde{Y} = fY$ may be extended to vector fields on all of M which vanish outside U.
- We then set

$$(\nabla_X^U Y)_p = (\nabla_{\widetilde{X}} \widetilde{Y})_p.$$

- The left hand side is defined at every point of V by this equation.
- By the lemma this definition is independent of the choices.
- It can be verified that ∇^U is a connection and is Riemannian, if ∇ is, using the induced Riemannian metric on U.

On Uniqueness

Lemma

Suppose that a Riemannian connection ∇ exists for every Riemannian manifold. Suppose ∇ is unique for manifolds covered by a single coordinate neighborhood U. Then it is unique for all manifolds. Conversely, suppose there exists a uniquely determined (Riemannian) connection ∇^U, for every Riemannian manifold covered by a single coordinate neighborhood U. Then there exists a uniquely determined Riemannian connection ∇ on every Riemannian manifold.

- We suppose that ∇ is a Riemannian connection on M.

  By hypothesis there is a uniquely determined Riemannian connection ∇^U on each coordinate neighborhood U, φ of M (with the induced Riemannian metric).

On Uniqueness (Cont'd)

- Let X, Y be vector fields on M.

  Denote by X_U, Y_U their restrictions to U.

  By the definition of ∇^U, the restriction of ∇ to U, we get

$$\nabla^U_{X_U} Y_U = (\nabla_X Y)_U.$$

  By the uniqueness assumption, on each coordinate neighborhood,

$$\widetilde{\nabla}^U = \nabla^U.$$

  Thus, we have

$$(\nabla_X Y)_U = \widetilde{\nabla}^U_{X_U} Y_U.$$

  But M is covered by coordinate neighborhoods.

  So this proves the first statement.

On Uniqueness (Cont'd)

- Now suppose that ∇^U is uniquely determined on every coordinate neighborhood U, φ of M.

  If there is defined on M a ∇ satisfying Properties (1)-(4), it must be unique by the above.

  We define ∇ on M as follows.

  Let $X, Y \in \mathfrak{X}(M)$ and let $p \in M$.

  Choose a coordinate neighborhood U, φ containing p.

  Define
$$(\nabla_X Y)_U = \nabla^U_{X_U} Y_U.$$

  This defines $\nabla_X Y$ not only at p but on the neighborhood U.

On Uniqueness (Cont'd)

- We may verify Properties (1)-(4), since they hold for ∇^U.

  Suppose V, ψ is a coordinate neighborhood overlapping U.

  Let $W = U \cap V$.

  Then W is a coordinate neighborhood using either coordinate map φ or ψ.

  Thus, ∇^W is uniquely defined.

  So we have at every point q of W

$$(\nabla^U_{X_U} Y_U)_q = (\nabla^W_{X_W} Y_W)_q = (\nabla^V_{X_V} Y_V)_q.$$

Proof of the Fundamental Theorem

- The proof of the existence and uniqueness of a Riemannian symmetric connection is now reduced to the case of a manifold covered by a single coordinate neighborhood.

  Let U, φ cover the manifold M.

  Let $x^1, \ldots, x^n$ denote the local coordinates.

  Let $E_1, \ldots, E_n$ be the coordinate frames.

  Denote the inner product by (X, Y).

  We have as components of the metric tensor the C^∞ functions on $U = M$

  $$g_{ij}(q) = (E_{iq}, E_{jq}).$$

  The matrix $(g_{ij}(q))$ is symmetric and positive definite.

  Hence, it has a uniquely determined inverse $(g^{ij}(q))$.

  The entries of $(g^{ij}(q))$ are C^∞ functions on U also.

  We show there exists a unique Riemannian connection ∇ on M.

Proof of the Fundamental Theorem (Cont'd)

- First we note that if ∇ can be defined at all, then, by Properties (1) and (2), it is determined by the C^∞ functions Γ_{ij}^k on U, $1 \leq i, j, k \leq n$, defined by
$$\nabla_{E_i} E_j = \sum_{k=1}^{n} \Gamma_{ij}^k E_k.$$
In fact, suppose that, on U,
$$X = \sum b^i(x) E_i \quad \text{and} \quad Y = \sum a^j(x) E_j.$$
Then by Properties (1) and (2) and the definition of Γ_{ij}^k,
$$\nabla_X Y = \sum_k \left(X a^k + \sum_{i,j} \Gamma_{ij}^k a^j b^i \right) E_k.$$
Conversely, given functions Γ_{ij}^k on U, this formula defines a C^∞ connection satisfying Properties (1) and (2).

Proof of the Fundamental Theorem (Cont'd)

- A Riemannian connection also satisfies Properties (3) and (4).
  Consequently, the Γ_{ij}^k are not arbitrary C^∞ functions.
  For the coordinate frames, $[E_i, E_j] = 0$.
  So Property (3) is equivalent to
  $$0 = [E_i, E_j] = \nabla_{E_i} E_j - \nabla_{E_j} E_i = \sum_k (\Gamma_{ij}^k - \Gamma_{ji}^k) E_k.$$
  This is equivalent to the symmetry of Γ_{ij}^k in the lower indices:
  $$\Gamma_{ij}^k = \Gamma_{ji}^k.$$

Proof of the Fundamental Theorem (Cont'd)

- Property (4) is equivalent to

$$E_k g_{ij} = E_k(E_i, E_j) = (\nabla_{E_k} E_i, E_j) + (E_i, \nabla_{E_k} E_j).$$

Equivalently,

$$E_k g_{ij} = \sum_s (\Gamma^s_{ki} g_{sj} + \Gamma^s_{kj} g_{si}), \quad 1 \leq i, j, k \leq n.$$

Finally, we define

$$\Gamma_{ijk} = \sum_s \Gamma^s_{ij} g_{sk}.$$

Using the matrix (g^{ij}) inverse to (g_{ij}), we get

$$\Gamma^k_{ij} = \sum \Gamma_{ijs} g^{sk}.$$

Thus, the n^3 C^∞ functions Γ^k_{ij} determine the n^3 C^∞ functions Γ_{ijk} and conversely.

Proof of the Fundamental Theorem (Cont'd)

- If we write $E_k g_{ij} = \frac{\partial g_{ij}}{\partial x^k}$, i.e., if we consider g_{ij} as functions of the local coordinates, the properties obtained above become:

  (3') $\Gamma_{ijk} = \Gamma_{jik}$;

  (4') $\frac{\partial g_{ij}}{\partial x^k} = \Gamma_{kij} + \Gamma_{kji}$.

  In summary, suppose we are given a Riemannian connection on M, covered by a single coordinate neighborhood.

  If a Riemannian connection ∇ exists, it determines n^3 functions Γ_{ijk} of class C^∞ which satisfy Properties (3') and (4').

  Conversely, we may check, by reversing these steps, that any such functions determine a C^∞ Riemannian connection on M.

  Thus, the theorem is completely established by the following lemma.

Final Step in Proving Existence

Lemma

Let W be an open subset of $\mathbb{R}^n$. Let (g_{ij}) be a symmetric, positive definite matrix whose entries are C^∞ functions on W. Then, on W, there exists a unique family of C^∞ functions

$$\Gamma_{ijk}(x), \quad 1 \leq i, j, k \leq n,$$

satisfying the two sets of equations:

(3') $\Gamma_{ijk} = \Gamma_{jik}$;

(4') $\frac{\partial g_{ij}}{\partial x^k} = \Gamma_{kij} + \Gamma_{kji}$.

- Write Equation (4') twice more, each time permuting i, j, k cyclically,

$$\frac{\partial g_{ij}}{\partial x^k} = \Gamma_{kij} + \Gamma_{kji}; \quad \frac{\partial g_{jk}}{\partial x^i} = \Gamma_{ijk} + \Gamma_{ikj}; \quad \frac{\partial g_{ki}}{\partial x^j} = \Gamma_{jki} + \Gamma_{jik}.$$

Final Step in Proving Existence (Cont'd)

- Then subtract the second of these equations from the sum of the first and third,

$$\frac{\partial g_{ij}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} = \Gamma_{kij} + \Gamma_{kji} - \Gamma_{ijk} - \Gamma_{ikj} + \Gamma_{jki} + \Gamma_{jik}.$$

Using Equation (3'), $\Gamma_{ijk} = \Gamma_{jik}$, we get

$$\frac{\partial g_{ij}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} = 2\Gamma_{jki}.$$

So we get the unique solutions

$$\Gamma_{jki} = \frac{1}{2}\left(\frac{\partial g_{ij}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j}\right).$$

This completes the last step in the proof of the fundamental theorem.

Formula for $(\nabla_X Y)_p$

- Suppose that U, φ is a local coordinate system.
- Let $x^1, \ldots, x^n$ be local coordinates.
- Let $E_1, \ldots, E_n$ be the coordinate frames.
- Let

  $$Y = \sum a^k E_k$$

  be the expression on U of the vector field Y.
- Let $p \in U$ and

  $$X_p = \sum b^k E_{kp}.$$
- The following corollary supplies a formula for $\nabla_X Y$ on U.

Formula for $(\nabla_X Y)_p$ (Cont'd)

Corollary

For each $p \in U$, using the above notation, we have

$$(\nabla_X Y)_p = \nabla_{X_p} Y = \sum_k \left(\sum_j b^j \frac{\partial a^k}{\partial x^j} + \sum_{i,j} \Gamma^k_{ij} a^i b^j \right) E_k,$$

with

$$\Gamma^k_{ij} = \frac{1}{2} g^{ks} \left(\frac{\partial g_{si}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^s} + \frac{\partial g_{js}}{\partial x^i} \right).$$

- As we have seen in the proof, $(\nabla_X Y)_U$ is the same as $\nabla^U_{X_U} Y_U$.
  The latter is ∇^U on X, Y, restricted to U.
  For this reason we use the same symbol ∇ for all cases.
  The formula of the corollary follows at once from applying Properties (1) and (2) defining a connection to $\nabla_{\sum b^j E_j}(\sum a^k E_k)$.

Formula for $(\nabla_X Y)_p$ (Cont'd)

- The preceding formula is the same formula we obtained earlier for a manifold M in Euclidean space.
- In fact we have an obvious corollary of the uniqueness of ∇.

Corollary

In the case of an imbedded (or immersed) manifold in Euclidean space, the differentiation defined in a previous theorem depends only on the Riemannian metric induced by the imbedding (but is otherwise independent of the imbedding).

Remark

- In the preceding sections we used the concept of differentiation of vector fields along curves $\frac{dY}{dt}$ to define $\frac{DY}{dt}$ and then $\nabla_X Y$ on submanifolds of $\mathbb{R}^n$.
- In this section we showed quite independently of the earlier discussion that there is a uniquely determined Riemannian connection ∇ on every Riemannian manifold M.
- Using this result we come full circle.
- We define, for a vector field Y and curve $p(t)$ on M, the covariant derivative $\frac{DY}{dt}$ of $Y(t) = Y_{p(t)}$ by

$$\frac{DY}{dt} = \nabla_{\frac{dp}{dt}} Y.$$

- Let Y be given locally by

$$Y = \sum b^k(x) E_k.$$

Remark (Cont'd)

- If $p(t)$ is given by $x(t) = (x^1(t), \ldots, x^n(t))$, we have

$$X_{p(t)} = \sum \dot{x}^j(t) E_j = \frac{dp}{dt}.$$

- By the corollary, we can rederive formula

$$\frac{DY}{dt} = \sum_{k=1}^{n} \left(\frac{db^k}{dt} + \sum_{i,j=1}^{n} \Gamma_{ij}^k(x(t)) b^i(x(t)) \frac{dx^j}{dt} \right) E_k.$$

- $\frac{db^k}{dt}$ depend only on the values of the components $b^1, \ldots, b^n$ of Y along the curve.
- So the formula is valid when Y is defined only at points of the curve.
- Of course on any interval of the curve, Y may be extended to a vector field on M.
- But $\frac{DY}{dt}$ is independent of the extension by the displayed formula.

Constant Vector Fields

- A vector field Y on M is said to be **constant** if, for all $p \in M$ and $X_p \in T_p(M)$,
$$\nabla_X Y = 0.$$
- In general there do not exist such vector fields, even on small open subsets of M.
- On the other hand, consider a differentiable curve
$$p(t), \quad 0 \leq t \leq T.$$
- Then, there is be a vector field
$$X(t) = X_{p(t)}$$
constant or **parallel** along $p(t)$ (by which we mean $\frac{DX}{dt} \equiv 0$).

Constant Vector Fields

Theorem

Let $p = p(0)$, the initial point of the curve $p(t)$, $0 \leq t \leq T$.
Let $X_p \in T_{p(0)}(M)$ be given arbitrarily.
Then there exists a unique constant vector field $X_{p(t)}$ along $p(t)$, such that $X_{p(0)}$ has the given value.
Suppose $E_{1p}, \ldots, E_{np}$ is an orthonormal frame at $p(0)$.
Then there is a unique, parallel field of orthonormal frames on $p(t)$ which coincide with the given one at $p = p(0)$.

- The proof depends on a previous existence theorem which was not fully proved.
- Moreover, we need a special fact about systems which are linear in the unknown functions.

Partial Proof

- To prove the existence and uniqueness of $X(t) = X_{p(t)}$, it is enough to demonstrate it for arcs of $p(t)$ lying in single coordinate neighborhoods.

  This is because:

  - The curve can be partitioned into a finite number of such arcs;
  - $X(t)$ can then be defined on each in turn beginning with $t = 0$..

  Now suppose that U, φ is such a coordinate neighborhood.

  Suppose U, φ contains $p(t)$, for $c \leq t \leq d$, and that $X_{p(c)}$ is given.

  We wish to determine $X_{p(t)} = \sum a^k(t)E_k$ so that it is parallel.

  By virtue of the formula in the preceding remark, this occurs if and only if

$$\frac{da^k}{dt} = -\sum_{i,j} \Gamma_{ij}^k a^k \frac{dx^j}{dt}, \quad k = 1, \ldots, n.$$

Partial Proof (Cont'd)

- In this system of ordinary differential equations:
  - The $a^k(t)$ are unknown except at $t = c$;
  - The Γ_{ij}^k depend on t through $x(t)$.

  Thus, $a^k(t)$ satisfy a system of first-order equations.

  We know the system has a unique solution satisfying arbitrarily given initial conditions $X_{p(c)} = \sum a^k(c)E_k$.

  So $a^k(t)$ are defined and unique for some interval of values of t.

  Moreover, they are necessarily C^r if the curve is C^r.

  We need to know that the solutions $a^k(t)$ are defined for all values of t in the given interval $c \leq t \leq d$.

  This is so (as mentioned above) because the equations are linear.

  That is, the right-hand sides are linear in the unknown functions $a^i(t)$.

Partial Proof (Cont'd)

- The second part of the proposition is a consequence of the first and of the inner product rule for differentiation.
  We extend each of the $E_{ip(0)}$ to a parallel vector field $E_{ip(t)}$.
  Then, by definition,
  $$\frac{DE_i}{dt} \equiv 0, \quad 1 \leq i \leq n.$$
  Differentiating (E_i, E_j), we find that
  $$\frac{D}{dt}(E_i, E_j) = \left(\frac{DE_i}{dt}, E_j\right) + \left(E_i, \frac{DE_j}{dt}\right) = 0.$$
  Thus (E_i, E_j) is for each i, j a constant function along $p(t)$.
  At $p(0)$, we have
  $$(E_i, E_j) = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j. \end{cases}$$
  So the same is true everywhere on $p(t)$.

Remark

- We remark that it is sufficient for the curve to be piecewise differentiable, for then we can move X_p along each piece separately.
- Therefore, it follows from this theorem that, given a piecewise differentiable curve $p(t)$, there exists an isomorphism, in fact isometry,

$$\tau_t : T_{p(0)}(M) \to T_{p(t)}(M)$$

  determined by the condition that $\tau_t(X_{p(0)})$ be a parallel (constant) vector field along $p(t)$.
- It is clear from our initial discussion of $\frac{dX}{dt}$ along a curve $p(t)$ in Euclidean space that this would enable us to define the derivative of vector fields along curves on a Riemannian manifold M by comparing vectors at different points of the curve.
- The notion of parallel displacement along curves is sometimes taken as the starting point in studying differentiation on manifolds.

Subsection 4

Addenda to the Theory of Differentiation on a Manifold

Order of Differentiation

- It is a standard theorem of Advanced Calculus that second-order partial derivatives are independent of the order of differentiation,

$$\frac{\partial}{\partial x^i}\left(\frac{\partial f}{\partial x^j}\right) = \frac{\partial}{\partial x^j}\left(\frac{\partial f}{\partial x^i}\right).$$

- For functions on manifolds the analogous property $X(Yf) = Y(Xf)$ does not hold in general.
- Indeed $[X, Y]$ measures the extent by which it fails,

$$[X, Y]f = X(Yf) - Y(Xf).$$

- The property still holds if $X = E_i$ and $Y = E_j$
- Allowing $\widetilde{f}$ to denote the expression for the function on M in local coordinates $x^1, \ldots, x^n$, $E_k f$ may be identified with $\frac{\partial \widetilde{f}}{\partial x^k}$.

Generalization to an Arbitrary Manifold

- So, in the case of functions, interchangeability of order of differentiation is measured by an interesting object $[X, Y]$.
- It is natural to study the same question for ∇_X and ∇_Y derivatives of a vector field Z on M with respect to vector fields X, Y.
- We may show by example that, in general,

$$\nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) \neq 0.$$

- Hence, it determines a vector field on M.
- $\nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z)$ may be thought of as analogous to $[X, Y]$.
- An even more important expression, which involves also the measure of noninterchangeability of derivatives of functions $[X, Y]$, is the following related vector field, denoted by $R(X, Y)Z$ or $R(X, Y) \cdot Z$,

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

Properties of $R(X,Y)\cdot Z$

- It is readily verified that the formula

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

  defines a multilinear mapping of $\mathfrak{X}(M)\times\mathfrak{X}(M)\times\mathfrak{X}(M)\to\mathfrak{X}(M)$.
- That is, $R(X,Y)\cdot Z$ is $\mathbb{R}$-linear in each variable.
- From another point of view, in this expression, $R(X,Y)$ is an operator, determined by the vector fields X and Y, and assigning to each vector field Z a new C^∞-vector field $R(X,Y)\cdot Z$.
- Note that if $[X,Y]=0$, as is the case when $X=E_i$, $Y=E_j$ are vectors of a coordinate frame, then

$$R(X,Y)\cdot Z=\nabla_X(\nabla_Y Z)-\nabla_Y(\nabla_X Z).$$

- It follows that, if $R(X,Y)=0$ on M, then ∇_{E_i} and ∇_{E_j} are interchangeable for all Z.

Properties of $R(X, Y) \cdot Z$

Theorem

At any point p, the vector $(R(X, Y) \cdot Z)_p$ depends only on X_p, Y_p, Z_p, the values of the three vector fields at p, and not their values in a neighborhood or on M. Thus,

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

assigns to each pair of vectors $X_p, Y_p \in T_p(M)$ a linear transformation

$$R(X_p, Y_p) : T_p(M) \to T_p(M).$$

In fact, $(X_p, Y_p) \to R(X_p, Y_p)$ is a linear mapping of $T_p(M) \times T_p(M)$ into the space of operators on $T_p(M)$.

Properties of $R(X, Y) \cdot Z$ (Cont'd)

- By definition, $R(X, Y) \cdot Z$ depends $\mathbb{R}$-linearly on each of the three arguments X, Y, Z.

  Let f be a C^∞ function on M (not necessarily constant).

  Then, by direct computation, we have

  $$R(fX, Y) \cdot Z = R(X, fY) \cdot Z = R(X, Y) \cdot fZ = fR(X, Y) \cdot Z.$$

  Let U, φ is a coordinate neighborhood.

  Let $(x^1, \ldots, x^n)$ denote the local coordinates.

  Let $E_1, \ldots, E_n$ be the coordinate frames.

Properties of $R(X,Y)\cdot Z$ (Cont'd)

- Suppose that

$$X=\sum \alpha^i E_i, \quad Y=\sum \beta^j E_j, \quad Z=\sum \gamma^k E_k.$$

By the remarks above,

$$R(X,Y)\cdot Z=\sum_{i,j,k}\alpha^i\beta^j\gamma^k R(E_i,E_j)\cdot E_k.$$

So at a given point p of U, the right-hand side involves:
  - $R(E_i,E_j)\cdot E_k$, which is independent of the vector fields;
  - The values of $\alpha^i,\beta^i,\gamma^i$ only at the point p itself, not at nearby points.

This proves the theorem.
- Note that we used only Properties (1) and (2) of the connection ∇.
- The next fact, on the other hand, uses the Riemannian metric.

A Covariant Tensor of Order Four

Corollary

The formula $R(X,Y,Z,W) = (R(X,Y)\cdot Z, W)$ defines a C^∞-covariant tensor of order 4. This tensor depends only on the Riemannian metric on M. That is, if M_1, M_2 are Riemannian manifolds and $F : M_1 \to M_2$ is an isometry, then

$$F^* R_2 = R_1.$$

- $R(X_p, Y_p)\cdot Z_p$ is defined as an element of $T_p(M)$, for any $p \in M$. So its inner product

$$(R(X_p, Y_p)\cdot Z_p, W_p)$$

with any $W_p \in T_p(M)$ is a well-defined real number.

A Covariant Tensor of Order Four (Cont'd)

- Thus, for each p,

$$R_p(X_p, Y_p, Z_p, W_p) = (R(X_p, Y_p) \cdot Z_p, W_p)$$

defines a multilinear function of four variables on $T_p(M)$.
That is, $R_p(X_p, Y_p, Z_p, W_p) \in \mathscr{T}^4(T_p(M))$.
Both inner product and $R(X, Y) \cdot Z$ are C^∞ for $X, Y, Z, W \in \mathfrak{X}(M)$.
Consequently,

$$R_p(X_p, Y_p, Z_p, W_p) = (R(X_p, Y_p) \cdot Z_p, W_p)$$

defines a C^∞-tensor field.
We have defined an isometry of Riemannian manifolds to be a diffeomorphism which preserves the Riemannian metric.
That is,

$$F_* : T_p(M_1) \to T_{F(p)}(M_2)$$

preserves inner products (and is an isomorphism onto).

A Covariant Tensor of Order Four (Cont'd)

- Parenthetically, if we do not suppose that the C^∞ mapping F is one-to-one onto, but only that F_* is onto and preserves inner products, then it is called a **local isometry**.

  This is an isometry on some neighborhood of each point (for example, covering spaces).

  The last statement is valid for local isometries also.

  Now ∇ is uniquely determined by the Riemannian metric.

  So F_* preserves the connection.

  More precisely

  $$F_*(\nabla^1_X Y) = \nabla^2_{F_*(X)} F_*(Y).$$

  From this we deduce that

  $$R_2(F_* X, F_* Y) \cdot F_* Z = R_1(X, Y) \cdot Z.$$

  Since inner products are preserved, this implies $F^* R_2 = R_1$.

Curvature Operator and Riemann Curvature Tensor

Definition

The operator $R(X,Y)$ is called the **curvature operator**.
The tensor $R(X,Y,Z,W)$ is called the **Riemann curvature tensor**.

- It is not difficult to see that each one determines the other.
- Let $E_1,\dots,E_n$ be a field of frames on U, an open set of M.
- Then the Riemann curvature tensor is uniquely determined on U by either of the n^4 sets of functions $R^j_{ik\ell}$ or $R_{ijk\ell}$ defined by the equations

$$\begin{aligned} R(E_k,E_\ell)\cdot E_i &= \sum_j R^j_{ik\ell}E_j; \\ R(E_k,E_\ell,E_i,E_j) &= R_{ijk\ell}=\sum_s g_{js}R^s_{ik\ell}, \quad g_{js}=(E_j,E_s). \end{aligned}$$

Setup for Connection Forms

- Let U be an open subset of a manifold M.
- Suppose U has defined over it a field of C^∞ frames $E_1, \ldots, E_n$.
- The most usual case is when these are the coordinate frames of a coordinate neighborhood U, φ.
- However, in the case of a Riemannian manifold, which is our present interest, we might find it convenient to consider a neighborhood with orthonormal frames.
- Corresponding to $E_1, \ldots, E_n$, we have at each $p \in U$ the dual basis $\theta^1, \ldots, \theta^n$ of $T_p^*(M)$, characterized by

$$\theta^i(E_j) = \delta^i_j.$$

- It is a field of dual coframes on U and is clearly C^∞.
- Conversely, if $\theta^1, \ldots, \theta^n$ are given, then $E_1, \ldots, E_n$ are determined.

Connection Forms

- In defining $\nabla_X Y$ on a manifold so as to satisfy properties (1) and (2), we saw that it is enough to know $\nabla_{E_i} E_j$.
- Then $\nabla_X Y$ may be computed.
- We obtained
$$\nabla_{E_i} E_j = \sum_k \Gamma_{ij}^k E_k,$$
where the Γ_{ij}^k were determined above.
- If a connection is given, so that Γ_{ij}^k are known on U, then we may define n^2 one-forms θ_j^k by
$$\theta_j^k = \sum_\ell \Gamma_{\ell j}^k \theta^\ell.$$

Connection Forms (Cont'd)

- Conversely, given these one-forms, we have
$$\Gamma_{ij}^k = \theta_j^k(E_i).$$
- Hence the $\nabla_{E_i} E_j$ are determined.
- This determines also the connection.
- Indeed one checks at once that
$$\nabla_X E_j = \sum_k \theta_j^k(X) E_k.$$
- That is, the values of the forms $\theta_j^1, \ldots, \theta_j^n$ on X are the components of $\nabla_X E_j$ relative to the given frames.
- Therefore, given U and $\theta^1, \ldots, \theta^n$, a field of coframes on U, then the connection is determined on U by the n^2 forms θ_j^k.
- The forms θ_j^k are called the **connection forms**.

Fundamental Theorem: Neighborhoods with Frame Fields

- We have the following restatement of the Fundamental Theorem of Riemannian Geometry in terms of forms.
- However, we restrict ourselves only to the case in which the manifold is covered by a neighborhood on which is defined a frame field.

Theorem

Let M be a Riemannian manifold such that it has a covering by a C^∞ field of coframes $\theta^1, \ldots, \theta^n$. Then there exists a uniquely determined set of n^2 C^∞ one-forms

$$\theta_j^k, \quad 1 \le j, k \le n,$$

on M satisfying the two equations:

(i) $d\theta^i - \sum_j \theta^j \wedge \theta_j^i = 0$;

(ii) $dg_{ij} = \sum_k (\theta_i^k g_{kj} + \theta_j^k g_{ki})$, where $g_{ij} = (E_i, E_j)$, with $E_1, \ldots, E_n$ the uniquely determined field of frames dual to $\theta^1, \ldots, \theta^n$.

Fundamental Theorem (Cont'd)

Theorem (Cont'd)

The forms θ_j^k so determined define the Riemannian connection satisfying Properties (1)-(4) of the Fundamental Theorem by the formulas:

(iii) $\nabla_X E_j = \sum \theta_j^k(X) E_k$;

(iv) $\nabla_X (fY) = (Xf)Y + f\nabla_X Y$, for $f \in C^\infty(U)$.

Conversely, the Riemannian connection determines θ_j^k, as explained above, and these θ_j^k satisfy Properties (i) and (ii).

Case of Orthonormal Frame Fields

- If we recall that a Riemannian manifold M of the type described may be covered by an orthonormal frame field $E_1, \ldots, E_n$ with $g_{ij} = (E_i, E_j) = \delta_{ij}$, then we have a nicer version of the above.
- In this case we denote θ^i by ω^i and θ_j^k by ω_j^k.
- Using $g_{ij} \equiv \delta_{ij}$ (hence $dg_{ij} = 0$), we obtain

Corollary

Let M be a Riemannian manifold which has a covering by a field $\omega^1, \ldots, \omega^n$ of coframes whose dual frames $E_1, \ldots, E_n$ are orthonormal. Then there exists a unique set of n^2 one-forms ω_j^k, $1 \leq j, k \leq n$, on M satisfying:

(i) $d\omega^i - \sum_j \omega^j \wedge \omega_j^i = 0$;

(ii) $\omega_j^k + \omega_k^j = 0$.

These ω_j^k determine the Riemannian connection (as above) and conversely.

Remark and Preview

- θ_j^k are uniquely determined by $\theta^1, \ldots, \theta^n$ and the Riemannian metric, that is by the coframe field and the metric.
- Thus, the exterior derivatives $d\theta_j^k$ are also uniquely determined.
- The same holds for their expressions as linear combinations of the basis

$$\theta^i \wedge \theta^j, \quad 1 \leq i < j \leq n,$$

  of two-forms on the domain U of $\theta^1, \ldots, \theta^n$.
- We shall see in the next chapter that the coefficients in these linear combinations determine the components of the curvature tensor.

Subsection 5

Geodesic Curves on Riemannian Manifolds

Geodesics

- Consider a curve on M,

$$p(t), \quad a < t < b.$$

- Let $\frac{dp}{dt}$ be its velocity vector, defined for $a < t < b$.
- We assume $p(t)$ is of class C^2 at least.

Definition

The (parametrized) curve $p(t)$ is said to be a **geodesic** if its velocity vector is constant (parallel). That is, if it satisfies the condition

$$\frac{D}{dt}\left(\frac{dp}{dt}\right) = 0, \quad a < t < b.$$

$\frac{D}{dt}\left(\frac{dp}{dt}\right) = 0$ is called the **equation of a geodesic**.

Examples

- As we saw previously, when $M = \mathbb{R}^n$, with its usual metric, this implies that the curve is a straight line.
- But for a submanifold of $\mathbb{R}^n$ this can mean something quite different.
- An example is the great circles on $S^{n-1} \subseteq \mathbb{R}^n$.

The Parameter on a Geodesic

- The fact that a curve is a geodesic depends both on its shape and its parametrization.
- Consider a (geometric) straight line in $\mathbb{R}^2$ given parametrically by

$$x^1 = t^3, \quad x^2 = t^3.$$

- We write $p(t) = (t^3, t^3)$.
- Then

$$\frac{dp}{dt} = 3t^2 \frac{\partial}{\partial x^1} + 3t^2 \frac{\partial}{\partial x^2}.$$

- Now $\frac{D}{dt} = \frac{d}{dt}$ in $\mathbb{R}^2$.
  So we have

$$\frac{D}{dt}\left(\frac{dp}{dt}\right) = \frac{D}{dt}\left(3t^2 \frac{\partial}{\partial x^1} + 3t^2 \frac{\partial}{\partial x^2}\right) = 6t \frac{\partial}{\partial x^1} + 6t \frac{\partial}{\partial x^2} \neq 0.$$

- Therefore, this curve is not a geodesic, although the path traversed is the line $x^1 = x^2$.

Permissible Parametrizations for Geodesics

Lemma

Let $p(t)$, $a < t < b$, be a nontrivial geodesic. Let t' be a new parameter. With respect to t' the curve will be a geodesic if and only if

$$t = ct' + d, \quad c \neq 0, \quad d \text{ constant}.$$

In particular, the arclength is always such a parameter.

- Introduce a new parameter t' by $t = ct' + d$, $c \neq 0$.
  Then $\frac{dp}{dt'} = c\frac{dp}{dt}$.
  So we get
  $$\frac{D}{dt'}\left(\frac{dp}{dt'}\right) = c^2 \frac{D}{dt}\left(\frac{dp}{dt}\right) = 0.$$
  So the curve remains a geodesic relative to t'.

Permissible Parametrizations for Geodesics (Cont'd)

- Now let s be arclength measured from a point $p(t_0)$ on the curve.

  Then
  $$\frac{ds}{dt} = \left\| \frac{dp}{dt} \right\|.$$
  Now $\frac{dp}{dt}$ is constant along the curve.

  By the inner product rule for derivatives, its length $\|\frac{dp}{dt}\|$ is constant.

  If $\|\frac{dp}{dt}\|$ is identically zero, then $p(t)$ is a single point and $s = 0$.

  Otherwise, $\frac{ds}{dt} = \|\frac{dp}{dt}\| = c$, a nonzero constant.

  So $s = ct + d$.

  This means that the curve is a geodesic when parametrized by arclength.

  Since any other permissible parameter is related to arclength by a similar (linear) relation, any two parameters are linearly related.

Observations

- The equation of a geodesic imposes only a local condition on the curve.

  If each point of a curve C has a neighborhood in which it may be written in the form
  $$p(t), \quad a < t < b,$$
  with $\frac{D}{dt}(\frac{dp}{dt}) = 0$, then it is a geodesic.

  Use arclength from some fixed point as parameter on all of C.

  It must satisfy the equation $\frac{D}{ds}(\frac{dp}{ds}) = 0$ over its entire length.
- The property of being a geodesic is preserved by isometries.

  This is because covariant differentiation is preserved.

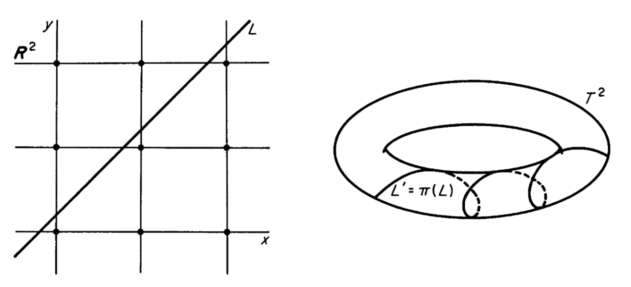
  As a result, so is parallelism of a vector field (for example, $\frac{dp}{dt}$) along a curve.

Example

- Let $\pi : \mathbb{R}^2 \to T^2$ be the standard covering.
- We take $\mathbb{R}^2$ with its usual Riemannian metric.
- Since the covering transformations are translations, they are isometries of $\mathbb{R}^2$.
- It follows that we may define on T^2 a Riemannian metric which makes the projection π a local isometry, meaning that π_* is an isometry of each tangent space $T_p(\mathbb{R}^2)$ onto $T_{\pi(p)}(T^2)$.
- With this metric the geometry of T^2 is locally equivalent to that of Euclidean space.
- This Riemannian metric should not be confused with the metric induced on a torus imbedded in $\mathbb{R}^3$ by the standard Riemannian metric of $\mathbb{R}^3$.
- Combining the two preceding observations, it follows that even a local isometry, as, e.g., this map π, carries geodesics onto geodesics.

Example (Cont'd)

- Thus, the images of straight lines of $\mathbb{R}^2$ on T^2 are geodesics of T^2.



  - Lines of rational slope map to closed geodesics on T^2;
  - Lines of irrational slope do not - they are dense on T^2.
- By contrast, in $\mathbb{R}^2$, geodesics can be neither closed curves nor dense.
- Thus "straight lines", even on spaces locally isometric to Euclidean space, present some fascinating variations from what we might expect.

Geodesics and Local Coordinates

- Let M be a Riemannian manifold.
- Let U, φ be a connected coordinate neighborhood.
- Let $(x^1, \ldots, x^n)$ be local coordinates.
- Then the equation of a geodesic $\frac{D}{dt}(\frac{dp}{dt}) = 0$ is equivalent to the system of second-order differential equations

$$\frac{d^2x^k}{dt^2} + \sum_{i,j=1}^{n} \Gamma_{ij}^k(x) \frac{dx^i}{dt} \frac{dx^j}{dt} = 0, \quad k = 1, \ldots, n.$$

- A solution is a curve given in local coordinates by n functions $(x^1(t), \ldots, x^n(t))$ which satisfy the system.
- As usual let $E_1, \ldots, E_n$ denote the coordinate frames.
- Using the Existence Theorem, we prove the existence and uniqueness of a geodesic through each $p \in U$ with given tangent direction at p.
- Then we study its dependence on p and the tangent direction.

Existence and Uniqueness of Solution

Lemma

Let $q \in U$. Consider the system of second-order differential equations

$$\frac{d^2x^k}{dt^2} + \sum_{i,j=1}^{n} \Gamma_{ij}^k(x) \frac{dx^i}{dt} \frac{dx^j}{dt} = 0, \quad k = 1, \ldots, n.$$

We can find a neighborhood V of q, with $V \subseteq U$, and positive numbers r, δ, such that, for each $p \in V$ and each tangent vector

$$X_p = \sum b^i E_i, \text{ with } \|X_p\| < r,$$

there exists a unique solution $(x^1(t), \ldots, x^n(t))$ of the system, defined for $-\delta < t < \delta$, which satisfies

$$x^i(0) = x^i(p), \quad \dot{x}^i(0) = b^i, \quad i = 1, \ldots, n.$$

Existence and Uniqueness of Solution (Cont'd)

Lemma (Cont'd)

Let $p(t) = \varphi^{-1}(x^1(t), \ldots, x^n(t))$, as just defined.
Then $p(t) \in U$ for $|t| < \delta$.

- Consider the system of $2n$ first-order ordinary differential equations

$$\begin{array}{rcll} \frac{dx^k}{dt} & = & y^k, & k = 1, \ldots, n, \\ \frac{dy^k}{dt} & = & -\sum_{i,j=1}^n \Gamma_{ij}^k(x) y^j y^k, & k = 1, \ldots, n, \end{array}$$

defined on the open subset

$$W = \varphi(U) \times \mathbb{R}^n \subseteq \mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}.$$

The right sides are C^∞ functions of $(x, y) = (x^1, \ldots, x^n; y^1, \ldots, y^n)$ on W.

Existence and Uniqueness of Solution (Cont'd)

- By the Existence Theorem for Ordinary Differential Equations, for each point in W, there exists a $\delta > 0$ and a neighborhood V of the point, such that, given

$$(a; b) = (a^1, \ldots, a^n; b^1, \ldots, b^n) \in V,$$

there are $2n$ unique functions

$$x^k = f^k(t, a; b), \quad y^k = g^k(t, a; b), \quad k = 1, \ldots, n,$$

and $|t| < \delta$, satisfying the system of equations and the initial conditions

$$f^k(0, a; b) = a^k, \quad g^k(0, a; b) = b^k, \quad k = 1, \ldots, n.$$

These functions are C^∞ in all variables and have values in W.

Existence and Uniqueness of Solution (Cont'd)

- If $p \in U$, we consider

$$(\varphi(p); 0) = (x^1(p), \ldots, x^n(p); 0, \ldots, 0) \in W.$$

Then there is a $\delta > 0$ and a neighborhood $\widetilde{V}$ of $(\varphi(p), 0)$ as described.

This neighborhood may be chosen to be of the form

$$\varphi(V) \times B^n_{r'}(0),$$

for some V, with:

- $\overline{V} \subseteq U$ compact;
- $r' > 0$.

Since $\overline{V}$ is compact, we may find a number $r > 0$, such that

$$\left(\sum g_{ij}(x) b_i b_j\right)^{1/2} = \|X_p\| < r \text{ and } p \in V \text{ imply } \left(\sum (b^i)^2\right)^{1/2} < r'.$$

This follows from inequalities used in the proof of a previous theorem.

Existence and Uniqueness of Solution (Cont'd)

- By the special nature of the preceding system,

$$\frac{df^k}{dt} = g^k.$$

Hence,

$$\frac{d^2 f^k}{dt^2} = -\sum_{i,j} \Gamma_{ij}^k \frac{df^i}{dt}\frac{df^j}{dt}.$$

I.e., $x^k(t) = f^k(t, a; b)$ are solutions of the system of equations.

Therefore, they are the equations in local coordinates of geodesics satisfying

$$x^k(0) = a^k, \quad \left.\frac{dx^k}{dt}\right|_{t=0} = b^k, \quad k = 1, \ldots, n.$$

Existence and Uniqueness of Solution (Cont'd)

- Finally, according to the Existence Theorem cited, the image of $I_\delta \times \widetilde{V}$ under the map

$$(t, a, b) \to (f^1(t, a; b), \ldots, f^n(t, a; b); g^1(t, a; b), \ldots, g^n(t, a; b))$$

is in W.

This proves that

$$p(t) = \varphi^{-1}(f(t, a; b)) \in U.$$

Existence and Uniqueness of Open Geodesic Arcs

Corollary

Let M is a Riemannian manifold and $p \in M$.
Let Y_p a nonzero tangent vector at p.
Then there is a $\lambda > 0$ and a geodesic curve $p(t)$ on M, defined on some interval $-\delta < t < \delta$, $\delta > 0$, such that

$$p(0) = p, \quad \left.\frac{dp}{dt}\right|_{t=0} = \lambda Y_p.$$

Any two geodesic curves satisfying these two initial conditions coincide in a neighborhood of p.

- Take a neighborhood U, φ of p.

  Choose $\lambda > 0$ so that $\|\lambda Y_p\| < r$, as in the first lemma.

  Then apply the preceding (second) lemma.

Maximal Geodesics

- Recall the remark that "being a geodesic" is a local property of parametrized curves.
- So, if two geodesic curves C_1 and C_2 coincide (as sets) over some interval, then their union, suitably parametrized, is a geodesic.
- Further, we now see that, if two geodesics have a single point in common and are tangent at that point, then their union is a geodesic.
- So each geodesic is contained in a unique maximal geodesic.
- A **maximal geodesic** is one that is not a proper subset of any geodesic.
  - If it is parametrized by a parameter t, with $a < t < b$, then a and b (which can be $-\infty$ and/or $+\infty$) are determined by the curve and the choice of parameter.
  - It is not possible to extend the definition of $p(t)$ (with the given parameter) so as to include either of these values and so that it will still be a geodesic.

Extending Geodesics

- We shall be interested in determining conditions on M which ensure that $a = -\infty$ and $b = +\infty$ for every geodesic, or that every geodesic can be extended indefinitely in either direction.
- By a previous lemma this property would be independent of parameter.
- It is easy to see that this is not always possible.

  Let M be $\mathbb{R}^2$ with the origin removed.

  Then radial straight lines cannot be extended to the origin.
- However, given a geodesic through a point p, clearly we can always reparametrize it so that $p = p(0)$ and $p(t)$ is defined for $|t| < 2$, say.
- Making use of this fact, we modify a previous lemma slightly to obtain our basic existence and uniqueness theorem for geodesics.

Existence and Uniqueness of Geodesics

Theorem

Let M be a Riemannian manifold. Let U, φ be a coordinate neighborhood of M. Let $q \in U$. Then there exists a neighborhood V of q and an $\varepsilon > 0$, such that, if $p \in V$ and $X_p \in T_p(M)$, with $\|X_p\| < \varepsilon$, then there is a unique geodesic

$$p(t) = p(t, p, X_p), \quad -2 < t < +2,$$

with

$$p(0) = p, \quad \left.\frac{dp}{dt}\right|_{t=0} = X_p.$$

The mapping into M defined by

$$(t, p, X_p) \to p(t, p, X_p)$$

is C^∞ on the open set $|t| < 2$, $p \in V$, $\|X_p\| < \varepsilon$ and has its values in U.

Existence and Uniqueness of Geodesics (Cont'd)

- By a previous lemma, we may find a neighborhood V of q and numbers $r, \delta > 0$, such that, given any $p \in V$ and vector $X_p \in T_p(M)$, with $\|X_p\| < r$, then there is a geodesic

$$p(t), \quad |t| < \delta,$$

satisfying the initial conditions

$$p(0) = p, \quad \left.\frac{dp}{dt}\right|_0 = X_p.$$

If we change to a parameter $t = ct'$, $c \neq 0$ a constant, then

$$\widetilde{p}(t') = p(ct')$$

is again a geodesic, with:

  - $\widetilde{p}(0) = p$;
  - $\frac{d\widetilde{p}}{dt'} = \frac{dp}{dt}\frac{dt}{dt'} = c\frac{dp}{dt} \;\Rightarrow\; \left.\frac{d\widetilde{p}}{dt'}\right|_0 = cX_p$.

Existence and Uniqueness of Geodesics (Cont'd)

- If $\delta > 2$, we may use $\varepsilon = r$, and we have no more to prove.

  Suppose $\delta < 2$. Let $\varepsilon = \frac{\delta r}{2}$.

  Let $p \in V$ and X_p a tangent vector at p, with $\|X_p\| < \varepsilon$.

  We know from the choice of ε that $\|\frac{2X_p}{\delta}\| < r$.

  Thus, there is a geodesic $p(t)$ with:
  - $p(0) = p$;
  - $\frac{dp}{dt}|_0 = \frac{2X_p}{\delta}$, defined for $|t| < \delta$ at least.

  Consider the curve $\widetilde{p}(t') = p\frac{\delta t'}{2}$.

  It is again a geodesic and satisfies:
  - $\widetilde{p}(0) = p$;

    $\frac{d\widetilde{p}}{dt'}|_0 = \frac{\delta}{2} \times \frac{dp}{dt}|_0 = X_p$.

  Moreover it is defined for $|\frac{\delta t'}{2}| < \delta$, that is, for $-2 < t' < +2$.

  This completes the proof, since the last statement is already contained in the previous lemma.

Subsection 6

The Tangent Bundle, Exponential Mapping. Normal Coordinates

Summary of the Process

- We started with a second-order system of equations

$$\frac{d^2x^k}{dt^2} + \sum_{i,j=1}^{n} \Gamma_{ij}^k(x)\frac{dx^i}{dt}\frac{dx^j}{dt} = 0, \quad k = 1, \ldots, n.$$

- We passed to a first-order system

$$\begin{array}{rcll} \frac{dx^k}{dt} & = & y^k, & k = 1, \ldots, n, \\ \frac{dy^k}{dt} & = & -\sum_{i,j=1}^{n} \Gamma_{ij}^k y^i y^k, & k = 1, \ldots, n, \end{array}$$

- The method involved introducing new variables which corresponded to the components of tangent vectors at points of a coordinate neighborhood U, φ.

Summary of the Process (Cont'd)

- The vectors X_p, $p \in U$, are in one-to-one correspondence with points $(x; y)$ of the open set $W = \varphi(U) \times \mathbb{R}^n \subseteq \mathbb{R}^n \times \mathbb{R}^n$.
- The correspondence $\widetilde{\varphi}$ is given by

$$\widetilde{\varphi}(X_p) = (\varphi(p); y^1, \ldots, y^n),$$

  where:
  - $\varphi(p) = (x^1, \ldots, x^n)$ are the coordinates of p;
  - $X_p = \sum y^i E_{ip}$, with $E_1, \ldots, E_n$ the coordinate frames.
- The differential equations of geodesics were interpreted as a system of first-order differential equations on W.
- Like all such systems, they correspond to a vector field on W.

Freeing From Local Coordinates

- We would like to avoid working exclusively with local coordinates.
- So we think of W as the image under $\widetilde{\varphi}$ of a coordinate neighborhood $\widetilde{U}, \widetilde{\varphi}$ on a manifold.
- This is possible and requires that we define a manifold structure on the set of all tangent vectors at all points of M.
- We shall denote this structure by $T(M)$.
- When this is done,

$$T(M) = \{X_p \in T_p(M) : p \in M\} = \bigcup_{p \in M} T_p(M)$$

  becomes a space.
- It is in fact a C^∞ manifold, whose points are tangent vectors to M.

Freeing From Local Coordinates

- In view of the preceding remarks, we require the subset $\widetilde{U}$, consisting of all X_p, such that $p \in U$, to be a coordinate neighborhood, with:
  - $\widetilde{\varphi}$ as coordinate map;
  - W as image.
- That is,
$$\widetilde{\varphi} : \widetilde{U} \to W.$$
- This virtually dictates the choice of topology and differentiable structure.
- We denote by $\pi : T(M) \to M$ the natural mapping taking each vector to its initial point
$$\pi(X_p) = p.$$
- Then we have $\pi^{-1}(p) = T_p(M)$.

The Space $T(M)$

Lemma

Let M be a C^∞-manifold of dimension n. There is a unique topology on $T(M)$, such that, for each coordinate neighborhood U, φ of M:

- The set $\widetilde{U} = \pi^{-1}(U)$ is an open set of $T(M)$;
- $\widetilde{\varphi} : \widetilde{U} \to \varphi(U) \times \mathbb{R}^n$, defined as above, is a homeomorphism.

With this topology $T(M)$ is a topological manifold of dimension $2n$. Moreover, the neighborhoods $\widetilde{U}, \widetilde{\varphi}$ determine a C^∞-structure relative to which π is an (open) C^∞-mapping of $T(M)$ onto M.

- Let U, φ and U', φ' be coordinate neighborhoods on M, such that
$$U \cap U' \neq \emptyset.$$
Then $\widetilde{U} \cap \widetilde{U}' \neq \emptyset$.

The Space $T(M)$ (Cont'd)

- We compare:
  - The coordinates of $p \in U \cap U'$;
  - The components of any $X_p \in T_p(M)$ relative to the two coordinate systems.

  Suppose
  $$x'^i = f^i(x^1, \ldots, x^n), \quad i = 1, \ldots, n,$$
  are the formulas for change of coordinates $\varphi' \circ \varphi^{-1}$ on $U \cap U'$.
  By a previous corollary, the change of components are
  $$y'_i = \sum_{i=1}^n y^i \frac{\partial f^n}{\partial x^i}, \quad i = 1, \ldots, n.$$
  Thus, we obtain the formulas for change of coordinates in $\widetilde{U} \cap \widetilde{U}'$,
  $$\widetilde{\varphi}' \circ \widetilde{\varphi}^{-1}(x^1, \ldots, x^n; y^1, \ldots, y^n) = (f^1(x), \ldots, f^n(x); \textstyle\sum_{i=1}^n y^i \frac{\partial f^1}{\partial x^i}, \ldots, \sum_{i=1}^n y^i \frac{\partial f^n}{\partial x^i}).$$

The Space $T(M)$ (Cont'd)

- The maps
$$\widetilde{\varphi}' \circ \widetilde{\varphi}^{-1} : \widetilde{\varphi}(\widetilde{U} \cap \widetilde{U}') \to \widetilde{\varphi}'(\widetilde{U} \cap \widetilde{U}')$$
are diffeomorphisms.

  Skipping some steps, we turn to the dimension.

  In local coordinates, π corresponds to projection of $\mathbb{R}^n \times \mathbb{R}^n$ onto its first factor.

  Further, locally, on the domain $\widetilde{U}$ of each coordinate neighborhood of the type above, $T(M)$ is a product manifold.

  That is, as an open submanifold of $T(M)$, $\widetilde{U}$ is diffeomorphic to $\varphi(U) \times \mathbb{R}^n$.

  In the case of Euclidean space, U, φ may be taken to be all of $M = \mathbb{R}^n$ so that $T(\mathbb{R}^n)$ is diffeomorphic to $\mathbb{R}^n \times \mathbb{R}^n$.

  It is clear that for every manifold M, $\dim T(M) = 2\dim M$.

The Tangent Bundle

Definition

The space

$$T(M)$$

with the topology and C^∞ structure just defined is called the **tangent bundle** of M.

The mappng

$$\pi : T(M) \to M,$$

is the **natural projection**.

The Function Exp

- We define Exp, the exponential mapping.
- Its domain $\mathscr{D}$ is some subset of $T(M)$.
- The range of Exp is M itself.
- Thus $\text{Exp} : \mathscr{D} \to M$ maps a vector X_p to a point of M.
- Let U, φ be a coordinate neighborhood of M and $q \in U$.
- Choose a neighborhood V of q and $\varepsilon > 0$ are chosen as in a previous theorem.
- Then, for each X_p, with $p \in V$, and $\|X_p\| < \varepsilon$, or equivalently, in the open subset
$$\{X_p : p \in V, \|X_p\| < \varepsilon\} \subseteq T(M),$$
the geodesic $p(t)$, with $p(0) = p$ and $\frac{dp}{dt}|_0 = X_p$, is defined for $|t| < 2$.
- On this open set of $T(M)$ we define Exp as follows.

The Function Exp (Cont'd)

Definition

$\mathrm{Exp}X_p = p(1)$, that is, the image of X_p under the exponential mapping is defined to be that point on the unique geodesic determined by X_p at which the parameter takes the value $+1$.

- Thus each $q \in M$ has a neighborhood V, such that Exp is defined on the open subset

$$\{X_p : p \in V, \|X_p\| < \varepsilon\} \subseteq \pi^{-1}(V),$$

where ε depends on q and its neighborhood V.

The Domain of Definition

- We restate the information on $\mathscr{D}$.
- Let M_0 be the submanifold of $T(M)$ consisting of all zero vectors

$$0_p, \quad p \in M.$$

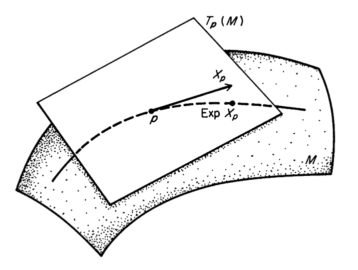
- Then $p \to 0_p$ maps M onto M_0 diffeomorphically.
- Moreover, $\pi : M_0 \to M$ is its inverse.
- The application of the same theorem then guarantees that the domain $\mathscr{D}$ of Exp contains an open neighborhood of M_0 in $T(M)$.

Geometric Interpretation of $\mathrm{Exp} X_p$

- We know $\|\frac{dp}{dt}\|$ is constant along a geodesic $p(t)$.
- So its length L from $p(0)$ to $p(1)$ is

$$L = \int_0^1 \left\|\frac{dp}{dt}\right\| dt = \int_0^1 \|X_p\| dt = \|X_p\|.$$

- Thus $\mathrm{Exp} X_p$ is the point on the unique geodesic $p(t)$ determined by X_p whose distance from p along the geodesic is the length of X_p.



A Property of Exp

Lemma

Assume that $q \in M$ and that $X_q \in T_q(M)$ for which $\mathrm{Exp} X_q$ is defined. Then $\mathrm{Exp} t X_q$ is defined at least for each t with $|t| < 1$.
Moreover,

$$q(t) = \mathrm{Exp} t X_q$$

is the geodesic through q at $t = 0$, with $\frac{dq}{dt}|_0 = X_q$.

- Let $q(t)$ be the unique geodesic with $q(0) = q$ and $\frac{dq}{dt}|_0 = X_q$.
  Then, by definition, $\mathrm{Exp} X_q = q(1)$.
  Given c, with $|c| < 1$, consider the geodesic $\widetilde{q}(t) = q(ct)$.
  We have $\widetilde{q}(0) = q$ and $\frac{d\widetilde{q}}{dt}|_{t=0} = cX_q$.
  This means that

  $$\mathrm{Exp} c X_q = \widetilde{q}(1) = q(c).$$

  Replacing c by t in this equality, we get the statement above.

Normal Neighborhood Theorem

Normal Neighborhood Theorem

Every point q of a Riemannian manifold M has a neighborhood N which is the diffeomorphic image under Exp_q of a star-shaped neighborhood $\widetilde{N}$ of the zero vector 0_q of the vector space $T_q(M)$.

- We revert to local coordinates U, φ around $q \in M$.

  Let $V \subseteq U$ and $\varepsilon > 0$ be as in a previous theorem.

  So, for $p \in V$ and $\|X_p\| < \varepsilon$, $\mathrm{Exp} X_p$ is defined.

  As in the proof of a previous lemma, the geodesic determined by p and X_p is given in local coordinates by

$$t \to (f^1(t, a; b), \ldots, f^n(t, a; b)),$$

  with:
  - $\varphi(p) = a = (a^1, \ldots, a^n)$;
  - $X_p = b^1 E_{1p} + \cdots + b^n E_{np}$.

Normal Neighborhood Theorem (Cont'd)

- This means that

$$\begin{aligned}\varphi(\mathrm{Exp}X_p) &= (f^1(1,a;b),\ldots,f^n(1,a;b));\\ \varphi(\mathrm{Exp}tX_p) &= (f^1(1,a;tb),\ldots,f^n(1,a;tb)), \quad |t|<1.\end{aligned}$$

The preceding lemma and the meaning of the functions $f^i(t,a;b)$ then give us the following identities, valid for $|t|<1$,

$$f^i(1,a^1,\ldots,a^n;tb^1,\ldots,tb^n) = f^i(t,a^1,\ldots,a^n;b^1,\ldots,b^n).$$

First, note that the f^i are C^∞ on their domain.

Hence, $X_p \to \mathrm{Exp}X_p$ is C^∞ on

$$\{X_p : p\in V, \|X_p\|<\varepsilon\}.$$

For brevity, we denote by Exp_q the restriction of Exp to $T_q(M)\cap\mathscr{D}$.

We may compute the Jacobian of Exp_q at $X_q = 0_q$.

Normal Neighborhood Theorem (Cont'd)

- Now q is fixed.

  $(a^1, \ldots, a^n)$ are constants.

  The Jacobian matrix at this point has as entries $\frac{\partial f^i}{\partial b^j}$ evaluated at $(1, a^1, \ldots, a^n, 0, \ldots, 0)$,

  $$\frac{\partial f^i}{\partial b^j} = \lim_{h \to 0} \frac{1}{h}(f^i(1, a; 0, \ldots, h, \ldots, 0) - f^i(1, a; 0, \ldots, 0)).$$

  We use the previously obtained identities, with $b^j = 1$, $b^k = 0$, for $k \neq j$, first with $t = h$, then with $t = 0$.

  Then, we get

  $$\begin{array}{rcl} \frac{\partial f^i}{\partial b^j} & = & \lim_{h \to 0} \frac{1}{h}(f^i(h, a; 0, \ldots, 1, \ldots, 0) - f^i(0, a; 0, \ldots, 1, \ldots, 0)) \\ & = & \dot{f}^i(0, a^1, \ldots, a^n; 0, \ldots, 1, \ldots, 0). \end{array}$$

Normal Neighborhood Theorem (Cont'd)

- Now
$$x^i = f^i(t, a^1, \ldots, a^n; 0, \ldots, 1, \ldots, 0), \quad i = 1, \ldots, n,$$
(with $b^j = 1$ and $b^k = 0$, if $k \neq j$), considered as functions of t, are the equations of the geodesic through q with E_{jq} as initial vector.

  So the Jacobian matrix reduces to the identity at $X_q = 0_q$.

  That is, we have
$$\frac{\partial f^i}{\partial b^j} = \delta^i_j.$$
So, for q fixed and for some $\varepsilon' < \varepsilon$, the mapping $X_q \to \mathrm{Exp} X_q$ is a diffeomorphism of the open set $\widetilde{N} = \{X_q : \|X_q\| < \varepsilon'\}$ of $T_q(M)$ onto an open set N containing $q = \mathrm{Exp} 0_q$.

  Retaining the notation Exp_q for Exp restricted to that part of its domain in $T_q(M)$, we obtain the result.

Normal Coordinate Neighborhoods

- We have defined $\widetilde{N}$ by $\|X_q\| < \varepsilon'$.
- The norm in $T_q(M)$ is given by the Riemannian metric.
- So we may choose an orthonormal basis of $T_q(M)$,

$$F_1, \ldots, F_n.$$

- Write

$$X_q = \sum_{i=1}^{n} y^i F_i.$$

- Then we have

$$\|X_q\| = \sum_{i=1}^{n} (y^i)^2.$$

Normal Coordinate Neighborhoods

- With these choices, the mapping

$$\psi : \mathrm{Exp}_q \left(\sum_{i=1}^{n} y^i F_i \right) \mapsto (y^1, \ldots, y^n)$$

takes the open neighborhood N of q diffeomorphically onto $B^n_{\varepsilon'}(0) \subseteq \mathbb{R}^n$.

Definition

The coordinate neighborhood N, ψ of q defined in this way is called a **normal coordinate neighborhood**.

Properties of Normal Coordinates

- Normal coordinates have special features that make them useful in the study of the geometry of the manifold.
- Of these the most important are the following.

(i) For all i, j,

$$g_{ij}(0) = \delta_{ij};$$

(ii) The equations of the geodesics through q take the form

$$y^i = a^i t, \quad i = 1, \ldots, n,$$

where a^i constants;

(iii) The coefficients of the connection vanish at q,

$$\Gamma_{ij}^k(0) = 0, \quad i, j, k = 1, \ldots, n.$$

Properties of Normal Coordinates (Cont'd)

- Statements (i) and (ii) are immediate consequences of the definition and a previous lemma.

  The third follows from the second.

  Consider $a^1, \ldots, a^n$ close to zero.

  Substitute the solutions

  $$y^i = a^i t$$

  in the equations of the geodesics.

  We get

  $$\sum_{ij} \Gamma_{ij}^k(0) a^i a^j = 0, \quad k = 1, \ldots, n.$$

A Strengthening

- Let U, φ be a coordinate neighborhood of $q \in M$.
- Let $E_1, \ldots, E_n$ denote the coordinate frames.
- Let $X_p = \sum b^i E_{ip}$ be the tangent vectors to $p \in U$.
- Let $\varphi(p) = (x^1, \ldots, x^n)$ be the local coordinates.
- We have shown that there exists a relatively compact neighborhood V of q, $\overline{V} \subseteq U$, and an $\varepsilon > 0$, such that $\text{Exp} X_p$ is defined and in U, for each X_p, with $p \in V$ and with $\|X_p\| < \varepsilon$.
- Then in local coordinates

$$\varphi(\text{Exp} X_p) = (f^1(1, x^1, \ldots, x^n; b^1, \ldots, b^n), \ldots, \\ f^n(1; x^1, \ldots, x^n, b^1, \ldots, b^n)),$$

  with $f^i(t, x, b)$ being C^∞ in all variables.
- We held p fixed at q to study the map Exp_q from $T_q(M)$ to M.

A Strengthening (Cont'd)

- Now, however, we consider the mapping F of the open set

$$\widetilde{\varphi}(\{X_p : p \in V, \|X_p\| < \varepsilon\}) \subseteq \mathbb{R}^n \times \mathbb{R}^n$$

to

$$\varphi(U) \times \varphi(U) \subseteq \mathbb{R}^n \times \mathbb{R}^n$$

which is defined by

$$F : (x^1, \ldots, x^n; b^1, \ldots, b^n) \to (x^1, \ldots, x^n; f^1(1, x, b), \ldots, f^n(1, x, b)).$$

- This map corresponds to the map

$$X_p = \sum b^i E_{ip} \to (p, \mathrm{Exp} X_p),$$

with domain in $T(M)$.

A Strengthening (Cont'd)

- We have already seen that, when $b^1 = \cdots = b^n = 0$,
$$\frac{\partial f^i}{\partial b^j} = \delta^i_j.$$
- So the Jacobian matrix of F is nonsingular at any point $(x^1, \ldots, x^n; 0, \ldots, 0)$ of $\mathbb{R}^n \times \{0\}$ for which $(x^1, \ldots, x^n) = \varphi(p)$, with $p \in V$.
- Therefore, by the Inverse Function Theorem, for each pair $(p, 0_p)$, 0_p the zero vector at $p \in V$, there is a neighborhood which is mapped diffeomorphically onto an open subset of $U \times U \subseteq M \times M$ by this mapping.
- The mapping takes the pair "p and vector X_p at p" to a pair of points of U,
$$(p, X_p) \to (p, \mathrm{Exp} X_p).$$

A Strengthening (Cont'd)

- Now V was originally chosen as a relatively compact neighborhood of q lying in a coordinate neighborhood U, φ.
- It was used to obtain an $\varepsilon > 0$ for which the open set $\{X_p : p \in V \text{ and } \|X_p\| < \varepsilon\}$ of $T(M)$ was in the domain $\mathscr{D}$ of Exp.
- This is also a set on which the mapping $(p, X_p) \to (p, \text{Exp} X_p)$ is given in local coordinates by F.
- From what we have just said we may restrict V and ε further (without changing notation) so that the resulting neighborhood $N(V, \varepsilon) = \{(p, X_p) : p \in V \text{ and } \|X_p\| < \varepsilon\}$ of q, 0_q is mapped diffeomorphically onto an open set $W \subseteq U \times U$.
- Although W is not of the form $B \times B$, it does contain the diagonal set $\{(p, p) : p \in V\}$.

A Strengthening (Cont'd)

- We now let $B \subseteq V$ be a neighborhood of q such that $B \times B \subseteq W$.
- Then $B \times B$ is the diffeomorphic image of some open subset of $N(V, \varepsilon)$ which can be described by
$$N_B = \{(p, X_p) : p \in B, \text{Exp} X_p \in B\}.$$
- Putting these facts together gives the following result.

Theorem

Let U, φ be a coordinate neighborhood of M and $q \in U$. Then there exists a neighborhood $B \subseteq U$ of q and an $\varepsilon > 0$, such that any two points p, p' of B can be joined by a unique geodesic of length less than ε. This geodesic is of the form $\text{Exp} tX_p$, $0 \leq t \leq 1$, and lies entirely in U. It follows that for each $p \in B$, Exp_p maps $\{X_p : \|X_p\| < \varepsilon\}$ diffeomorphically into an open set N_p, such that $B \subseteq N_p \subseteq U$.

Remarks

- Our choice of the neighborhood N_B does not allow us to conclude that whenever $(p, X_p) \in N_B$, then $(p, tX_p) \in N_B$, for all $0 < t < 1$.
- Thus, in general, B does not necessarily have the property that $p, p' \in B$ are joined by a geodesic lying entirely in B.
- We have made our choices so that for each $p \in V$, Exp_p maps the ε ball $\{X_p : \|X_p\| < \varepsilon\}$ into U diffeomorphically and clearly has B in its image.
- Thus, each $p \in B$ has a normal neighborhood N_p, with $B \subseteq N_p \subseteq U$.
- With somewhat more effort one can show that it is, in fact, possible to select a neighborhood B of each point q on a Riemannian manifold with the property that each pair of points $p, p' \in B$ may be joined by a unique (minimizing) geodesic segment lying entirely in B.
- Such neighborhoods are called **geodesically convex**.

Subsection 7

Some Further Properties of Geodesics

Properties of $\mathscr{D}$ and Exp

Theorem

$\mathscr{D}$ is an open subset of $T(M)$ and $\text{Exp} : \mathscr{D} \to M$ is a C^∞ mapping.

- We keep the same notation.

  Recall that to each coordinate neighborhood U, φ of M corresponds a coordinate neighborhood $\widetilde{U}, \widetilde{\varphi}$ of $T(M)$.

  We have

  $$\widetilde{U} = \pi^{-1}(U) \quad \text{and} \quad \widetilde{\varphi}(\widetilde{U}) = \varphi(U) \times \mathbb{R}^n \subseteq \mathbb{R}^n \times \mathbb{R}^n.$$

  In fact, let:
  - $\varphi(p) = (x^1, \ldots, x^n)$;
  - $E_1, \ldots, E_n$ be the coordinate frames.

  Then

  $$\widetilde{\varphi}(X_p) = \widetilde{\varphi}\left(\sum y^i E_i\right) = (x^1, \ldots, x^n; y^1, \ldots, y^n).$$

Properties of $\mathscr{D}$ and ExP (Cont'd)

- The natural mapping $\pi : T(M) \to M$ is given in local coordinates by
$$\varphi(\pi(X_p)) = (x^1, \ldots, x^n).$$
It is an open C^∞ mapping and has rank n at every point.
Suppose that $p(t)$ is a geodesic on M.
Then its velocity vector $X_{p(t)} = \frac{dp}{dt}$ defines a curve $t \to X_{p(t)}$ on $T(M)$ with
$$\pi(X_{p(t)}) = p(t).$$
An examination of the method by which we passed from the equations of geodesics to first-order equations reveals that on $\widetilde{\varphi}(\widetilde{U})$ we considered the first-order system corresponding to the vector field
$$Z' = \sum_i y^i \frac{\partial}{\partial x^i} + \sum_k \left(\sum_{i,j} \Gamma^k_{ij}(x) y^i y^j \right) \frac{\partial}{\partial y^k}.$$

Properties of $\mathscr{D}$ and ExP (Cont'd)

- Now we define a vector field Z on $\widetilde{U} \subseteq T(M)$ so that
$$\widetilde{\varphi}_*(Z) = Z'.$$
Suppose, as in a previous lemma, the solutions of the first-order equations are given by
$$x^i(t) = f^i(t, a, b) \quad \text{and} \quad y^i(t) = \frac{dx^i}{dt}, \quad i = 1, \ldots, n.$$
Then on $\widetilde{U}$ the integral curves (solutions) of the system of equations defined by Z are of the form
$$\widetilde{\varphi}^{-1}\left(x^1(t), \ldots, x^n(t); \frac{dx^1}{dt}, \ldots, \frac{dx^n}{dt}\right),$$
where $\varphi^{-1}(x^1(t), \ldots, x^n(t)) = p(t)$ is a geodesic in $U = \pi(\widetilde{U})$.
In brief, $X_{p(t)} = \frac{dp}{dt}$ is a solution curve of Z on $\pi^{-1}(U) \subseteq T(M)$ if and only if $p(t)$ is a geodesic on U.

Properties of $\mathscr{D}$ and ExP (Cont'd)

- From its geometric meaning, or by a tedious computation for change of coordinates, we see that Z is a vector field defined intrinsically on all of $T(M)$, independent of the particular expression in a coordinate system.

  That is, the components

$$\left(x^1, \ldots, x^n, \sum_{i,j} \Gamma_{ij}^1 y^i y^j, \ldots, \sum_{i,j} \Gamma_{ij}^n y^i y^j\right)$$

  transform as they should for a vector field when we pass to other coordinates.

  So Z is globally defined and depends only on the Riemannian connection and metric.

  The geodesics on M are therefore exactly the projections by $\pi : T(M) \to M$ of the integral curves of Z.

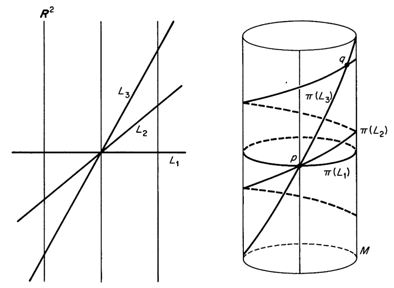
  Thus, the conclusion follows from a previous theorem.

Comparison With $\mathbb{R}^n$

- We have seen that geodesics on Riemannian manifolds generalize straight lines in $\mathbb{R}^n$ in the following sense.
- Their unit tangent vector as we move along the curve is constant.
- But another basic property which characterizes straight lines in $\mathbb{R}^n$ is the famous minimizing property of being the shortest curve joining any two of its points.
- We now examine in some detail the extent to which this property generalizes.

Example

- Consider the right circular cylinder M with the Riemannian metric obtained by considering the plane $\mathbb{R}^2$, with its usual metric, as universal covering.
- Then the geodesics on the cylinder are exactly those curves which go into straight lines if we roll the cylinder along the plane.
- That is, vertical generators and helices.



- Thus, two points not on a circle whose plane is orthogonal to the axis will be joined by an infinite number of distinct geodesics of different lengths.

Examples

- Consider the sphere S^2.

  On S^2 consider the larger of the two arcs of a great circle which join two points p and q (which are not at opposite ends of a diameter).

  Such a path is not of minimal length, even among nearby circular arcs.

- Finally, consider the plane with the origin removed.

  The points $(-1, 0)$ and $(+1, 0)$ cannot be joined by a shortest curve at all.

Conclusions

- In view of the preceding examples, it is remarkable that we are able to salvage something, in fact almost everything, if we limit ourselves to points close together and short geodesics.
- Let us recall that we have defined the length of a piecewise differentiable curve $p(t)$ (of class D^1), over $a \leq t \leq b$, by

$$L = \int_a^b \left\| \frac{dp}{dt} \right\| dt.$$

- This is the Riemann integral of a piecewise continuous function.
- It is, by definition, equal to the sum of the integrals over the intervals of continuity [on each of which $p(t)$ is of class C^1].

Geodesic Spheres

- According to a previous theorem, given $q \in M$, there exist B and $\varepsilon > 0$, such that each pair of points p, p' of B can be joined by a unique geodesic of length $L < \varepsilon$.
- In fact, the equation $p(t)$ of the geodesic is given by

$$p(t) = \mathrm{Exp}\, tX_p, \quad 0 \leq t \leq 1,$$

  and $\|X_p\| = L$.
- The open set B lies in a coordinate neighborhood U, φ which contains this geodesic.
- Exp_p is a diffeomorphism of the open ball of vectors X_p of $T_p(M)$ of length $\|X_p\| < \varepsilon$ onto an open set N_p of U containing B.
- This means that any sphere

$$\{X_p : \|X_p\| = r < \varepsilon\}$$

  maps diffeomorphically to a submanifold of U, denoted by S_r (and called a **geodesic sphere**).

Uniqueness of Geodesics: Lemma 1

Lemma

Let $p \in B$ and suppose Exp_p maps the open ε-ball of $T_p(M)$ diffeomorphically onto $N_p \supseteq B$. Then the geodesics through p are orthogonal to the geodesic spheres S_r, determined by

$$\mathrm{Exp}_p X_p \quad \text{and} \quad \|X_p\| = r, \quad r < \varepsilon.$$

- Let $X(t)$ be a curve in $T_p(M)$ with $\|X(t)\| \equiv 1$, $a \leq t \leq b$.
  Any geodesic from the point p may be written

  $$r \to \mathrm{Exp}_p rX, \quad 0 \leq r \leq \varepsilon,$$

  with $\|X\| = 1$.
  Any curve on S_r may be written in the form

  $$t \to \mathrm{Exp}_p rX(t).$$

Uniqueness of Geodesics: Lemma 1 (Cont'd)

- The mapping
$$(r,t) \to p(r,t) = \mathrm{Exp}_p rX(t)$$
maps the rectangle $[0,\varepsilon] \times [a,b]$ differentiably into M.
We will show that the inner product
$$\left(\frac{\partial p}{\partial r}, \frac{\partial p}{\partial t}\right) = 0, \quad \text{for each } r_0, t_0.$$
$\frac{\partial p}{\partial r}$ is the tangent vector to $p(r,t_0)$, the geodesic curve.
$\frac{\partial p}{\partial t}$ is the tangent vector to $p(r_0,t)$ a curve on the geodesic sphere S_r.
They intersect at $p(r_0,t_0)$.
It suffices to show this inner product vanishes for every (r_0,t_0).

Uniqueness of Geodesics: Lemma 1 (Cont'd)

- We first show that $(\frac{\partial p}{\partial r}, \frac{\partial p}{\partial t})$ is independent of r.

  By a basic property of differentiation,

$$\frac{D}{\partial r}\left(\frac{\partial p}{\partial r}, \frac{\partial p}{\partial t}\right) = \left(\frac{D}{\partial r}\frac{\partial p}{\partial r}, \frac{\partial p}{\partial t}\right) + \left(\frac{\partial p}{\partial r}, \frac{D}{\partial r}\frac{\partial p}{\partial t}\right).$$

  Of these, $\frac{D}{\partial r}\frac{\partial p}{\partial r} = 0$, since holding t fixed and allowing r to vary gives a geodesic through q with $\frac{\partial p}{\partial r}$ as its unit tangent vector.

  In the second term, if we interchange the order of differentiation, we obtain

$$\left(\frac{\partial p}{\partial r}, \frac{D}{\partial t}\frac{\partial p}{\partial r}\right) = \frac{1}{2}\frac{\partial D}{\partial t}\left(\frac{\partial p}{\partial r}, \frac{\partial p}{\partial r}\right).$$

Uniqueness of Geodesics: Lemma 1 (Cont'd)

- Now, we have
$$\left\|\frac{\partial p}{\partial r}\right\| = \|X(t)\| \equiv 1.$$
So we get
$$\frac{\partial D}{\partial t}\left(\frac{\partial p}{\partial r}, \frac{\partial p}{\partial r}\right) = 0.$$
Therefore, $(\frac{\partial p}{\partial r}, \frac{\partial p}{\partial t})$ is independent of r.
But $p(0,t) \equiv q$. So $\frac{\partial p}{\partial t} = 0$ at $r = 0$.
Thus,
$$\left(\frac{\partial p}{\partial r}, \frac{\partial p}{\partial t}\right) = 0, \quad \text{for all } r.$$
Hence, for each (r_0, t_0), the inner product
$$\left(\frac{\partial p}{\partial r}, \frac{\partial p}{\partial t}\right) = 0.$$

Uniqueness of Geodesics: Lemma 2

- We consider a (piecewise) differentiable curve in $N_p - \{p\}$,

$$\widetilde{p}(t), \quad a \leq t \leq b.$$

- It has a unique expression of the form

$$\widetilde{p}(t) = \mathrm{Exp}_p r(t) X(t), \quad \|X(t)\| \equiv 1. vspace{-0.15in}$$

Lemma

We have

$$\int_a^b \left\| \frac{d\widetilde{p}}{dt} \right\| dt \geq |r(b) - r(a)|.$$

Equality holds if and only if $r(t)$ is monotone and $X(t)$ is constant.

Uniqueness of Geodesics: Lemma 2 (Cont'd)

- Again we consider the map

$$\begin{array}{rcl}[0,\varepsilon]\times[a,b] & \to & U; \\ (r,t) & \to & p(r,t)=\mathrm{Exp}_p rX(t).\end{array}$$

The curve $\widetilde{p}(t)$ connects the spherical shells S_r in U_q of radius

$$r=r(a) \quad \text{and} \quad r=r(b).$$

We have

$$\widetilde{p}(t)=p(r(t),t).$$

Moreover,

$$\frac{d\widetilde{p}}{dt}=\frac{\partial p}{\partial r}r'(t)+\frac{\partial p}{\partial t}.$$

Uniqueness of Geodesics: Lemma 2 (Cont'd)

- Now $\|\frac{\partial p}{\partial r}\| = \|X(t)\| = 1$ and $(\frac{\partial p}{\partial r}, \frac{\partial p}{\partial t}) = 0$.
  So we have
  $$\left\|\frac{\partial \widetilde{p}}{dt}\right\|^2 = |r'(t)|^2 + \left\|\frac{\partial p}{\partial t}\right\|^2 \geq |r'(t)|^2.$$
  Equality holds if and only if $\frac{\partial p}{\partial t} \equiv 0$, that is, $X(t) = \text{constant}$.
  Consequently,
  $$\int_a^b \left\|\frac{d\widetilde{p}}{dt}\right\| dt \geq \int_a^b |r'(t)| dt \geq \left|\int_a^b r'(t) dt\right| = |r(b) - r(a)|.$$
  In the last inequality, we have equality only if $r(t)$ is monotone.
  Thus, $\int_a^b \|\frac{d\widetilde{p}}{dt}\| dt = |r(b) - r(a)|$ if and only if $r(t)$ is monotone and $X(t) = \text{constant}$.

Proof of Uniqueness of Geodesics

- We continue the notation of the lemmas.

  Suppose
  $$\widetilde{p}(t), \quad 0 \leq t \leq 1,$$
  is a piecewise smooth curve joining:
  - $p = \widetilde{p}(0)$.
  - $p' = \widetilde{p}(1) = \mathrm{Exp}_p rX_p \in N_p$, $0 < r < \varepsilon$ and $\|X_p\| = 1$.

  Let δ satisfy $0 < \delta < \varepsilon$, and consider the segment of the curve joining the shell of radius δ around p to that of radius r.

  According to the preceding lemma:
  - The length of this segment is $\geq r - \delta$;
  - Equality holds only if the curve coincides as a point set with segment of the radial geodesic from p cut off by these shells, its length being $r - \delta$.

  Thus, the portion of the curve between these shells has length $> r - \delta$, unless it coincides as a point set with a radial geodesic.

  Letting δ approach zero gives the result of the theorem.

Comments

- The statement of the theorem is bound up with the notion of distance on M, that is, the metric $d(p,p')$ which we considered previously.
- Recall that $d(p,p')$ is the infimum of the lengths of all piecewise differentiable curves from p to p'.
- Moreover, we showed that the metric topology and the usual topology coincided.
- The theorem just proved guarantees that, for each point $q \in M$, there is an $\varepsilon > 0$ and a neighborhood B of diameter less than ε (in terms of d), such that, for every pair of points $p, p' \in B$, there is a unique geodesic segment from p to p' whose length is the distance $d(p,p')$.

Minimal Geodesics

Corollary

If a piecewise differentiable path (of class D^1) from p to q on M has length equal to $d(p,q)$, then it is a geodesic when parametrized by arclength.

- Note that it follows that the path is C^∞!
- Of course the hypothesis and the definition of $d(p,q)$ imply that the path has minimum length among all such curves.
- For the proof, note that any segment of the path lying in a sufficiently small neighborhood (as above) must also have as length the distance between its endpoints (or it could be replaced by a shorter path).

  So the path must be a geodesic.

  It follows that the curve is a geodesic locally.

  So it is a geodesic.

Minimal Geodesics and Indefinite Extendibility

Definition

A geodesic segment whose length is the distance between its endpoints is called a **minimal geodesic**.

- We recall that each geodesic and geodesic segment is contained in a maximal geodesic, that is, a geodesic $p(t)$ such that $p(t)$ is defined for $a < t < b$ and not for any larger interval of values.
- If $a = -\infty$ and $b = +\infty$, we say that the geodesic can be **extended indefinitely**.
- This is always true of a closed geodesic (a geodesic which is the image of a circle, for example, a great circle on S^2).
- If every geodesic from $p \in M$ can be extended indefinitely, then the domain $\mathscr{D}$ of Exp contains all of $T_p(M)$ and conversely.

The Hopf and Rinow Theorem

Theorem (Hopf and Rinow)

Let M be a connected Riemannian manifold. Then the following two properties are equivalent:

(i) Any geodesic segment can be extended indefinitely.

(ii) With the metric $d(p,q)$, M is a complete metric space.

- The proof will be based on a lemma.

  Assume any geodesic segment $t \to p(t)$, $a \leq t \leq b$, can be extended to a maximal geodesic curve $t \to p(t)$, defined for $-\infty < t < +\infty$.

  To see that M is complete (every Cauchy sequence converges), it is sufficient to show that every closed and bounded set is compact.

  To prove this we need the following lemma, of interest in itself.

The Hopf and Rinow Theorem (Lemma)

Lemma

Suppose M has the property that every geodesic from some point $p \in M$ can be extended indefinitely. Then any point q of M can be joined to p by a minimal geodesic [whose length is necessarily $d(p, q)$].

- Let q be an arbitrary point of M and let $a = d(p, q)$.
  Any geodesic from p may be written $p(s) = \text{Exp} s X_p$ with:
  - X_p a unit tangent vector at p;
  - s arclength measured from $p = p(0)$.

  We must show that, for some X_p, with $\|X_p\| = 1$,

  $$p(a) = \text{Exp} a X_p = q.$$

  Then $s \mapsto \text{Exp} s X$, $0 \leq s \leq a$, would be the minimal geodesic segment.

  We will use the following fact, which is also of some interest.

The Hopf and Rinow Theorem (Fact)

Fact

Suppose that $p_0, p_1, \ldots, p_n$ are points of M and that

$$d(p_0, p_1) + d(p_1, p_2) + \cdots + d(p_{n-1}, p_n) = d(p_0, p_n)$$

If a piecewise differentiable curve contains $p_i, p_{i+1}, \ldots, p_{i+r}$ and has length equal to $d(p_i, p_{i+1}) + \cdots + d(p_{i+r-1}, p_{i+r})$, then it is a geodesic segment from p_i to p_{i+r}. Conversely, if $p_0, \ldots, p_n$ lie on a minimal geodesic segment, in that order, then the equation holds for them.

- It is easily seen that it is enough to verify this for $r = 2$.
  The curve C from p_i to p_{i+1} to p_{i+2} has length

$$L = d(p_i, p_{i+1}) + d(p_{i+1}, p_{i+2}).$$

  By the triangle inequality $L \geq d(p_i, p_{i+2})$.
  If equality holds, C is a geodesic segment from p_i to p_{i+2}.

The Hopf and Rinow Theorem (Fact Cont'd)

- We show that $L \geq d(p_i, p_{i+2})$.

  Otherwise, by the triangle inequality, we have

  $$d(p_i, p_{i+1}) + d(p_{i+1}, p_{i+2}) > d(p_i, p_{i+2}).$$

  Then, substituting in the statement, we get

  $$d(p_0, p_1) + \cdots + d(p_i, p_{i+2}) + \cdots + d(p_{n-1}, p_n) < d(p_0, p_n).$$

  This contradicts the triangle inequality.

  Finally, the last statement follows immediately from the fact that any subsegment of a minimal geodesic segment is also minimal.

The Hopf and Rinow Theorem (Lemma Cont'd)

- Using a previous theorem, let $\delta > 0$ be chosen so that

$$S_\delta = \{p' : d(p, p') = \delta\}$$

is a geodesic sphere in some normal neighborhood of p, sufficiently small to ensure that each radial geodesic from p to S_δ is minimal.

Now S_δ is compact.

So, there exists a $p_0 \in S_\delta$, satisfying

$$d(p_0, q) = \inf_{p' \in S_\delta} d(p', q).$$

Let X_p be the unit vector at p, such that

$$p_0 = \text{Exp}\delta X_p.$$

The Hopf and Rinow Theorem (Lemma Cont'd)

- We must have
$$d(p, p_0) + d(p_0, q) = d(p, q).$$
Otherwise, there is a piecewise differentiable curve joining p to q whose length is less than
$$d(p, p_0) + d(p_0, q) = \delta + d(p_0, q).$$
It must intersect S_δ at some point p' and its length from p to p' can be no less than δ.
So we have $d(p', q) < d(p_0, q)$.
This contradicts our choice of p_0.

The Hopf and Rinow Theorem (Lemma Cont'd)

- We now consider all s', $0 \leq s' \leq a$, such that:
  - The geodesic segment $s \mapsto \text{Exp} s X_p$, $0 \leq s \leq s'$, is minimizing;
  - $d(p, \text{Exp} s' X_p) + d(\text{Exp} s' X_p, q) = d(p, q)$.

  Both conditions are continuous.

  So the collection of all such s' forms a closed interval $0 \leq s' \leq b$.

  If $b = a$, then $\text{Exp} a X_p = q$, which proves the lemma..

  Suppose, on the other hand, that $b < a$.

  Let $p_1 = \text{Exp} b X_p$.

  Then
  $$d(p, p_1) + d(p_1, q) = d(p, q).$$

  We may obtain a contradiction by repeating the arguments above as follows.

The Hopf and Rinow Theorem (Lemma Cont'd)

- Let S_η, $\eta > 0$, be a small geodesic sphere (with radial geodesics minimizing) in a normal neighborhood of $p_1 = \text{Exp} bX_p$.
  Choose a point p_2 on S_η, such that
  $$d(p_2, q) = \inf_{p'' \in S_\eta} d(p'', q).$$
  Then, as before
  $$d(p_1, p_2) + d(p_2, q) = d(p_1, q).$$
  Therefore,
  $$d(p, p_1) + d(p_1, p_2) + d(p_2, q) = d(p, p_1) + d(p_1, q) = d(p, q).$$
  By the Fact, the geodesic $p(s) = \text{Exp} sX_p$ from p to p_1 together with the (radial) geodesic in S_η from p_1 to p_2 is a single (minimizing) geodesic segment from p to p_2 of length $d(p, p_2) > b$.
  This contradicts the definition of b.
  Therefore, $b = a$ and the lemma follows.

The Hopf and Rinow Theorem ((i)⇒(ii))

- We show first that (i) implies (ii).

  Let K be a closed and bounded subset of M.

  We show that K is compact.

  Suppose $p \in K$ and let

$$a = \sup_{q \in K} d(p, q).$$

  a is finite, since K is bounded.

  By the lemma, for any $q \in K$, there is a minimizing geodesic from p to q.

  Its length is $d(p, q)$, which must be no greater than a.

The Hopf and Rinow Theorem ((i)⇒(ii) Cont'd)

- It follows that $K \subseteq \mathrm{Exp}_p \overline{B}_a$, where $\overline{B}_a$ is the closed ball of radius a in T_p,
$$\overline{B}_a = \{Y_p : \|Y_p\| \leq a\}.$$
Now $\overline{B}_a$ is compact and Exp is continuous.
So $\mathrm{Exp}_p \overline{B}_a$ is compact.
K is a closed subset of $\mathrm{Exp}_p \overline{B}_a$.
So it must be compact.
However, a metric space in which every bounded set is relatively compact (has compact closure) is complete.
So M is a complete metric space.

Remark

- Any manifold M having Property (i) has the property that the domain $\mathscr{D}$ of the exponential function is all of $T(M)$.
- That is, that the vector field Z of a previous theorem is complete.
- Actually, in proving that (i) implies (ii), we used only the weaker hypothesis of the lemma.
- I.e., that every geodesic from some point $p \in M$ can be extended indefinitely, that is, $\mathscr{D} \supseteq T_p(M)$ for some $p \in M$.
- It was not necessary to assume $p \in K$, for if K is bounded, then for any $p \in M$ the distances $d(p, q)$ are bounded for all $q \in K$.

The Hopf and Rinow Theorem ((ii)⇒(i))

- For (ii) implies (i), we suppose that every Cauchy sequence on M converges and show that this implies the extendability of geodesics.
  Suppose to the contrary that there is a geodesic ray,
  $$p(t), \quad 0 \leq t < t_0,$$
  which cannot be extended to $t = t_0$.
  We may assume, changing parameter if necessary, that t is arclength.
  Let $\{t_n\}$ be an increasing sequence of values with $\lim_{n\to\infty} t_n = t_0$.
  Denote by p_n the points $p(t_n)$.
  The expression $|t_n - t_m|$ is the length of a curve from p_n to p_m.
  So we have
  $$d(p_n, p_m) \leq |t_n - t_m|.$$
  Thus, $\{p_n\}$ is a Cauchy sequence.
  We let
  $$q = \lim_{n\to\infty} p_n = \lim_{t_n\to t_0} p(t_n).$$

The Hopf and Rinow Theorem ((ii)⇒(i) Cont'd)

- Let B be a neighborhood of q.

  Let $\varepsilon > 0$ be so chosen that each pair of points p, p' of B are joined by a unique geodesic of length less than ε.

  This geodesic is minimizing, or equivalently its length is $d(p, p')$.

  Let N be an integer large enough so that, for $n, m \geq N$,

$$d(p_n, p_m) < \varepsilon, \quad d(p_n, q) < \varepsilon \quad \text{and} \quad p_n, p_m \in B.$$

  Consider $n \geq N$ fixed and suppose $m > n$.

  Then we have

$$d(p_n, p_m) + d(p_m, q) = (t_m - t_n) + d(p_m, q).$$

  $t_m - t_n$ is the length of our geodesic from p_n to p_m.

  Moreover, it is less than ε.

  So this segment of the geodesic is minimal.

The Hopf and Rinow Theorem ((ii)⇒(i) Cont'd)

- Now let $m \to \infty$.

  By continuity, we have

  $$d(p_n, q) = t_0 - t_n, \quad \text{for } n > N.$$

  Applying this to $m > n$, we have, for all $m > n > N$,

  $$d(p_n, p_m) + d(p_m, q) = t_m - t_n + t_0 - t_m = t_0 - t_n = d(p_n, q).$$

  Choose a fixed $m > n$.

  We see that the unique geodesic segment from p_n to p_m of length $d(p_n, p_m)$ together with the unique geodesic segment from p_n to q of length $d(p_n, q)$ has length equal to the distance $d(p_n, q)$.

  Therefore, it is a single (unbroken) geodesic from p_n to q.

The Hopf and Rinow Theorem ((ii)⇒(i) Cont'd)

- However, it coincides with the given geodesic

$$p(t), \quad t_n \le t \le t_m.$$

That is, from p_n to p_m.

Thus, it is an extension of this to a geodesic segment from p to q.

This shows that $p(t)$ can be extended to $t = t_0$.

We note that it is immediate that a geodesic segment $p(t)$, $0 \le t \le t_0$, can be extended beyond its endpoints.

This follows at once from the fundamental existence theorems.

So any geodesic on a complete manifold can be extended indefinitely.

This implies that Exp_p is defined on all of $T_p(M)$, for every p.

Hence, Exp has the entire tangent bundle $T(M)$ as its domain, that is, $\mathscr{D} = T(M)$.

Consequences

- The following corollary depends on the fact that a compact metric space is complete.

Corollary

Let M be a compact connected Riemannian manifold.
Then any pair of points $p, q \in M$ may be joined by a geodesic whose length is $d(p, q)$.

Consequences (Cont'd)

Corollary

Let $F_1, F_2 : M \to M$ be isometries of a complete, connected Riemannian manifold. Suppose that, for some $p \in M$:

- $F_1(p) = F_2(p)$;
- $F_{1*} = F_{2*}$ on $T_p(M)$.

Then $F_1 = F_2$.

- Let $q \in M$ and let
$$p(s), \quad 0 \le s \le \ell,$$
be a geodesic from $p = p(0)$ to $q = p(\ell)$.
Then, for $i = 1, 2$,
$$F_i(p(s))$$
is a geodesic from $F_i(p)$ to $F_i(q)$.

Consequences (Cont'd)

- But, by hypothesis,

$$F_1(p) = F_2(p) \quad \text{and} \quad F_{1*}(\dot{p}(0)) = F_{2*}(\dot{p}(0)).$$

So these geodesics coincide.

So we have

$$F_1(q) = F_1(p(\ell)) = F_2(p(\ell)) = F_2(q).$$

Subsection 8

Symmetric Riemannian Manifolds

Symmetric Connected Riemannian Manifolds

Definition

A connected Riemannian manifold M is said to be **symmetric** if, to each $p \in M$, there is associated an isometry

$$\sigma_p : M \to M,$$

which:

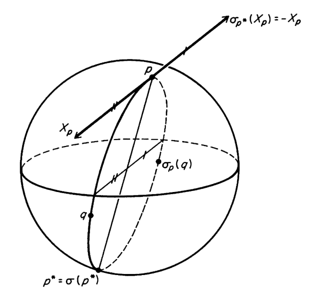
(i) Is involutive (σ_p^2 is the identity);

(ii) Has p as an isolated fixed point, that is, there is a neighborhood U of p in which p is the only fixed point of σ_p.

- An example is Euclidean n-space.
- In that case σ_p is reflection in p.

Example of a Symmetric Manifold

- Consider S^n, the unit sphere in $\mathbb{R}^{n+1}$, with the metric induced by $\mathbb{R}^{n+1}$.
- In the case of the sphere, σ_p is again reflection in p.
- For each q, $\sigma_p(q) = q'$, where q and q' are equidistant from p on a geodesic (great circle) through p.



- In the case of S^n we note that $\sigma_p(p) = p$ and $\sigma_p(p^*) = p^*$, where p^* denotes the point antipodal to p.
- Thus, in general, σ_p may have other fixed points than p.
- Note also that the previous example is a noncompact manifold whereas this is compact.
- A symmetric space, as we will see, is always complete.

Properties of the Isometry

Lemma

Let M be a Riemannian manifold. Let $p \in M$ and σ_p an involutive isometry, with p as isolated fixed point. Then, for all $X_p \in T_p(M)$,

$$\sigma_{p*}(X_p) = -X_p \quad \text{and} \quad \sigma_p(\text{Exp}X_p) = \text{Exp}(-X_p).$$

- We know that σ_p^2 is the identity.

  So the same holds for $(\sigma_{p*})^2$ on $T_p(M)$.

  This means that the eigenvalues of σ_{p*} on $T_p(M)$ are ± 1.

  Suppose $+1$ is an eigenvalue.

  Then, there exists a vector $X_p \neq 0$ such that $\sigma_{p*}(X_p) = X_p$.

  Isometries preserve geodesics.

  So, for any isometry $F : M \to M$, $F \circ \text{Exp} = \text{Exp} \circ F_*$.

Properties of the Isometry (Cont'd)

- For any isometry $F : M \to M$, $F \circ \text{Exp} = \text{Exp} \circ F_*$.

  This means that $\sigma_p(\text{Exp}tX) = \text{Exp}tX$.

  So the geodesic through p with initial direction X_p is pointwise fixed.

  This means that p is not an isolated fixed point of a p.

  Thus $+1$ is not an eigenvalue and $\sigma_{p*} = -I$, I being the identity.

  Now σ is an isometry.

  So

  $$\sigma_p(\text{Exp}X_p) = \text{Exp}\sigma_{p*}(X_p) = \text{Exp}(-X_p).$$

  This means that σ_p takes each geodesic through p onto itself with direction reversed, exactly as in the two examples.

Uniqueness of the Isometry for Complete Manifolds

- By the preceding lemma and a previous corollary, we get

Corollary

Given any complete Riemannian manifold M and point $p \in M$, there can be at most one involutive isometry σ_p with p as isolated fixed point.

Completeness of Symmetric Manifolds

Theorem

A symmetric Riemannian manifold M is necessarily complete. Moreover, if $p, q \in M$, then there is an isometry σ_r, corresponding to some $r \in M$, such that

$$\sigma_r(p) = q.$$

- First we show that M is complete.

  We prove that every geodesic can be extended to infinite length.

  Suppose $p(s)$, $0 \leq s < b$, is a geodesic ray, with s as arclength.

  We will show that it can be extended to a length $\ell > b$.

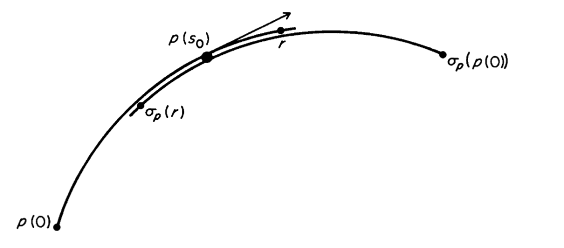
  Let $s_0 = \frac{3}{4}b$, and let $\sigma_{p(s_0)}$ be the symmetry in $p(s_0)$.

  It takes the geodesic $p(s)$ to another geodesic through $p(s_0)$ whose:
  - Tangent vector at $p(s_0)$ is $-\frac{dp}{ds}|_{s_0}$;
  - Length is the same as that of $p(s)$.

Completeness of Symmetric Manifolds (Cont'd)

- Since it has a common tangent with $p(s)$ at $p(s_0)$, it coincides with $p(s)$ on the interval $\frac{1}{2} < s < b$.

  Thus it extends it to a length $> \frac{3}{2}b$, which proves the statement.



  Using this it follows easily that given any $p, q \in M$ there is an isometry of M taking p to q.

  In fact, let r be the midpoint of a geodesic from p to q.

  Then the isometry σ_r takes this geodesic onto itself and carries p to q.

Group of Isometries

- It is easy to verify that the isometries of a Riemannian manifold M form a group $I(M)$.
- It is a subgroup of the group of all diffeomorphisms of M.
- A classical theorem due to Myers and Steenrod asserts that:
  - It is a Lie group
  - Acts differentiably on M.
- By the theorem, it is also transitive when M is a symmetric space.

Symmetry of Compact Connected Lie Groups

Theorem

Every compact connected Lie group G is a symmetric space with respect to the bi-invariant metric.

- Let $\psi : G \to G$ denote the diffeomorphism which takes each element to its inverse, $\psi(x) = x^{-1}$.

  This map is clearly involutive.

  It is an isometry of G with e, the identity, as isolated fixed point.

  To see this, recall that, to each $X_e \in T_e(G)$, corresponds a uniquely determined one-parameter subgroup $t \mapsto g(t)$, with $\dot{g}(0) = X_e$.

  Since $\psi(g(t)) = g(-t)$, by the chain rule, we obtain

$$\psi_*(X_e) = \psi_*(\dot{g}(0)) = \frac{d}{dt}\psi(g(t))|_{t=0} = -\dot{g}(0) = -X_e.$$

  This means that $\psi_{*e} = -I$, which is an orthogonal linear transformation (or isometry) of any inner product on $T_e(G)$.

Symmetry of Compact Connected Lie Groups (Cont'd)

- Let $a \in G$ be arbitrary.
  Given any $g \in G$, denote by:
  - L_g left translation by g;
  - R_g right translation by g.

  We may write

  $$\psi(x) = x^{-1} = (a^{-1}x)^{-1}a^{-1} = R_{a^{-1}} \circ \psi \circ L_{a^{-1}}(x).$$

  Hence $\psi_{*a} : T_a(G) \to T_{a^{-1}}(G)$ may be written

  $$\psi_{*a} = (R_{a^{-1}*})_\circ \psi_{*e} \circ (L_{a^{-1}*})_a.$$

  This is a composition of three linear mappings each of which is an isometry of the inner product determined by the bi-invariant metric ($R_{a^{-1}}$ and $L_{a^{-1}}$ induce isometries on every tangent space and ψ_{*e} is an isometry as shown above).
  It follows that $\psi : G \to G$ is an isometry.

Symmetry of Compact Connected Lie Groups (Cont'd)

- Now consider a normal neighborhood of e.

  Then, by a previous lemma, ψ is given in local coordinates by reflection in the origin.

  Hence, e is an isolated fixed point.

  Now let $g \in G$.

  We define the isometry $\sigma_g : G \to G$ which has g as an isolated fixed point by

$$\sigma_g = L_g \circ R_g \circ \psi.$$

  That is,

$$\sigma_g(x) = gx^{-1}g.$$

  It is an isometry since R_g, L_g and ψ are isometries.

  We can check that it is involutive and has g as isolated fixed point.

Example

- Let $G = SO(n)$ be the group of $n \times n$ orthogonal matrices of determinant $+1$.
- According to a previous example, the tangent space $T_e(G)$, $e = I$, the $n \times n$ identity matrix, may be identified with the skew symmetric matrices $A = (a_{ij}) = -A'$.
- The identification means that

$$X_e = \sum_{i,j} a_{ij} \frac{\partial}{\partial x_{ij}}$$

  is tangent at I to $SO(n)$ considered as a submanifold of $Gl(n, R) \subseteq \mathbb{R}^{n^2}$.
- The one-parameter subgroups are of the form

$$Z(t) = e^{tA}.$$

Example (Cont'd)

- In this case we may compute

$$\mathrm{Ad}B : T_e(G) \to T_e(G)$$

as follows.

- First one verifies, from the definition of e^{tA}, that

$$Be^{tA}B^{-1} = e^{tBAB^{-1}}.$$

- But $\mathrm{Ad}(B)$ acting on $T_e(G)$ is just the linear map of the tangent space induced by the mapping $Z \to BZB^{-1}$ on $SO(n)$.
- It follows that $\mathrm{Ad}(B)$ takes the component matrix $A = (a_{ij})$ of X_e to BAB^{-1}.

Example (Cont'd)

- Now define on $T_e(G)$ an inner product (X_e, Y_e).
- For $X_e = \sum a_{ij} \frac{\partial}{\partial x_{ij}}$ and $Y_e = \sum c_{ij} \frac{\partial}{\partial x_{ij}}$, define

$$(X_e, Y_e) = \text{tr} A'C = \sum_{i,j=1}^{n} a_{ij} c_{ij}.$$

- This product is clearly bilinear and symmetric.
- Moreover, we have

$$(X_e, X_e) = \text{tr} A'A = \sum_{i,j} a_{ij} a_{ij} = \sum a_{ij}^2.$$

- So the product is also positive definite.

Example (Cont'd)

- Finally for $B \in SO(n)$, we have

$$\begin{aligned}(\mathrm{Ad}(B)X_e, \mathrm{Ad}(B)Y_e) &= \mathrm{tr}((BAB^{-1})'BCB^{-1}) \\ &= \mathrm{tr}(BACB^{-1}) \\ &= \mathrm{tr}AC \\ &= (X_e, Y_e).\end{aligned}$$

- This means that this inner product determines a bi-invariant Riemannian metric on G.
- By a previous theorem, G is a symmetric space with this Riemannian metric.
- A similar procedure may be employed to obtain the bi-invariant Riemannian metric for other compact matrix groups.

Isometry Associated to a Geodesic

- Let M be any symmetric Riemannian manifold.
- Let $p(t)$, $-\infty < t < \infty$, be any geodesic on M.
- The symmetry $\sigma_{p(t)}$ associated with any point of this geodesic maps the geodesic onto itself and reverses its sense.
- Let c be a fixed real number.
- We denote by τ_c the following composition of two such isometries

$$\tau_c = \sigma_{p(c)} \circ \sigma_{p(c/2)}.$$

- τ_c maps the geodesic onto itself and preserves its sense.
- So its restriction to the geodesic must be of the form

$$\tau_c(p(t)) = p(t + \text{constant}).$$

- But we have

$$\tau_c(p(0)) = \sigma_{p(c)} \circ \sigma_{p(c/2)}(p(0)) = \sigma_{p(c)} p(c) = p(c).$$

- So we see that the constant is c and $\tau_c(p(t)) = p(t + c)$.

Action on the Tangent Space at a Point

- We consider how τ_c acts on the tangent space at a point of $p(t)$.
- Suppose that $X_{p(0)} \in T_{p(0)}(M)$.
- Define a vector field $X_{p(t)}$ along $p(t)$ by the formula

$$X_{p(t)} = \tau_{t*} X_{p(0)}.$$

- Let $X'_{p(t)}$ be the unique vector field satisfying

$$X'_{p(0)} = X_{p(0)}$$

  which is constant along the geodesic $p(t)$.
- We wish to show that these two vector fields coincide.

Action on the Tangent Space at a Point (Cont'd)

- For any real number t_0, $\sigma_{p(t_0)}$ is an isometry.
- Therefore, $\sigma_{p(t_0)*}X'_{p(t)}$ is a parallel vector field along $p(t)$.
- On the other hand, $p(t_0)$ is the fixed point of the symmetry.
- So
$$\sigma_{p(t_0)*}X'_{p(t_0)} = -X'_{p(t_0)}.$$
- But $-X'_{p(t)}$ is also a constant vector field along $p(t)$ and agrees with the field $\sigma_{p(t_0)*}X'_{p(t)}$ at one point.
- So it must agree with $\sigma_{p(t_0)*}X'_{p(t)}$ everywhere.
- Applying this argument twice we see that, for all t and each constant c,
$$\tau_{c*}X'_{p(t)} = X'_{p(t+c)}.$$
- Letting $t = 0$ and $c = t$ proves our assertion.

Summarizing in a Theorem

Theorem

Let M be a symmetric manifold. Let $p(t)$, $-\infty < t < \infty$, be a geodesic of M. Let τ_c be the associated isometry, for each real number c. Then

$$\tau_c(p(t)) = p(t+c).$$

If $X_{p(0)}$ is any element of $T_{p(0)}(M)$, then

$$X_{p(t)} = \tau_{t*} X_{p(0)}$$

is the associated parallel (constant) vector field along $p(t)$. That is, as t varies,

$$\tau_{t*} : T_{p(0)}(M) \to T_{p(t)}(M)$$

is the parallel translation along the geodesic.

Remark

- Let $p_1 = p(c_1)$ and $p_2 = p(c_2)$ be any two points of a geodesic
$$p(t), \quad -\infty < t < \infty.$$
- Then, by the same argument,
$$\sigma_{p_2} \circ \sigma_{p_1}(p(t)) = p(t + 2(c_2 - c_1)).$$
- Moreover,
$$(\sigma_{p_2} \circ \sigma_{p_1})_*$$
maps any parallel vector field along $p(t)$ to a parallel vector field.

Geodesics and One-Parameter Groups

Theorem

Let $M = G$, a compact, connected Lie group with the biinvariant metric. Let $X \in T_e(G)$. Then the unique geodesic $p(t)$ with $p(0) = e$ and $\dot{p}(0) = X_e$ is precisely the one-parameter subgroup determined by X_e. All other geodesics are left (or right) cosets of these one-parameter subgroups.

- Given a geodesic $p(t)$ with $p(0) = e$, we consider the isometry of G
$$\sigma_{p(s)}\sigma_{p(0)}.$$
By the remark above, this maps the geodesic onto itself with
$$p(t) \mapsto p(t + 2s).$$
Using our formula for σ_p on G together with $p(0) = e$, we have
$$\sigma_{p(s)}\sigma_{p(0)}p(t) = p(s)p(t)p(s),$$
the right-hand side being the group product of $p(s)$, $p(t)$ and $p(s)$.

Geodesics and One-Parameter Groups (Cont'd)

- Thus for all t, s,

$$p(s)p(t)p(s) = p(t+2s).$$

Using various t and mathematical induction, this gives, for arbitrary s and any integer n,

$$(p(s))^n = p(ns).$$

In particular, if a, b, c, d are integers with $bd \neq 0$, we have

$$p\left(\frac{a}{b}+\frac{c}{d}\right) = p\left(\frac{1}{bd}\right)^{ad+bc} = p\left(\frac{1}{bd}\right)^{ad} p\left(\frac{1}{bd}\right)^{bc} = p\left(\frac{a}{b}\right) \cdot p\left(\frac{c}{d}\right).$$

Thus, for any rational numbers, we have

$$p(r_1 + r_2) = p(r_1)p(r_2).$$

Geodesics and One-Parameter Groups (Cont'd)

- Now $p(t)$ depends continuously on t.

  So, for all real numbers,

$$p(r_1 + r_2) = p(r_1)p(r_2).$$

  Thus, any geodesic with $p(0) = e$ is a one-parameter subgroup.

  However, there is exactly one geodesic and one such subgroup with given $\dot{p}(0) = X_e$.

  So the first sentence of the theorem is true.

  The second follows at once if we use the following facts:
  - Either left or right translations are isometries, and hence preserve geodesics;
  - A geodesic through any $g \in G$ is uniquely determined (with its parametrization) by its tangent vector at g.

Lie Groups and One-Parameter Subgroups

Corollary

Let G be a compact Lie group. Then any $g \in G$ lies on a one-parameter subgroup.

- With the bi-invariant Riemannian metric G is a symmetric Riemannian manifold.

  Moreover, it is complete.

  Hence, any pair of points can be joined by a geodesic.

  If $g \in G$, then, by the theorem, the geodesic segment from e to g is on a one-parameter subgroup.

Example (Cont'd)

- Consider again $G = SO(n)$.
- Then the geodesics, relative to the bi-invariant metric of a previous example are the curves
$$p(t) = e^{tA},$$
where A any skew symmetric matrix, and their cosets.

Lie Derivative and Riemannian Differentiation

Theorem

Let G be a group with a bi-invariant metric.
Let X and Y are left-invariant vector fields on G.
Let ∇ be the Riemannian differentiation operator of vector fields.
Then we have

$$\nabla_X Y = \frac{1}{2}[X, Y] = \frac{1}{2}L_X Y.$$

- Suppose that Z is any left-invariant vector field.

  Then we will compute $\nabla_{Z_e} Z$.

  Let $g(t)$ is the uniquely determined one-parameter group with

$$g(0) = e \quad \text{and} \quad \dot{g}(0) = Z_e.$$

Lie Derivative and Riemannian Differentiation (Cont'd)

- Then for any vector field Y we have

$$\nabla_{Z_e} Y = \frac{DY_{g(t)}}{dt}\Big|_{t=0}.$$

On the other hand,

$$Z_{g(t)} = \frac{dg}{dt}$$

and $g(t)$ is a geodesic.

Thus,

$$\frac{DZ_{g(t)}}{dt} = \frac{D}{dt}\frac{dg}{dt} = 0.$$

So $\nabla_{Z_e} Z = 0$.

Now Z and the metric are left-invariant.

Lie Derivative and Riemannian Differentiation (Cont'd)

- So we get
$$\nabla_Z Z = 0$$
everywhere on G.
Thus,
$$\nabla_{X+Y}(X+Y) = 0.$$
We conclude that
$$\nabla_X Y + \nabla_Y X = 0.$$
On the other hand, we know that for any pair of vector fields a Riemannian connection satisfies the identity
$$\nabla_X Y - \nabla_Y X = [X, Y].$$
Combining these two identities gives the conclusion.

Subsection 9

Some Examples

Lie Groups, Manifolds and Riemannian Metrics

Theorem

Let G be a Lie group acting transitively on a manifold M.
Then M has a Riemannian metric such that the transformation determined by each element of G is an isometry if the isotropy group H of a point $p \in M$ is a connected compact (Lie) subgroup of G.

- Let $\theta : G \times M \to M$ denote the action.

  For each $g \in G$, $\theta_g : M \to M$ denotes the diffeomorphic transformation of M onto itself determined by g,

  $$\theta_g(q) = \theta(g, q).$$

  If $g \in H$, then $\theta_g(p) = p$.

  So θ_g induces a linear mapping $\theta_{g*} : T_p(M) \to T_p(M)$.

Lie Groups, Manifolds and Riemannian Metrics (Cont'd)

- We have $\theta_{g_1} \circ \theta_{g_2} = \theta_{g_1 g_2}$.

  So we have

  $$\theta_{g_1 *} \circ \theta_{g_2 *} = \theta_{g_1 g_2 *}.$$

  So $g \to \theta_{g*}$ is a homomorphism of H into the group of linear transformations on $T_p(M)$.

  By hypothesis, θ is C^∞.

  So $g \to \theta_{g*}$ is a C^∞ homomorphism.

  Thus, it is a representation of H on $T_p(M)$.

  Now H is compact and connected.

  So, by previous results, there must be an invariant inner product, which we shall denote by $\Phi_p(X_p, Y_p)$ on $T_p(M)$.

Lie Groups, Manifolds and Riemannian Metrics (Cont'd)

- If $q \in M$, there is a $g \in G$ such that $\theta_g(q) = p$.
  We define $\Phi_q(X_q, Y_q)$ by
  $$\Phi_q(X_q, Y_q) = \theta_g^* \Phi_p(X_q, Y_q) = \Phi_p(\theta_{g*} X_q, \theta_{g*} Y_q).$$
  If $\theta_{g_1}(q) = p$ also, then $gg_1^{-1} \in H$.
  Hence $\theta^*_{gg_1^{-1}} \Phi_p = \Phi_p$ and
  $$\theta^*_{g_1} \Phi_p = \theta^*_{g_1} \theta^*_{gg_1^{-1}} \Phi_p = \theta^*_{g_1} \circ \theta^*_{g_1^{-1}} \circ \theta^*_g \Phi_p = \theta^*_g \Phi_p.$$
  It follows that Φ_q is well defined.
  It is positive definite, since θ_g is a diffeomorphism.
  It is easily verified that Φ is C^∞ and G-invariant on M.
  Thus Φ defines a Riemannian metric on M with respect to which each θ_g is an isometry of M.

Lie Groups, Manifolds and Symmetry

Theorem

Let G be a Lie group acting transitively and effectively on a manifold M. Let $p \in M$, with its isotropy group H a connected compact subgroup. Let $\alpha : G \to G$ be an involutive automorphism of G whose fixed set is H. Then the correspondence

$$\widetilde{\alpha}(\theta(g, p)) = \theta(\alpha(g), p)$$

defines an involutive isometry of M onto M with p as an isolated fixed point.

- First we check that $\widetilde{\alpha}$ actually defines a mapping of M onto itself.

  Let q be an arbitrary point of M.

  By transitivity, there is at least one $g \in G$, such that $\theta(g, p) = q$.

Lie Groups, Manifolds and Symmetry (Cont'd)

- If g' is a second such element, then $g' = gh$ and

$$\alpha(g') = \alpha(g)\alpha(h) = \alpha(g)h.$$

Hence

$$\begin{array}{rcl} \widetilde{\alpha}(\theta(g', p)) & = & \theta(\alpha(g'), p) \\ & = & \theta(\alpha(g)h, p) \\ & = & \theta(\alpha(g), \theta(h, p)) \\ & = & \theta(\alpha(g), p). \end{array}$$

Therefore $\widetilde{\alpha}$ is defined independently of any choices.

Since $\widetilde{\alpha}^2$ is the identity, $\widetilde{\alpha}$ is onto.

Lie Groups, Manifolds and Symmetry (Cont'd)

- Let us assume for the moment that we have proved that:
  - $\widetilde{\alpha}$ is C^∞;
  - $\widetilde{\alpha}$ has p as an isolated fixed point;
  - $\widetilde{\alpha}_* : T_p(M) \to T_p(M)$ is $-I$, that is, $\widetilde{\alpha}_*(X_p) = -X_p$.

  Then, clearly, $\widetilde{\alpha}_*$ preserves the inner product Φ_p on $T_p(M)$.

  If $q \in M$, $q \neq p$, then choose $g \in G$, such that $\theta_g(p) = q$.

  Then
  $$\widetilde{\alpha}(q) = \theta(\alpha(g), p) = \theta_{\alpha(g)}(\theta_{g^{-1}}(q)).$$

  Hence $\widetilde{\alpha}_{*q} : T_q(M) \to T_{\widetilde{\alpha}(q)}(M)$ is given by
  $$\widetilde{\alpha}_{*q} = \theta_{\alpha(g)*} \circ \theta_{g^{-1}*}.$$

  Both $\theta_{\alpha(g)*}$ and $\theta_{g^{-1}*}$ are isometries on the tangent spaces.

  Thus, subject to checking the other properties, $\widetilde{\alpha}$ is an isometry.

Lie Groups, Manifolds and Symmetry (Cont'd)

- We aim to verify the remaining properties.

  For this, we need to use the fact that the natural identification of M with G/H, given by the mapping $F : G/H \to M$,

  $$F(gH) = \theta(g, p),$$

  is C^∞ and commutes with left translation on G/H.

  Thus, we use a previous application of Frobenius' Theorem.

  First we recall that, if $gH \in G/H$, then there is a C^∞ section S defined on a neighborhood V of $gH, S : V \to G$, with $\pi \circ S = \text{id}$ ($\pi : G \to G/H$ is the natural projection and id the identity on V).

  Using the diffeomorphism F, obtain a C^∞ section

  $$\widetilde{S} = S \circ F^{-1}$$

  on $\widetilde{V} = F(V)$ into G.

Lie Groups, Manifolds and Symmetry (Cont'd)

- This means a C^∞ mapping such that

$$\theta(\widetilde{S}(q), p) = q, \quad \text{for all } q \in \widetilde{V}.$$

Every point of M is contained in the domain $\widetilde{V}$ of such a section.

Moreover, $\widetilde{\alpha}|_{\widetilde{V}}$ is given by

$$\widetilde{\alpha}(q) = \widetilde{\alpha}(\theta(\widetilde{S}(q), p)) = \theta(\alpha(\widetilde{S}(q)), p).$$

This is a composition of C^∞ mappings.

It follows that $\widetilde{\alpha}$ is C^∞.

Lie Groups, Manifolds and Symmetry (Cont'd)

- Finally we wish to show that $\widetilde{\alpha}$ has p as an isolated fixed point and that $\widetilde{\alpha}_{*p} = -I$.

  We use facts demonstrated previously concerning the exponential mapping $\exp : T_e(G) \to G$ (not to be confused with the exponential mapping Exp of Riemannian manifolds).

  Given any $X_p \in T_e(G)$, then

$$\exp tX_p = g(t)$$

  is the one-parameter subgroup of G with $\dot{g}(0) = X_e$.

  Further, $\exp X_p = g(1)$.

  By a previous theorem, there is an $\varepsilon > 0$ such that an ε-ball $B^n_\varepsilon(0) \subseteq T_e(M)$ is mapped diffeomorphically onto a neighborhood U of e, the identity of G.

Lie Groups, Manifolds and Symmetry (Cont'd)

- Now $\alpha : G \to G$ is a Lie group automorphism with α^2 the identity.

  So $\alpha_* : T_e(G) \to T_e(G)$ splits $T_e(G)$ into the direct sum of two subspaces $V^{\pm}$ of characteristic vectors belonging to the characteristic values ± 1 of α_*.

  We have $\alpha(\exp t X_e) = \exp t \alpha_*(X_e)$.

  So $\alpha_*(X_e) = X_e$ if and only if $X_e \in T_e(H)$.

  Thus
  $$T_e(G) = V^+ \oplus V^-, \quad V^+ = T_e(H).$$

  $\pi : G \to G/H$ defines $\pi_* : T_e(G) \to T_{\pi(e)}(G/H)$, with:
  - $\ker \pi_* = T_e(H)$;
  - $\pi_*|_{V^-}$ an isomorphism onto.

  So $\pi \circ \exp$ maps a neighborhood W of $V^- \cap B^n_\varepsilon(0) \subseteq T_e(G)$ diffeomorphically onto a neighborhood of H in G/H.

Lie Groups, Manifolds and Symmetry (Cont'd)

- Composing with $F : G/H \to M$ gives a diffeomorphism onto an open set around p.

  Thus, for $X_e \in W$, the mapping

  $$X_e \to \theta(\exp X_e, p)$$

  is a diffeomorphism.

  Moreover,

  $$\widetilde{\alpha}(\theta(\exp X_e, p)) = \theta(\alpha(\exp X_e), p) = \theta(\exp(-X_e), p).$$

  It follows that:
  - p is the only fixed point of $\widetilde{\alpha}$ in this neighborhood;
  - $\widetilde{\alpha}_* : T_p(M) \to T_p(M)$ is $-I$, that is, each vector is taken to its negative.

  This, taken with preceding work, completes the proof.

Corollary

- The following corollary is immediate, since each $\theta_g : M \to M$ is an isometry.

Corollary

Let G be a Lie group acting transitively and effectively on a manifold M. Let $p \in M$, with its isotropy group H a connected compact subgroup. Let $\alpha : G \to G$ be an involutive automorphism of G whose fixed set is H. Consider the correspondence

$$\widetilde{\alpha}(\theta(g,p)) = \theta(\alpha(g),p).$$

The manifold M is a symmetric space, with involutive isometries

$$\sigma_p = \widetilde{\alpha} \quad \text{and} \quad \sigma_q = \theta_g \circ \widetilde{\alpha} \circ \theta_{g^{-1}}, \quad q = \theta(g,p).$$

Example

- Let M be the collection of all $n \times n$, symmetric, positive definite, real matrices of determinant $+1$.
- Let $G = Sl(n, \mathbb{R})$ be the $n \times n$ matrices of determinant $+1$.
- Then G acts on M by
$$\theta(g, s) = gsg',$$
where g' denotes the transpose of $g \in Sl(n, \mathbb{R})$.
- Let p, the base point of the theorems be I, the $n \times n$ identity.
- We then note that
$$H = \{g \in Sl(n, \mathbb{R}) : \theta(g, I) = I\}$$
is given by the equivalent condition $gg' = I$.
That is, $g \in SO(n)$, the group of orthogonal $n \times n$ matrices.
Hence, $H = SO(n)$.
- So M is canonically identified with $Sl(n, \mathbb{R})/SO(n)$.

Example (Cont'd)

- The automorphism α which we consider is defined by

$$\alpha(g) = (g^{-1})',$$

the transpose of the inverse of $g \in Sl(n, \mathbb{R})$.

- Note that $\alpha(g) = g$ if and only if $g \in SO(n)$.
- Thus all of the conditions of the theorem are met if $Sl(n, \mathbb{R})$ is transitive on M.
- However, any positive definite, symmetric matrix q may be written in the form

$$q = gg' = gIg',$$

where $g \in Sl(n, \mathbb{R})$, by standard theorems of linear algebra.

- From the corollary above M is a symmetric space relative to an $Sl(n, \mathbb{R})$ invariant metric.

Example (Cont'd)

- Note that $\widetilde{\alpha} : M \to M$ can be seen, quite directly, to be C^∞ and to have the identity $p = I$ as its only fixed point on M.
- In fact, using $q = sIs'$, we see that

$$\widetilde{\alpha}(q) = \widetilde{\alpha}\theta(s, I) = \theta(s'^{-1}, I) = s'^{-1}s^{-1} = (ss')^{-1} = q^{-1}.$$

- Thus $\widetilde{\alpha} : M \to M$ simply takes each positive definite symmetric matrix to its inverse.
- The only such matrix which is equal to its inverse is the identity I.

Example

- We look at a variant on the above which is a particularly important case.
- Consider the upper half-plane

$$M = \{(x, y) \in \mathbb{R}^2 : y > 0\}.$$

- We define an action of $Sl(2, \mathbb{R})$ on M as follows.
- We identify $\mathbb{R}^2$ with $\mathbb{C}$, the complex numbers in the usual way.
- Let $z = x + iy$ and let $w = u + iv$, $i = \sqrt{-1}$.
- Let $g \in Sl(2, \mathbb{R})$, that is, $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $ad - bc = +1$.
- We then define

$$w = \theta(g, z) = \frac{az + b}{cz + d}.$$

Example (Cont'd)

- It is not difficult to verify directly that:
  - If $y = \text{Im}(z) > 0$, then $v = \text{Im}(w) > 0$;
  - $\theta(g_1, \theta(g_2, z)) = \theta(g_1 g_2, z)$.
- Moreover the Riemannian metric defined (in the local coordinates (x, y) - or $z = x + iy$ - which cover M) by the matrix of components

$$(g_{ij}) = \begin{pmatrix} \frac{1}{y^2} & 0 \\ 0 & \frac{1}{y^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{(\text{Im}(z))^2} & 0 \\ 0 & \frac{1}{(\text{Im}(z))^2} \end{pmatrix}$$

  is invariant under the action of $Sl(2, \mathbb{R})$.
- Thus this group acts on M as a group of isometries of this metric.
- Let the complex number i which corresponds to $(0, 1)$ in $\mathbb{R}^2$, play the role of p in the general discussion above.

Example (Cont'd)

- The action of $Sl(2,\mathbb{R})$ is transitive.
- Consider any $z_0 = u + vi$ with $v > 0$.
- Then an element of $G = Sl(2,\mathbb{R})$ taking i to z_0 is

$$g = \begin{pmatrix} \sqrt{v} & \frac{u}{\sqrt{v}} \\ 0 & \frac{1}{\sqrt{v}} \end{pmatrix}.$$

- This gives, in general,

$$\theta(g,z) = \frac{\sqrt{v}z + \frac{u}{\sqrt{v}}}{0z + \frac{1}{\sqrt{v}}} = vz + u.$$

- When $z = i$,

$$\theta(g,i) = u + iv.$$

Example (Cont'd)

- The isotropy group of i consists of all $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, such that
$$i = \frac{ai + b}{ci + d}.$$
- We get $ai + b = -c + di$.
- Equivalently, $a = d$ and $b = -c$.
- Since in addition $ad - bc = +1$, we have also
$$a^2 + b^2 = 1.$$
- Hence
$$g = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$
- This gives $H = SO(2)$.

Example (Cont'd)

- It follows from our general theory that the upper half-plane with this geometry and the 2×2 positive definite matrices are equivalent, both as manifolds and as homogeneous spaces, with

$$Sl(2, \mathbb{R})/SO(2).$$

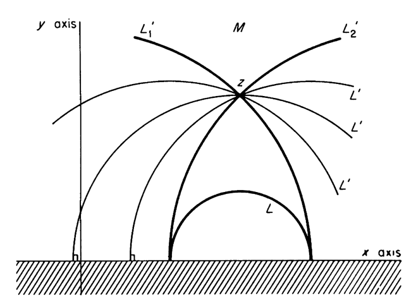
- This shows that the identification of a homogeneous space with a coset space of a Lie group as a prototype is a deeper and more interesting result than it might appear to be.

Non-Euclidean Geometries

- The example of the upper half-plane is a realization (due to Poincaré) of the space of non-Euclidean geometry of Bolyai, Lobachevskii and Gauss.
- We can use our results to check that the lines $x =$ constant are geodesics in this geometry.
- Another problem consists of showing that the upper halves of circles with centers on the x-axis are - when suitably parametrized - also geodesics.
- This is done by showing that each such circle is an image by one of the isometries of G of a vertical line.
- Through a point z there is such a circle tangent to any direction.
- Hence, these must be all of the geodesics.
- Using this fact it is easy to see that Euclid's postulate of parallels does not hold in this geometry.

Non-Euclidean Geometries (Illustration)

- There are more than one, in fact an infinite number of lines through a point z not on the line L which are parallel to L, that is, do not intersect L at any point of the upper half-plane M.
- The possibilities are shown in the figure.



- L_1' and L_2' indicate parallel lines (geodesics) which bound the infinite collection (faint lines) of lines L' parallel to L through z.

Example

- As a last example of a symmetric space, we mention the Grassmann manifold $G(k, n)$ of k-planes through the origin of $\mathbb{E}^n$.
- We have noted that this is a homogeneous manifold.
- Moreover, it is acted on in a natural way by $Gl(n, \mathbb{R})$.
- It is easy to see that the subgroup $SO(n, \mathbb{R})$ also acts transitively on the k-planes in $\mathbb{R}^n$.
- In fact, a k-plane contains an orthonormal basis $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k$ which can be completed to an orthonormal, oriented basis $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n$ of $\mathbb{R}^n$.
- Then there exists an orthogonal transformation of determinant $+1$ taking the standard basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ to this one.
- Hence the k-plane P_0 spanned by $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k$ is carried onto any k-plane P by at least one element of $SO(n, \mathbb{R})$ acting in the natural way.

Example (Cont'd)

- The isotropy group H of P_0 is $S(O(k) \times O(n-k))$.
- That is, it consists of the matrices in $SO(n)$ of the form

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}, \quad A \in O(k), \quad B \in O(n-k),$$

with

$$\det A \, \det B = +1.$$

- One can show that, in this case, α is the automorphism

$$\alpha : x \mapsto gxg^{-1},$$

determined by the element $g = \begin{pmatrix} -I_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$ of $Gl(n, \mathbb{R})$.
- Moreover, $\alpha(x) = x$ if and only if $x \in H$.