CHAPTER 1

Basic Concepts and Results

1. Notation and preliminaries

We will begin by fixing some notation and recalling some standard facts about connections. M will denote a smooth (i.e. C^∞) finite-dimensional manifold and $T(M)$ its tangent bundle. Sometimes we write M^n to indicate that M has dimension n. For basic concepts about manifolds, such as the notions of submanifolds, immersion, embedding and diffeomorphism, we refer to Kobayashi and Nomizu [1963, 1969]. For $p \in M$, M_p denotes the tangent space to M at p. $\chi(M)$ or χ is the linear space of smooth vector fields on M and $\mathfrak{F}(M)$ the ring of smooth functions on M. We will usually denote a tangent vector at a point by a lower case letter and its extension to a vector field by the corresponding capital.

A Riemannian metric is an assignment to each $p \in M$ of a symmetric positive-definite bilinear form $\langle\ ,\ \rangle_p$ on M_p such that for any $V, W \in \chi(M)$ the function $p \to \langle V, W\rangle_p$ is in $\mathfrak{F}(M)$. Also, $\langle V, V\rangle^{\frac{1}{2}}$ is denoted by $||V||$.

An *affine connection* is a bilinear map

$$\nabla : \chi(M) \times \chi(M) \to \chi(M),$$

which has the following properties:

$$\nabla_{fV} W = f \nabla_V W, \tag{1.1}$$

$$\nabla_V fW = (Vf)W + f\nabla_V W \tag{1.2}$$

for any $f \in \mathfrak{F}(M)$, $V, W \in \chi(M)$.

We call $\nabla_V W$ the *covariant derivative* of W in the direction of V. We refer to Bishop and Crittenden [1964] and Kobayashi and Nomizu [1963, 1969] for more detailed versions of connection theory.

The Fundamental Theorem of Riemannian Geometry states that for each Riemannian metric there is a unique affine connection, called the *Riemannian connection*, with the following two properties:

$$X\langle V, W\rangle = \langle \nabla_X V, W\rangle + \langle V, \nabla_X W\rangle, \tag{1.3}$$

$$\nabla_V W - \nabla_W V - [V, W] = 0, \tag{1.4}$$

where $[\ ,\ ]$ signifies the Lie bracket, $[X, Y]f = (XY - YX)f$. The first of these properties is a condition of compatibility between the affine connection and the metric, while the second is a symmetry condition on the connection alone. In general, the quantity set equal to zero in (1.4) is called the *torsion* of the connection, $\mathrm{Tor}(v, w)$. It is a tensor of type (1,2). Hence the Fundamental Theorem may be paraphrased by saying that there is a unique torsion-free connection compatible with any given metric. We recall the proof.

To show uniqueness, it suffices to show that $\langle \nabla_X Y, Z\rangle$ is determined by (1.3), (1.4). Using (1.4),

$$\begin{aligned} X\langle Y, Z\rangle &= \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle, \\ Y\langle X, Z\rangle &= \langle \nabla_Y X, Z\rangle + \langle X, \nabla_Y Z\rangle, \\ Z\langle X, Y\rangle &= \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle, \end{aligned}$$

Subtracting the third of these equations from the sum of the first two and using (1.4) yields

$$\begin{aligned} 2\langle \nabla_X Y, Z\rangle = X\langle Y, Z\rangle &+ Y\langle X, Z\rangle - Z\langle X, Y\rangle \\ &+ \langle [X, Y], Z\rangle - \langle [X, Z], Y\rangle - \langle [Y, Z], X\rangle. \end{aligned}$$

Conversely, if we use this formula to define $\langle \nabla_X Y, Z\rangle$, it is straightforward to check that we get a connection satisfying (1.3) and (1.4).

From (1.1), (1.2) it is easy to see that $(\nabla_V W)(p)$ is determined by W and $V(p)$. In fact it is not hard to show that it is determined by $V(p)$ and W restricted to any curve through p in direction $V(p)$. If α is a 1-form, we define $\nabla_V \alpha$ by the equation

$$(\nabla_V \alpha)(W) + \alpha(\nabla_V W) = V(\alpha(W)).$$

We then define $\nabla_V \sigma$ for any tensor field σ by extending ∇ as a derivation.

Let $c : [0, 1] \to M$ be a smooth curve, and let $c'(t)$ denote the tangent vector to $c(t)$. For any $v \in M_{c(0)}$ there is a unique vector $V(t) \in M_{c(t)}$ such that $V(0) = v$ and $\nabla_{c'(t)} V(t) \equiv 0$ We call $V(t)$ the *parallel field* and $V(1)$ the *parallel translate* of v along $c(t)$. In fact, one can easily find smooth vector fields along $c(t)$, $E_1(t), \dots, E_n(t)$ such that $\{E_i(t)\}$ is an orthonormal basis for $M_{c(t)}$. In terms of these fields, the above equation may be written as a first-order system of ordinary differential equations. We have

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle V, E_i\rangle = c'\langle V, E_i\rangle = \langle \nabla_{c'} V, E_i\rangle + \langle V, \nabla_{c'} E_i\rangle = \langle \nabla_{c'} E_i, V\rangle.$$

Therefore

$$\begin{bmatrix} \langle V, E_1\rangle \\ \vdots \\ \langle V, E_n\rangle \end{bmatrix}' = \langle \nabla_{c'} E_i, E_j\rangle \begin{bmatrix} \langle V, E_1\rangle \\ \vdots \\ \langle V, E_n\rangle \end{bmatrix}.$$

The existence theorem for ordinary differential equations tells us that we can solve for $V(t)$. Because the above equation is linear, we obtain a linear map $P_c : M_{c(0)} \to M_{c(1)}$ defined by $v \to V(1)$. It follows easily from (1.3) that P_c is an isometry.

In the sequel we will often be confronted with the following situation: Let $\phi : N \to M$ be a smooth map, and let M have a connection ∇. A *vector* *field along* ϕ is an assignment $x \to W$, where $W \in M_{\phi(x)}$. Let $\{E_i\}$ be a frame field in a neighborhood of $\phi(x)$. Then we can write

$$W(x) = \Sigma f_i(x) E_i(\phi(x)).$$

We say that $W(x)$ is *smooth* if the functions $f_i(x)$ are smooth. If $v \in N_x$, we can define $\tilde{\nabla}_v W \in M_{\phi(x)}$, the covariant derivative of W in the direction of v, by

$$\tilde{\nabla}_v W = \Sigma v(f_i(x)) E_i(\phi(x)) + f_i(x) \nabla_{\mathrm{d}\phi(v)} E_i(\phi(x)).$$

The definition is easily seen to be independent of the choice of $\{E_i(x)\}$. Let M be Riemannian and ∇ the Riemannian connection. If $v \in N_x$, W_1, W_2 are vector fields along ϕ, then one easily checks that

$$v\langle W_1, W_2\rangle = \langle \tilde{\nabla}_v W_1, W_2\rangle + \langle W_1, \tilde{\nabla}_v W_2\rangle. \qquad (*)$$

Also, if V_1, V_2 are vector fields in N, then $\mathrm{d}\phi(V_1)$, $\mathrm{d}\phi(V_2)$ are vector fields along ϕ and

$$\tilde{\nabla}_{V_1} \mathrm{d}\phi(V_2) - \tilde{\nabla}_{V_2} \mathrm{d}\phi(V_1) - \mathrm{d}\phi([V_1, V_2]) = 0. \qquad (**)$$

A vector field along ϕ is also called a *section* of the induced bundle $\phi^*(TM)$. We call $\tilde{\nabla}$ the *induced connection*. In the proofs of the first and second variation formulas and elsewhere, implicitly we will be using (*) and (**). However, for convenience we will suppress the notation $\tilde{\nabla}$ and proceed formally as if the vector fields along ϕ were actually defined on M.

2. First variation of arc length

Let M be a Riemannian manifold. We denote the length of the continuous piecewise smooth curve $c : [a, b] \to M$ by $L[c]$. By definition,

$$L[c] = \int_a^b ||c'(t)|| \mathrm{d}t.$$

It follows from the chain rule that $L[c]$ does not depend on a particular choice of parameterization. M becomes a metric space if we define the distance between two points as the infimum of the lengths of all curves between them. We denote the distance from p to q by $\rho(p, q)$. When proving that M is a metric space, the only point which is not entirely trivial to check is that if $p \neq q$, then $\rho(p, q) > 0$. To see this, let $x_1, \ldots, x_n$ be a local coordinate system with p at the origin, and let $B_r(p)^-$ denote the set $\Sigma x_i^2 \leq r$. Let g denote the given Riemannian metric and $\bar{g}$ the Euclidean metric $\bar{g}(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}) = \delta_{ij}$. At each point, g may be diagonalized with respect to $\bar{g}$, and the smallest eigenvalue $\lambda(x)$ is a positive continuous function on $B_r(p)^-$. It follows easily that any curve from p to the boundary of $B_r(p)^-$ has length at least $\frac{r}{\lambda_0}$, where λ_0 is the minimum of $\lambda(x)$. However, given $q \neq p$, we may assume that r has been chosen sufficiently small so that $q \notin B_r(p)^-$. Then any curve from p to q must contain an initial segment from p to the boundary of $B_r(p)^-$. Therefore $\rho(p, q) > \frac{r}{\lambda_0}$. A similar argument shows that the metric space and manifold topologies coincide.

Let $p, q \in M$. In the next three sections we will investigate the conditions under which there exists a curve γ from p to q such that $L[\gamma] = \rho(p,q)$. We are now going to derive a necessary condition for the existence of such a curve. We begin by assuming that c is smooth, $||c'(t)|| \neq 0$, and without loss of generality we may assume c to be parameterized proportional to arc length. In other words, $||c'(t)||$ is a constant l.

Let $\alpha : Q \to M$ be a smooth function, where Q is the rectangle $[a,b] \times (-\epsilon, \epsilon)$ and such that

$$\alpha|[a,b] \times \{0\} = c : [a,b] \to M.$$

α is said to define a *smooth variation* of the curve $\alpha|[a,b] \times \{0\}$. Let T, V be the fields of tangent vectors on Q corresponding to its first and second variables. We shall identify these vectors with their images under the differential of α. Our goal is to compute the change in arc length over the family of curves $c_s = \alpha|[a,b] \times \{s\}$, where $-\epsilon < s < \epsilon$. This is given by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s} L[c_s] &= \frac{\mathrm{d}}{\mathrm{d}s} \int_a^b \langle c_s'(t), c_s'(t) \rangle^{\frac{1}{2}} \mathrm{d}t = \int_a^b V \langle T, T \rangle^{\frac{1}{2}} \mathrm{d}t \\ &= \frac{1}{2} \int_a^b \langle T, T \rangle^{-\frac{1}{2}} V \langle T, T \rangle \mathrm{d}t = \int_a^b \langle T, T \rangle^{-\frac{1}{2}} \langle \nabla_V T, T \rangle \mathrm{d}t.\end{aligned}$$

Since $[T, V] = 0$ on Q, using (**) we may rewrite this as

$$\int_a^b \langle T, T \rangle^{-1/2} \langle \nabla_T V, T \rangle \mathrm{d}t.$$

Since $||c_0'|| = l$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s} L[c_s]|_{s=0} &= l^{-1} \int_a^b \langle \nabla_T V, T \rangle \mathrm{d}t \\ &= l^{-1} \int_a^b (T \langle V, T \rangle - \langle V, \nabla_T T \rangle \mathrm{d}t.\end{aligned}$$

By integrating the first term we obtain

$$\frac{\mathrm{d}}{\mathrm{d}s} L[c_s]|_{s=0} = l^{-1} \Big\{ \langle V, T \rangle|_a^b - \int_a^b \langle V, \nabla_T T \rangle \mathrm{d}t \Big\}. \tag{1.5}$$

This expression is called the *first variation formula*.

Suppose that rather than being smooth, the function α above is continuous, and has the property that $a = t_0 < t_1 < \cdots < t_n = b$ is some subdivision of $[a,b]$ such that $\alpha_{[t_i, t_{i+1}] \times (-\epsilon, \epsilon)}$ is smooth. In this case, α is said to define a *piecewise smooth* variation. The first variation formula for piecewise smooth variations is obtained by applying (1.5) to each segment $[t_i, t_{i+1}]$. In particular, if c_0 is actually smooth, the intermediate terms cancel, and (1.5) remains valid. A vector field V along c_0 is called piecewise smooth if V is continuous and there exists a subdivision as above such that $V|[t_i, t_{i+1}]$ is smooth vector field along c_0. It is important to remark that any piecewise smooth V arises from some variation. In fact, for sufficiently small s, the variation $\alpha : (t, s) \to \exp_{c_0(t)}(sV)$ will suffice. (See Section 3 for the definition of exp).

If all curves c_s have the same endpoints, then $V(a, 0) = V(b, 0) = 0$, so that

$$\frac{\mathrm{d}}{\mathrm{d}s} L[c_s]|_{s=0} = -l^{-1} \int_a^b \langle V, \nabla_T T \rangle \mathrm{d}t.$$

If $c = c_0$ is the shortest curve from $c(a)$ to $c(b)$, then

$$\frac{\mathrm{d}}{\mathrm{d}s} L[c_s]|_{s=0} = 0$$

for any map $\alpha : Q \to M$. Hence for V any vector field along c which vanishes at the endpoints, the right-hand side of (1.5) vanishes. Therefore, if c is smooth and minimal, by taking a variation such that $V = \phi(t)\nabla_T T$ for some function $\phi(t)$ such that $\phi(t) > 0$ for $a < t < b$ and $\phi(a) = \phi(b) = 0$, we conclude that

$$\nabla_T T = \nabla_{c'} c' \equiv 0.$$

The preceding calculation motivates the following definition.

Definition 1.6. We call a smooth curve c a *geodesic* if $\nabla_{c'} c' \equiv 0$.

If c is a geodesic,

$$c' \langle c', c' \rangle = 2 \langle \nabla_{c'} c', c' \rangle = 0,$$

so $\langle c', c' \rangle$ must be constant. Thus a geodesic is parameterized proportional to arc length, and by (1.5) it is always a critical point of the arc-length function under any variation with fixed end points. In fact, for this we only need to know that the variation is perpendicular to T at the endpoints, or, more generally, that $\langle T, V \rangle|_a^b = 0$. If $||c'|| = 1$, c is called a *normal* geodesic.

The following useful proposition illustrates how the first variation formula may be applied to obtain geometrical information.

Proposition 1.7. *Let N and $\bar{N}$ be two submanifolds of M, without boundary, and let $\gamma : [0, t] \to M$ be a geodesic such that $\gamma(0) \in N$, $\gamma(t) \in \bar{N}$ and γ is the shortest curve from N to $\bar{N}$. Then $\gamma'(0)$ is perpendicular to $N_{\gamma(0)}$ and $\gamma'(t)$ is perpendicular to $\bar{N}_{\gamma(t)}$*

Proof. 1If $\gamma'(0)$ is not perpendicular to $N_{\gamma(0)}$, choose $x \in N_{\gamma(0)}$ such that $\langle \gamma'(0), x \rangle > 0$, and let c be a curve in N starting at $\gamma(0)$ such that $c'(0) = x$.

Construct a variation $\alpha : [0, t] \times (-\epsilon, \epsilon) \to M$ such that

$$\alpha|[0, t] \times \{0\} = \gamma, \qquad \alpha(0, s) = c(s), \alpha(t, s) = \gamma(t).$$

Then if $\gamma_s = \alpha|[0, t] \times \{s\}$, formula (1.5) shows that

$$\frac{\mathrm{d}}{\mathrm{d}s} L[\gamma_s]|_{s=0} = -l^{-1} \langle \gamma'(0), x \rangle < 0.$$

Therefore, for small s, $L[\gamma_s] < L[\gamma]$, and γ is not minimal. A completely analogous argument shows that $\gamma'(t)$ must also be perpendicular to $\bar{N}_{\gamma(t)}$. (See Fig. 1.1) $\square$

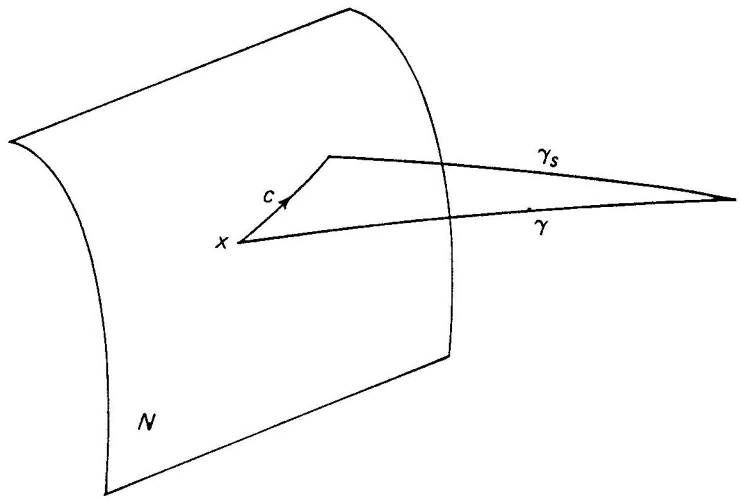

Fig. 1.1.

Note that in the above variation it is not necessary that the curves γ_s be geodesics. As long as α is a smooth map, our calculation shows that for small s, $L[\gamma_s] < L[\gamma]$.

3. Exponential map and normal coordinates

A fundamental property of geodesics is that, given a point $p \in M$ and a vector $v \in M_p$, there exists a unique geodesic γ_v through p whose tangent at p is v. This follows from the fact that the defining condition for geodesics, $\nabla_{\gamma'}\gamma' = 0$, is a second-order differential equation in the parameter t of γ. p and v are exactly the required initial conditions for existence and uniqueness of γ_v (see Milnor [1963], p.56, for details). If $\gamma_v : (-\epsilon, \epsilon) \to M$ is a geodesic parameterized by t, the curve $c : (-\frac{\epsilon}{s}, \frac{\epsilon}{s}) \to M$ defined by $c(t) = \gamma_v(st)$ (s fixed) is also a geodesic; for $\nabla_{c'}c' = s^2\nabla_{\gamma_v}\gamma_v = 0$. Also $c'(0) = sv$, so $c = \gamma_{sv}$.

The *exponential map* $\exp_p : M_p \to M$ is defined by $\exp_p(v) = \gamma_v(1)$ for all $v \in M_p$ such that 1 is in the domain of γ_v. From the above we know that for any fixed v there exists a number $s > 0$ such that $\gamma_{sv}(1)$ is defined, and by the existence theorem of second-order differential equations, we can pick s to vary continuously with v. It follows that $\exp_p$ is defined on a neighborhood of the origin in M_p. Furthermore, it is a smooth map and, by the implicit function theorem, a local diffeomorphism in a neighborhood of the origin. Actually, since we have $\exp_p : M_p \to M$ for each $p \in M$, we can define the union of these maps $\exp : T(M) \to M$. We shall define exp on the union over p of the domains of $\exp_p$, which is a neighborhood of the zero section of $T(M)$.

If we choose an orthonormal basis $\{e_i\}$ for M_p, we can define a coordinate system in a neighborhood of p by assigning to the point $\exp_p(\Sigma x^i e_i)$ the coordinates $(x_1, \ldots, x_n)$. Such coordinates are called *normal coordinates* at p. Since the rays through the origin are geodesics, normal coordinates have the property that $\nabla_{\frac{\partial}{\partial x_i}}(\frac{\partial}{\partial x_i}) \equiv 0$. It follows that for all v in M_p,

$$\left(\nabla_v \frac{\partial}{\partial x_i}\right)\Big|_p = 0.$$

For these reasons, normal coordinates are convenient to use.

If M is a Riemannian manifold, by definition each tangent space M_p comes equipped with an inner product. For each $v \in M_p$, the tangent space $(M_p)_v$ can be naturally identified with M_p. Hence $(M_p)_v$ inherits an inner product. Fixing $v \in M_p$, $\mathrm{dexp} : (M_p)_v \to M_{\exp(v)}$ is a linear map. dexp does not in general preserve the inner products on these spaces.

If we let $\rho(t) = tv$ be the ray from $0 \in M_p$ through v, and assume that exp is defined along ρ, it is easy to verify that $\mathrm{dexp}(\rho'(t)) = \gamma_v'(t)$ and that $||\rho'(t)|| = ||\gamma_v'(t)||$. Moreover, we have the following important result known as the Gauss Lemma.

Gauss Lemma 1.8. *If $\rho(t) = tv$ is a ray through the origin of M_p and $w \in (M_p)_{\rho(t)}$ is perpendicular to $\rho'(t)$, then $\mathrm{dexp}(w)$ is perpendicular to $\mathrm{dexp}(\rho'(t))$.*

Proof. Let $c(s)$ be a curve in M_p such that $c(0) = v$, $c'(0) = w$ and such that every point of c is at the same distance from the origin of M_p. Let $\alpha(t, s)$ be a rectangle in M defined by

$$\alpha(t, s) = \exp(\rho_s(t)),$$

where $\rho_s : [0, 1] \to M_p$ is the ray from 0 to $c(s)$ in M_p. From the definition of exp and α we know that the lengths of the curves $t \to \alpha(t, s)$ are independent of s. Also, $t \to \alpha(t, 0)$ is a geodesic, and

$$\mathrm{d}\alpha \frac{\partial}{\partial s}(0, 0) = 0, \qquad \mathrm{d}\alpha \frac{\partial}{\partial s}(0, 1) = \mathrm{dexp}(w).$$

The first variation formula gives

$$0 = \frac{\mathrm{d}}{\mathrm{d}s} L[\exp(\rho_s)]\Big|_{s=0} = \langle \mathrm{dexp}(w), \gamma_v' \rangle|_0^t = \langle \mathrm{dexp}(w), \gamma_v'(t) \rangle.$$

The lemma follows. $\square$

The Gauss Lemma is equivalent to the fact that on a normal coordinate ball with the origin deleted the *gradient* of the function $r = (\Sigma x_i^2)^{\frac{1}{2}}$ is $\frac{\partial}{\partial r}$. Recall that the gradient of a function f is the unique vector field defined by

$$\langle \mathrm{grad} f, x \rangle = \mathrm{d}f(x) = x(f).$$

In fact, using polar coordinates $r, \theta_1, \ldots, \theta_{n-1}$ (which are defined like normal coordinates above) the Gauss Lemma implies that $\langle \frac{\partial}{\partial \theta_i}, \frac{\partial}{\partial r} \rangle = 0$. Then if

$$x = h\frac{\partial}{\partial r} + \sum g_i \frac{\partial}{\partial \theta_i},$$

we have

$$\Big\langle \frac{\partial}{\partial r}, x \Big\rangle = h = h\frac{\partial r}{\partial r} + \sum g_i \frac{\partial r}{\partial \theta_i} = x(r).$$

So far we have not shown, even for sufficiently small ϵ, that the length of a normal geodesic segment $\gamma : [0, \epsilon] \to M$ equals the distance between its endpoints. The following corollary of the Gauss Lemma shows that this is indeed the case. In fact we show directly that γ has shorter length than any other curve between its endpoints. This suffices because distance is defined as the infimum of lengths of such curves.

COROLLARY 1.9. *Let $B_r(0) \subseteq M_p$ be a ball of radius r on which $\exp_p$ is an embedding. Then:*

(1) For $v \in B_r(0)$, $\gamma_v : [0,1] \to M$ is the unique curve satisfying

$$L[\gamma] = \rho(p, \exp_p(v)) = ||v||.$$

In particular for any curve c, if $L[c] = \rho(c(0), c(1))$, then, up to reparameterization, c is a smooth geodesic.

(2) If $q \notin \exp_p(B_r(0)) = B_r(p)$, then there exists $q' \in \partial B_r(p)$, the boundary of $B_r(p)$, such that $\rho(p,q) = r + \rho(q', q)$. In particular, $\rho(p,q) \geq r$.

PROOF. (1) Let $c : [0,1] \to M$ be a piecewise smooth curve from p to $\exp_p(v)$. Assume $c(t) \in \exp_p(B_r(0))$ for $t \leq t_0$, i.e. $r(c(t)) \leq r$ for $t \leq t_0$. Since $||\frac{\partial}{\partial r}|| \equiv 1$, it follows that where $c(t)$ is smooth we have

$$||c'(t)|| \geq \Big\langle c'(t), \frac{\partial}{\partial r} \Big\rangle,$$

with equality holding if and only if $c'(t) = \lambda(t)(\frac{\partial}{\partial r})$, with $\lambda(t) \geq 0$. Then

$$\begin{aligned} L[c] = \int_0^1 ||c'||\mathrm{d}t &= \int_0^{t_0} ||c'||\mathrm{d}t + \int_{t_0}^1 ||c'||\mathrm{d}t \\ &\geq \int_0^{t_0} \Big\langle c', \frac{\partial}{\partial r} \Big\rangle \mathrm{d}t + \int_{t_0}^1 ||c'||\mathrm{d}t. \end{aligned}$$

As we have seen, grad $r = \frac{\partial}{\partial r}$. Thus the right-hand side of the above is equal to

$$\int_0^{t_0} \frac{\mathrm{d}}{\mathrm{d}t} r(c(t))\mathrm{d}t + \int_{t_0}^1 ||c'||\mathrm{d}t = r(c(t_0)) + \int_{t_0}^1 ||c'||\mathrm{d}t.$$

By the Intermediate-Value Theorem, there will be a first value t_1 for which $r(c(t_1)) = ||v||$. For this choice we get

$$L[c] = ||v|| + \int_{t_1}^1 ||c'||\mathrm{d}t.$$

Thus $L(c) = ||v||$ if and only if wherever $c(t)$ is smooth we have $c'(t) = \lambda(t)\frac{\partial}{\partial r}$ with $\lambda(t) \geq 0$ and $||c'|| \equiv 0$ for $t > t_1$. Then we may as well assume $t_1 = 1$. Moreover, up to reparameterization, each smooth segment of c is a segment of a radial geodesic. But then, since c is continuous, it follows that c is actually, up to reparameterization, a (single smooth) radial geodesic. In particular, if $L[c] = \rho(c(0), c(1))$, then we conclude that up to reparameterization $c(t)$ is a (smooth) geodesic.

(2) Let $c(t)$ be a curve from p to q. Since $q \notin B_r(p)$, there is a first value t_0 such that $c(t_0) \in \partial B_r(p)$. By the above,

$$L[c] \geq r + \rho(c(t_0), q) \geq r + \rho(\partial B_r(p), q).$$

Therefore

$$\rho(p, q) = \inf_c L[c] \geq r + \rho(\partial B_r(p), q).$$

But by triangle inequality the opposite inequality is also true. Therefore

$$\rho(p, q) = r + \rho(\partial B_r(p), q).$$

Since $\partial B_r(p)$ is compact, there exists $q' \in \partial B_r(p)$ such that $\rho(q', q) = \rho(\partial B_r(p), q)$. This suffices to complete the proof. □

4. The Hopf-Rinow Theorem

The preceding discussion suggests two natural questions:

(1) When is $\exp_p$ defined on all of M_p?

(2) When is it possible to join two arbitrary points by a geodesic whose length is equal to the distance between them?

The answers to these questions are related and given by the Hopf-Rinow Theorem.

THEOREM 1.10. *The following are equivalent:*

(a) M is a complete metric space where the distance from p to q in M is defined as the minimum length of all curves from p to q.

(b) For some $p \in M$, $\exp_p$ is defined on all of M_p.

(c) For all $p \in M$, $\exp_p$ is defined on all of M_p.

Any of these conditions imply

(d) Any two points p, q of M can be joined by a geodesic whose length is the distance from p to q.

In practice, the implication $(a) \Rightarrow (d)$ is most important. (a) is a very natural hypothesis which holds in particular whenever M is compact. On the other hand, it is necessary to know (d) in order to apply geometrical and analytical tools to the study of M. Theorems 1.31, 1.39, 1.42, as well as various theorems in later chapters, illustrate this. We shall often use the implication $(a) \Rightarrow (d)$ without explicitly mentioning Theorem 1.10.

In the sequel we shall assume that all manifolds are complete.

PROOF OF THEOREM 1.10. $(b) \Rightarrow (a)$. We shall first show that if for some p, $\exp_p$ is defined on all M_p, then any point q can be connected to p

by a geodesic whose length is the distance from p to q. We then show that this statement together with (b) implies (a). Now, given p, let $B_r(p)$ be a normal coordinate ball. By Corollary 1.9(1) we may assume that $q \notin B_r(p)$, so by Corollary 1.9(2) let $q' \in \partial B_r(p)$ be such that

$$\rho(p, q) = r + \rho(q', q).$$

Let $\gamma : [0, \infty) \to M$ be the normal geodesic such that $\gamma|[0, r]$ is the unique minimal geodesic from p to q'. The set of t such that

$$\rho(p, \gamma(t)) + \rho(\gamma(t), q) = \rho(p, q)$$

is clearly closed, so let $t_0 \in [r, \rho(p, q)]$ be the last such value. Let $B_{r_1}(\gamma(t_0))$ be a normal coordinate ball about $\gamma(t_0)$ so that there exists $q'' \in \partial B_{r_1}(\gamma(t_0))$ such that

$$\rho(\gamma(t_0), q) = \rho(\gamma(t_0), q'') + \rho(q'', q).$$

Let σ be the unique minimal geodesic from $\gamma(t_0)$ to q''. Since

$$\rho(p, q) = \rho(p, \gamma(t_0)) + \rho(\gamma(t_0), q'') + \rho(q'', q),$$

by triangle inequality we have

$$\rho(p, \gamma(t_0)) + \rho(\gamma(t_0), q'') = \rho(p, q'').$$

But

$$L[\gamma]|[0, t_0] = \rho(p, \gamma(t_0)), \qquad L[\sigma] = \rho(\gamma(t_0), q'').$$

Therefore

$$L[\gamma \cup \sigma] = \rho(p, q'').$$

The the curves γ and σ must fit together to form a smooth geodesic $\gamma \cup \sigma = \gamma|[0, t_0 + r_1]$. Then

$$\rho(p, q) = \rho(p, \gamma(t_0 + r_1)) + \rho(\gamma(t_0 + r_1), q),$$

which is a contradiction.

To finish the proof that $(b) \Rightarrow (a)$, let q_i be a Cauchy sequence, and let $\gamma_i : [0, t_i] \to M$ be a sequence of minimal normal geodesics with $\gamma(t_i) = q_i$. Clearly $\{t_i\}$ is also a Cauchy sequence with limit say t_0. By compactness of the unit sphere at p we may pass to a subsequence such that $\gamma'_{i_j}(0) \to v$. Let $\gamma : [0, \infty) \to M$ be the normal geodesic such that $\gamma'(0) = v$. Then the theory of ordinary differential equations (continuous dependence of solutions on initial data)applied to the geodesic equation gives $q_{i_j} = \gamma(t_{i_j}) \to \gamma(t_0)$. Since q_i is a Cauchy sequence, in fact $q_i \to \gamma(t_0)$, which completes the proof.

$(a) \Rightarrow (c)$ We must show that given p and $v \in M_p$ there exists a geodesic $\gamma : [0, 1] \to M$ such that $\gamma'(0) = v$. Let $[0, t_0)$ be the largest open interval for which such a γ exists. Then if $t_i \uparrow t_0$, $\gamma(t_i)$ is a Cauchy sequence with limit say q. Define $\gamma(t_0) = q$. Then $\gamma|[0, t_0]$ is continuous. Let $B_r(p)$ be a normal coordinate ball. For i sufficiently large, $\gamma(t_i) \in B_r(q)$. Let $\sigma : (-r, r) \to M$ be the unique minimal geodesic such that $\gamma(t_i) \in \sigma$ and $\sigma(0) = q$. Then $\gamma \cup \sigma$ is continuous, piecewise smooth and

$$L\big[\gamma \cup \sigma|[t_0 - r, t_0 + r]\big] = 2r.$$

Hence $\gamma \cup \sigma$ is actually smooth, and γ extends past t_0, which is a contradiction.

$(c) \Rightarrow (b)$ is trivial.

$(c) \Rightarrow (d)$. This is the same argument as was already given in the first part of the proof that $(b) \Rightarrow (a)$. □

In Chapter 5 we will take up the question of finding conditions under which the minimizing geodesic between two points is unique.

5. The curvature tensor and Jacobi fields

From the Gauss Lemma we know that at $v \in M_p$ the deviation of dexp from being an isometry is measured by the extent to which it fails to preserve the inner product on vectors in P, the subspace of $(M_p)_v$ which is perpendicular to the direction v itself. This failure is in turn measured by the *curvature tensor*.

The curvature tensor R assigns to each $p \in M$ a trilinear map of $M_p \times M_p \times M_p \to M_p$. If x, y, z are elements of M_p, we extend them to vector fields X, Y, Z and define

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

It is easy to check that $R(x, y)z$ does not depend on the extension to vector fields and that it is antisymmetric in x and y. Also, a straightforward computation shows that

$$R(x, y)z + R(y, z)x + R(z, x)y = 0.$$

This is called the Jacobi (or the first Bianchi) identity. Moreover,

$$\langle R(x, y)z, w\rangle = \langle R(z, w)x, y\rangle,$$

as is straightforward to check.

We now examine the relationship between curvature and the exponential map.

Let v and w be orthonormal vectors in M_p. Using the natural identification of the tangent space at any $x \in M_p$ with M_p itself, v and w induce(parallel) vector fields V and W on all of M_p. Consider the family ρ_s of rays in M_p defined by

$$\rho_s(t) = (V + sW)t.$$

Then $\exp_p \circ \rho_s$ is a geodesic through p with initial tangent vector $v + sw$. In order to measure the effect of dexp_p on the lengths of vectors in P, we shall compute some terms of the Taylor expansion of $||\mathrm{dexp}(tW)||^2$.

First some preliminaries. It is clear from the definition that $\mathrm{dexp}(tW)$ arises as the variation field of the 1-parameter family of geodesics $\exp_p \circ \rho_s$. Fields of this type are called *Jacobi fields* and are characterized as the solutions of a certain second-order differential equation. More precisely, let $\alpha(t, s) : [a, b] \times (-\epsilon, \epsilon) \to M$ be such that for fixed s, $\alpha(t, s)$ is a geodesic. Let $T = \mathrm{d}\alpha\left(\frac{\partial}{\partial t}\right)$ and $V = \mathrm{d}\alpha\left(\frac{\partial}{\partial s}\right)$.

We now determine the differential equation satisfied by $V|\alpha(t,0)$. By the discussion at the end of Section 1,

$$\nabla_T V - \nabla_V T - \mathrm{d}\alpha\Big[\frac{\partial}{\partial t}, \frac{\partial}{\partial s}\Big] = 0.$$

But $\left[\frac{\partial}{\partial t}, \frac{\partial}{\partial s}\right] = 0$, so $\nabla_T V = \nabla_V T$. Therefore $\nabla_T \nabla_T V = \nabla_T \nabla_V T$. Since $\nabla_T T = 0$, we may write

$$\nabla_T \nabla_T V = \nabla_T \nabla_V T = \nabla_T \nabla_V T - \nabla_V \nabla_T T.$$

Using the definition of the curvature tensor and the fact that

$$[T, V] = \mathrm{d}\alpha\Big(\Big[\frac{\partial}{\partial t}, \frac{\partial}{\partial s}\Big]\Big) = 0,$$

we get the *Jacobi equation*

$$\nabla_T \nabla_T V = R(T, V)T.$$

A vector field V, along a geodesic γ with tangent vector T satisfying this equation is called a *Jacobi field*. Let $\{E_i(t)\}$ be orthogonal and parallel along $\alpha(t,0)$. Then the Jacobi equation may be written as the linear second-order system of ordinary differential equations

$$\begin{bmatrix} \langle J, E_1\rangle \\ \vdots \\ \langle J, E_n\rangle \end{bmatrix}'' = \langle R(T, E_i)T, E_j\rangle \begin{bmatrix} \vdots \\ \langle J, E_i\rangle \\ \vdots \end{bmatrix}.$$

From the theory of ordinary differential equations it follows that the space of solutions of this system is $2n$-dimensional and that there exists a unique solution with prescribed initial value and first derivative. This is equivalent to prescribing $J(0)$ and $J'(0) = \nabla_T J|_{t=0}$. Notice that since $\nabla_T T = 0$, we have

$$\langle J, T\rangle'' = \langle J'', T\rangle = \langle R(T, J)T, T\rangle = 0.$$

Therefore any Jacobi field J may be written uniquely as

$$J = J_0 + (at + b)T,$$

where $\langle J_0, T\rangle \equiv 0$, Finally, if J is a Jacobi field, then J comes from a variation of geodesics. In fact, let $c(s)$ be a curve such that $c'(0) = J(0)$, and let T and $J'(0)$ be extended to parallel fields along $c(s)$. Then the variation field of $\exp_{c(s)}(t(T + sJ'(0)))$ is a Jacobi field with the same initial conditions as J. Therefore it equals J by the uniqueness theorem above.

We are now ready to calculate the Taylor series for $||\mathrm{dexp}(tW)||^2$. Set $\mathrm{dexp}(V) = T$ and $\mathrm{dexp}(tW) = J$. Then it is easily seen that $J'(0) = w$. Then

$$\begin{aligned}
&\langle J, J\rangle|_{t=0} = 0, \\
&\langle J, J\rangle'|_{t=0} = 2\langle J, J'\rangle|_{t=0} = 0, \\
&\langle J, J\rangle''|_{t=0} = 2\langle J', J'\rangle|_{t=0} + 2\langle J'', J\rangle|_{t=0} = 2||W||^2 = 2.
\end{aligned}$$

Note that in fact $J''|_{t=0} = R(T,J)T|_{t=0} = 0$, so

$$\langle J, J\rangle''' = 6\langle J'', J'\rangle|_{t=0} + 2\langle J''', J\rangle|_{t=0} = 0.$$

Also,

$$J''' = \nabla_T(R(T,J)T)|_{t=0} = (\nabla_T R)(T,J)T|_{t=0} + R(T,J')T|_{t=0}.$$

So

$$J'''|_{t=0} = R(T,J')T|_{t=0} = R(T,w)T|_{t=0}.$$

Then

$$\begin{aligned}\langle J, J\rangle'''' &= 8\langle J''', J'\rangle|_{t=0} + 6\langle J'', J''\rangle|_{t=0} + 2\langle J'''', J\rangle|_{t=0}\\ &= 8\langle R(T,w)T, w\rangle = -8\langle R(w,T)T, w\rangle.\end{aligned}$$

Therefore

$$||\mathrm{d}\exp(tW)||^2 = t^2 - \frac{1}{3}\langle R(w,T)T, w\rangle t^4 + O(t^5). \tag{1.11}$$

We find that when compared to the rays ρ_s geodesics $\exp\circ\rho_s$ come together to the order of $(\frac{1}{3}\langle w, R(w,v)v\rangle)^{\frac{1}{2}}t^2$. So if $\langle R(w,v)v, w\rangle$ is positive, geodesics locally converge, and if it is negative, they locally diverge by comparison with rays.

(1.11) may also be given the following interpretation. Let $g_{ij}(x)$ denote the metric expressed in terms of normal coordinates. Then at the origin, up to first order the metric looks like the Euclidean metric $g_{ij} = \delta_{ij}$. The deviation comes in with the second-order terms, which are in turn measured by curvature.

Given any plane σ in M_p and two vectors v and w which span σ, we define the *sectional curvature* $K(\sigma)$ to be

$$\frac{\langle R(v,w)w, v\rangle}{||v\wedge w||^2}.$$

Here $||v\wedge w||^2$ denotes the square of the area of the parallelogram spanned by v and w. One can easily check that $K(\sigma)$ does not depend on the choice of the spanning vectors. Furthermore, the curvature tensor R is completely determined by the inner product together with the function $K : G_{2,n}(M_m) \to \mathbb{R}$, where $G_{2,n}(M_m)$ denotes the space of all 2-dimensional subspaces of M_m, the Grassmann manifold. In fact, a straightforward computation shows that

(1.12)

$$\begin{aligned}\langle R(x,y)z,w\rangle = \frac{1}{6}\{&K(x+w,y+z)||(x+w)\wedge(y+z)||^2\\&-K(y+w,x+z)||(y+w)\wedge(x+z)||^2\\&-K(x,y+z)||x\wedge(y+z)||^2-K(y,x+w)||y\wedge(x+w)||^2\\&-K(z,x+w)||z\wedge(x+w)||^2-K(w,y+z)||w\wedge(y+z)||^2\\&+K(x,y+w)||x\wedge(y+w)||^2+K(y,z+w)||y\wedge(z+w)||^2\\&+K(z,y+w)||z\wedge(y+w)||^2+K(w,x+z)||w\wedge(x+z)||^2\\&+K(x,z)||x\wedge z||^2+K(y,w)||y\wedge w||^2\\&-K(x,y)||x\wedge w||^2-K(y,z)||y\wedge z||^2.\end{aligned}$$

Here $K(x,y)$ denotes the curvature of the plane spanned by x, y. If the curvatures of all plane sections are of the same sign, then this sign is a fundamental invariant. By studying in more detail its effect on the behavior of geodesics, we will derive topological and geometrical information. We will use the notation $K_M > H$ to indicate that for all plane sections at all points of M the sectional curvature is bigger than the constant H.

The condition $K \equiv 0$ is equivalent to the statement that in normal coordinates $g_{ij} \equiv \delta_{ij}$, as will be clear from the results of the next section and Section 14.

6. Conjugate points

As was shown in Section 2, in order for a curve γ to realize the distance between its endpoints, it is necessary that γ be a geodesic. However, if γ is too long, this condition is not sufficient. For example, on the unit sphere, geodesics are great circles. A geodesic of the length of more than π will not minimize. Suppose, for example, that γ starts at p. There are infinitely many geodesics σ having length π and going from p to the antipodal point q.

The path $\sigma \cup \tau$ shown in Fig. 1.2, consists of the segment of σ from p to r, and the minimal geodesic τ from r to s will have length shorter than the segment of γ from p to s. Now in this example q is a singular value of $\exp_p$. Although the above argument just used the fact that there were *two* distinct geodesics from p to q, the example suggests that there should also be a connection between failure of geodesics to minimize globally and singular values of exp. Intuitively, these are points at which distinct geodesics come together at least infinitesimally. To demonstrate this connection, we shall need an infinitesimal version of the argument. We begin with a characterization of the singularities of exp. We say that q is *conjugate* to p if q is a singular value of $\exp : M_p \to M$. The conjugacy is said to be along $\gamma = \gamma_v$ if dexp is singular at v. The *order* of a conjugate point is defined to be the

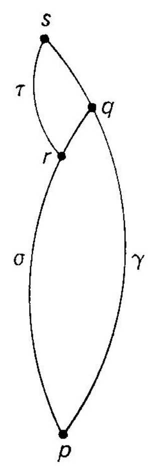

Fig. 1.2.

dimension of the null space of dexp : $(M_p)_v \to M_q$. Let us pick $v \in M_p$ and $w \in (M_p)_v$, and assume $\text{dexp}(w) = 0$. We know immediately that w (when identified with its corresponding vector in M_p) is perpendicular to v, because the length of any component of w in the v-direction is preserved by dexp. Also, if as in the previous section we form the rays $\rho_s(t) = (v + sw)t$ and geodesics $\gamma_s = \exp \circ \rho_s$, we find

$$\frac{\mathrm{d}}{\mathrm{d}s}(\exp \circ \rho_s(1)) = 0.$$

Hence there exists a one-parameter family of geodesics γ_s which come from p and whose associated Jacobi field, the variation vector field $\frac{\mathrm{d}}{\mathrm{d}s}(\gamma_s(t))$ vanishes at $q = \gamma_0(1)$. Thus we have proved half of the following proposition.

Proposition 1.13. *q is conjugate to p along a geodesic γ if and only if there exists a non-zero Jacobi field J along γ such that J(0) = J(1) = 0. Hence q is conjugate to p if and only if p is conjugate to q.*

Proof. Suppose there exists a nonzero J along γ with $J(0) = J(1) = 0$. $\alpha(t, s) = \exp_{\gamma(0)}(T + sJ'(0))t$ is a rectangle whose associated variation field is J. Then

$$\text{dexp} J'(0)|_{\gamma'(0) \in M_{\gamma(0)}} = J(1) = 0,$$

so p is conjugate to q. Since the Jacobi-field condition is symmetric in p and q, so is conjugacy. □

Now we give two elementary facts about Jacobi fields and conjugate points.

Proposition 1.14. *Let $\gamma : [a, b] \to M$ be a geodesic, $\gamma' = T$, and assume there is a Jacobi field J which vanishes at $\gamma(a)$ and $\gamma(b)$. Then $\langle J, T\rangle \equiv \langle J', T\rangle \equiv 0$.*

PROOF. We note that

$$T\langle T, J'\rangle = \langle T, J''\rangle = -\langle T, R(J,T)T\rangle = -\langle R(T,T)J, T\rangle = 0.$$

Therefore, $\langle T, J'\rangle$ is constant on γ. But $\langle T, J'\rangle = T\langle T, J\rangle$, so that $\langle T, J\rangle$ as a function of the parameter of γ has constant derivative. But

$$\langle T, J\rangle_{\gamma(a)} = \langle T, J\rangle_{\gamma(b)} = 0,$$

so $\langle T, J\rangle \equiv 0$ and $\langle T, J'\rangle \equiv 0$. □

REMARK 1.15. This argument actually shows that

$$\langle T, J\rangle = \langle T, J(0)\rangle + \langle T, J'(0)\rangle t.$$

PROPOSITION 1.16. *If $\gamma(a)$ and $\gamma(b)$ are not conjugate, then a Jacobi field J along γ is determined by its values at $\gamma(a)$ and $\gamma(b)$.*

PROOF. Let W and J be two Jacobi fields which coincide at the endpoints of γ. Then $W - J$ is a Jacobi field which vanishes at both endpoints. But since these points are not conjugate, $W - J$ must be identically zero, so $W = J$ □

Suppose that M has constant curvature K. Then the formula of the previous section specializes to

$$\langle R(x,y)z, w\rangle = -K(\langle x, z\rangle\langle y, w\rangle - \langle y, z\rangle\langle x, w\rangle).$$

Then the Jacobi equation expressed in terms of parallel fields becomes

$$\langle J'', E_i\rangle = \langle \nabla_T\nabla_T J, E_i\rangle = -\langle R(J,T)T, E_i\rangle = K(\langle J, T\rangle\langle T, E_i\rangle - \langle J, E_i\rangle).$$

So if $\langle J, T\rangle = 0$, we get simply

$$\langle J, E_i\rangle'' = -K\langle J, E_i\rangle.$$

The reader may verify that the general solution is

$$K > 0 : \sum(a_i \sin(\sqrt{K}t) + b_i \cos(\sqrt{K}t))E_i(t);$$
$$K = 0 : \sum(a_i t + b_i)E_i(t);$$
$$K < 0 : \sum(a_i \sinh(\sqrt{-K}t) + b_i \cosh(\sqrt{-K}t))E_i(t).$$

If $K < 0$ or $K = 0$, geodesics have no conjugate points, while if $K > 0$ conjugate points occur at $t = \frac{\pi l}{\sqrt{K}}$, where l is an integer.

7. Second variation of arc length

We have shown in Section 2 that geodesics are always critical points of the arc-length function. However, as we have observed in Section 6, they are not always local minima. Therefore we compute the second derivative of arc length with respect to a variation.

Let $\gamma : [a, b] \to M$ be a geodesic, and let $\alpha : Q \to M$ be a smooth map, where Q is the rectangular solid $[a, b] \times [-\epsilon, \epsilon] \times [-\delta, \delta]$ and $\alpha(t, 0, 0) = \gamma(t)$. This means that α is a 2-parameter variation of the geodesic γ. We will look at the arc-length function differentiated successively with respect to these two parameters. Let $L(v,w)$ be the arc length of the curve $t \to \alpha(t,v,w)$, so

$$L(v,w) = \int_a^b ||T||\mathrm{d}s.$$

Assume $||\gamma'|| \equiv 1$. Let T, V, W be vector fields corresponding to the first, second and third variables of α, respectively.

As in the proof of the first variation formula

$$\frac{\partial}{\partial v} L(v,w) = \int_a^b \frac{\langle \nabla_T V, T\rangle}{||T||}.$$

Then

$$\begin{aligned}
\frac{\partial^2}{\partial w \partial v} L(v,w) &= \frac{\partial}{\partial w}\int_a^b \frac{\langle \nabla_T V, T\rangle}{||T||} \\
&= \int_a^b \frac{\langle \nabla_W \nabla_T V, T\rangle + \langle \nabla_T V, \nabla_W T\rangle}{||T||} - \langle \nabla_T V, T\rangle \frac{\langle \nabla_W T, T\rangle}{||T||^3} \\
&= \int_a^b \frac{\langle R(W,T)V, T\rangle + \langle \nabla_T \nabla_W V, T\rangle + \langle \nabla_T V, \nabla_T W\rangle}{||T||} \\
&\quad - \frac{\langle \nabla_T V, T\rangle \langle \nabla_T W, T\rangle}{||T||^3}.
\end{aligned}$$

Using $||T|||_{(0,0)} \equiv 1$ and $\nabla_T T|_{(0,0)} \equiv 0$,

$$\begin{aligned}
\frac{\partial^2 L}{\partial w \partial v}\Big|_{(0,0)} &= \int_a^b \langle \nabla_T V, \nabla_T W\rangle - \langle R(W,T)T, V\rangle \\
&\quad - T\langle \nabla_W V, T\rangle - T\langle V, T\rangle T\langle W, T\rangle \\
&= -\langle \nabla_W V, T\rangle|_a^b + \int_a^b \langle \nabla_T V, \nabla_T W\rangle \\
&\quad - \int_a^b \langle R(W,T)T, V\rangle - T\langle V, T\rangle T\langle W, T\rangle.
\end{aligned} \tag{1.17}$$

This is the *second variation formula.* Note that (1.17) is valid for 2-parameter piecewise smooth variations (with the obvious definition). This follows by restricting attention to the subintervals on which V, W are smooth, then adding and observing that the endpoint terms cancel. In case the variation is through geodesics, by Proposition 1.14, $T\langle V, T\rangle$ and $T\langle W, T\rangle$ are constant. Then if $\langle V, T\rangle$ or $\langle W, T\rangle$ vanishes at both endpoints, the last term drops out. Moreover, if either V or W vanishes at the endpoints or, more generally, $\nabla_V W = 0$, we get

$$\frac{\partial^2 L}{\partial w \partial v}\Big|_{(0,0)} = \int_a^b \langle \nabla_T V, \nabla_T W\rangle + \langle R(W,T)V, T\rangle. \tag{1.18}$$

We remark that in this case the second variation depends only on the restrictions of V, W to γ. We call the above integral the *index form* $I(V,W)$.

It is a symmetric bilinear form on the space of piecewise smooth vector fields V, W along γ such that $\langle V, T\rangle \equiv \langle W, T\rangle \equiv 0$. Notice that I is independent of the orientation of γ. If I is positive definite on vector fields vanishing at $\gamma(a), \gamma(b)$, then γ is a minimum among all nearby curves with the same endpoints.

If $a = t_0 < t_1 < \cdots < t_n = b$ is such that $V, W|[t_i, t_{i+1}]$ is smooth, then using

$$\langle \nabla_T V, \nabla_T W\rangle = T\langle \nabla_T V, W\rangle - \langle \nabla_T \nabla_T V, W\rangle$$

and integrating, we get

$$I(V, W) = \sum_i \langle \Delta_{t_i}(\nabla_T V), W\rangle - \int_a^b \langle \nabla_T \nabla_T V, W\rangle + \langle R(V, T)T, W\rangle, \tag{1.19}$$

where

$$\Delta_{t_i}(\nabla_T V) = \lim_{t \to t_i^+} \nabla_T V - \lim_{t \to t_i^-} \nabla_T V.$$

In particular, if $V|[t_i, t_{i+1}]$ is a Jacobi field, then

$$I(V, W) = \sum_i \langle \Delta_{t_i}(\nabla_T V), W\rangle.$$

Proposition 1.20. *Let I be defined on all piecewise smooth vector fields along γ which vanish at the endpoints. Then the null space of I is exactly the set of Jacobi fields along γ which vanish at $\gamma(a)$ and $\gamma(b)$. Specifically, V is a Jacobi field if and only if $I(V, W) = 0$ for all W.*

Proof. By the above it suffices to show that $I(V, W) = 0$ for all W implies that V is a Jacobi field. Let f be a function vanishing at $\{t_i\}$ and positive elsewhere. Then setting

$$W = f(t)(-\nabla_T \nabla_T V + R(T, V)T)$$

we see that $V|[t_i, t_{i+1}]$ is a Jacobi field for all i. Now letting W_0 be a field such that $W_0(t_i) = \Delta_{t_i} \nabla_T V$, the claim follows. □

Corollary 1.21. *I has a non-trivial null space if and only if $\gamma(a)$ is conjugate to $\gamma(b)$ along γ. The dimension of the null space is the order of the conjugate point $\gamma(b)$.*

Proof. The first statement is merely Proposition 1.13. The second follows from the construction used in the proof of that proposition. It provides a linear isomorphism between the null space of dexp and the space of Jacobi fields along γ which vanish at the endpoints. □

8. Submanifolds and the second fundamental form

We leave our study of the second variation of arc length in order to present some facts about connections on submanifolds which we shall need later. We return to second variation in Section 9.

Let N be a submanifold of M. Then for each $p \in N$, let $P : M_p \to N_p$ be the orthogonal projection with respect to $\langle\ ,\ \rangle$. The normal bundle $\nu(M)$

(or simply ν) is the subset of $T(M)$ defined by: $x \in \nu$ if $x \in M_p$ for $p \in N$, and $P(x) = 0$. ν is a vector bundle over N whose dimension is the difference of the dimensions of M and N. Its fiber at $p \in N$ will be denoted ν_p.

The Riemannian metric of M also induces a positive-definite bilinear form on each tangent space N_p. So N inherits a Riemannian metric, and hence an affine connection, which we shall call ∇^0. In fact ∇^0 can be defined as follows: Given X and Y vector fields on N, extend them to M and set $\nabla^0_X Y$ equal to $P(\nabla_X Y)$. It is easy to check that $\nabla^0$is the unique torsion-free connection on N which is compatible with the metric.

The *second fundamental form* S of N is the difference between ∇ and ∇^0. Specifically if x and y are in N_p, we extend them to vector fields and define $S(x, y)$ to be $\nabla_X Y - \nabla^0_X Y$. One checks that the definition is independent of the extension and that $S : N_p \times N_p \to \nu_p$ is a bilinear map. It is also symmetric, because if we extend x and y so that $[X, Y](p) \in N_p$, then

$$\nabla_X Y - \nabla_Y X = [X, Y] = \nabla^0_X Y - \nabla^0_Y X.$$

Given $z \in \nu_p$, we define $S_z : N_p \times N_p \to \mathbb{R}$ by

$$S_z(x, y) = \langle S(x, y), z\rangle.$$

Of course, S_z is a symmetric bilinear form on N_p.

Generalizing the notion of conjugate point, we say that a *focal point* q of N is a singular value of $\exp|_\nu$. We call q a focal point of N at p if there is a singular inverse image of q somewhere in ν_p.

There is a particular submanifold of codimension one which will prove useful in what follows. Fix $p \in M$ and x a unit vector in M_p. Let

$$x^\perp = \{y \in M_p | \langle x, y\rangle = 0\}.$$

Since $\exp : M_p \to M$ is a local diffeomorphism at zero, there is a neighborhood U of zero in $x^\perp$ such that $\exp|_U$ is an embedding. Let N be the submanifold $\exp(U)$. We call N the *geodesic submanifold* defined by x.

Lemma 1.22. *$S_x = 0$, where S is the second fundamental form of N.*

Proof. Pick $z \in N_p$. Let γ be a geodesic from p in direction z. Then near p, $\gamma \subseteq N$, so we can extend z to a vector field Z on N such that Z is the tangent vector to γ near p. Then $(\nabla_Z Z)(p) = 0$, so $S_x(z, z) = 0$. Therefore, since S_x is symmetric, $S_x = 0$. □

Corollary 1.23. *Let x be extended to X, the unit normal field to N, $\langle X, X\rangle \equiv 1$, $\langle X, Y\rangle \equiv 0$ for all $Y \in T(N)$. Then $\nabla_z X = 0$ for $z \in N_p$.*

Proof. If Y is any vector field on N, then

$$\langle \nabla_Z Y, X\rangle_p = Z\langle Y, X\rangle_p - \langle Y, \nabla_Z X\rangle_p.$$

But $S_x = 0$ implies $\langle \nabla_Z Y, X\rangle_p = 0$, and $\langle Y, X\rangle$ is zero on N, so $z\langle Y, Z\rangle = 0$. Therefore $\langle y, \nabla_Z X\rangle_p = 0$ for all $y \in N_p$. But since X has constant length,

$$\langle \nabla_Z X, X\rangle = \frac{1}{2} Z\langle X, X\rangle = 0;$$

hence $\nabla_Z X = 0$. □

If S_X vanishes identically, then N is called *totally geodesic*. Let γ be a geodesic of such an N. The equation

$$\nabla_{\gamma'}\gamma' = \nabla^0_{\gamma'}\gamma' + S(\gamma', \gamma') = 0$$

shows that γ is a geodesic of M. By uniqueness of geodesics, this is equivalent to the statement that if γ is a geodesic of M tangent to N at $\gamma(0)$, then γ remains in N. Conversely, if N has the latter property, then as in Lemma 1.22, $S \equiv 0$.

9. Basic index lemmas

In this section we prove two lemmas which show that in a certain sense, Jacobi fields minimize the index form. The first lemma involves conjugate points, while the second is an analogous statement with conjugate points replaced by focal points of the submanifold N described above. These lemmas are of fundamental importance and are used in the proofs of Rauch Comparison Theorems and the Morse Index Theorem in Chapter 4.

LEMMA 1.24 (First Lemma). *Let γ be a geodesic in M from p to q such that there are no points conjugate to p on γ. Let W be a piecewise smooth vector field on γ and V the unique Jacobi field such that $V(p) = W(p) = 0$ and $V(q) = W(q)$. Then $I(V, V) \leq I(W, W)$, and equality holds only if $V{=}W$.*

LEMMA 1.25 (Second Lemma). *Let γ be a geodesic in M from p to q. Let x be the tangent vector to γ at p and let N be the geodesic submanifold defined by x. Assume that N has no focal points along γ. Let W be a piecewise smooth vector field along γ, V the unique Jacobi field such that $(\nabla_X V)(p) = 0$ and $V(q) = W(q)$. Then $I(V, V) \leq I(W, W)$, and equality holds only if $V = W$.*

PROOF OF FIRST LEMMA. Let $\{V_i\}$ be a basis of T_qM and extend each V_i to a Jacobi field along γ such that $V_i(p) = 0$. This is uniquely possible since γ has no conjugate points, and the V_i's are linearly independent except at p. Since $V_i(p) = 0$, we can write $V_i = tA_i$, where t is the parameter of $\gamma : [0, 1] \to M$ and A_i is some vector field on γ. Then $V_i'(p) = A_i$, so $\{A_i\}$ is also linearly independent, and thus there are functions $q_i(t)$ such that $W = \sum\limits_i q_i(t)A_i$. But since $W_p = 0$, there exist piecewise smooth functions f_i so that $W = \sum f_iV_i$. Then $V = \sum f_i(1)V_i$.

We shall make two preliminary calculations:

$$I(V, V) = \langle V'(1), V(1)\rangle = \sum f_i(1)f_j(1)\langle V_i'(1), V_j(1)\rangle. \tag{1.26}$$

In fact, since V is a Jacobi field, $V'' = R(T,V)T$, so

$$\begin{aligned} I(V,V) &= \int_0^1 \langle V', V'\rangle + \langle R(T,V)T, V\rangle \\ &= \int_0^1 \langle V', V\rangle' - \langle V'', V\rangle + \langle R(T,V)T, V\rangle = \langle V'(1), V(1)\rangle. \end{aligned}$$

If V_i, V_j are Jacobi fields, then

$$\langle V_i', V_j\rangle - \langle V_i, V_j'\rangle = c, \tag{1.27}$$

for some constant c. We compute

$$\begin{aligned} (\langle V_i', V_j\rangle - \langle V_i, V_j'\rangle)' &= \langle V_i'', V_j\rangle + \langle V_i', V_j'\rangle - \langle V_i', V_j'\rangle - \langle V_i, V_j''\rangle \\ &= \langle V_i'', V_j\rangle - \langle V_i, V_j''\rangle \\ &= \langle R(T,V_i)T, V_j\rangle - \langle V_i R(T,V_j)T\rangle = 0. \end{aligned}$$

The last step follows by the usual symmetry property of the curvature tensor. In our case the constant in (1.9.2) is zero since the expression vanishes at $t = 0$. Now

$$\begin{aligned} \nabla_T W &= \sum f_i' V_i + f_i V_i' = A + B, \\ I(W,W) &= \int \langle A, A\rangle + \langle A, B\rangle + \langle B, A\rangle + \langle B, B\rangle + \langle R(T,W)T, W\rangle, \\ \int \langle B, B\rangle &= \sum \int f_i f_j \langle V_i' V_j'\rangle = \sum \int f_i f_j (\langle V_i', V_j\rangle' - \langle V_i'', V_j\rangle). \end{aligned}$$

Integrating the first term by parts and applying the Jacobi equation to the second gives

$$\begin{aligned} \int \langle B, B\rangle = \sum & f_i(1) f_j(1) \langle V_i'(1), V_j(1)\rangle \\ & - \int (f_i' f_j \langle V_i', V_j\rangle + f_i f_j' \langle V_i', V_j\rangle + \langle R(T,W)T, W\rangle). \end{aligned}$$

By (1.26), the first term is $I(V,V)$. By (1.27), the second term is $\int \langle A, B\rangle$, while the third term equals $\int \langle B, A\rangle$. Hence

$$\int \langle B, B\rangle = I(V,V) - \int \langle A, B\rangle + \langle B, A\rangle + \langle R(T,W)T, W\rangle.$$

Therefore, referring to the original expression for $I(W,W)$,

$$I(W,W) = I(V,V) + \int \langle A, A\rangle.$$

The second term on the right is nonnegative and vanishes only if $W = V$. □

We emphasize that the above calculation, in particular the integration by parts, depends only on W being piecewise smooth.

To prove the second lemma we pick $\{v_i\}$ an orthonormal basis of T_pM such that $v_1 = T$, and extend its members to Jacobi fields $\{V_i\}$ such that $(\nabla_T V_i)_p = 0$. We need the following:

SUBLEMMA 1.28. *If $\{V_i\}$ are as above and N, γ are as in the second lemma, then $\{V_i\}$ are linearly independent along γ if and only if N has no focal point.*

PROOF. $V_1 = T$ all along γ and $V_2, \dots, V_n$ are everywhere perpendicular to T. Hence it suffices to show that $\{V_2, \dots, V_n\}$ are independent or to show that if Z is any nonzero Jacobi field on γ such that $\langle Z, T\rangle_p = 0$ and $(\nabla_T Z)_p = 0$, then Z is never zero.

Let η be the geodesic from p in direction Z, and extend T along η so that $||T||$ is constant and T is perpendicular to N. Construct a geodesic from each point of η in direction T. This one-parameter family of geodesics gives rise to a Jacobi field Z_0 on γ. Clearly $Z_0(p) = Z(p)$. Also $(\nabla_{Z_0} T)(p) = 0$ by Corollary 1.23, so $(\nabla_T Z_0)(p) = 0$ and $Z_0 = Z$. From the construction of Z_0 it is immediate that Z_0 has a zero implies exp ν has a singularity over γ. Also, if there is such a singularity, one constructs Z_0 in a straightforward manner. This proves the sublemma. □

Now the proof of the second lemma proceeds exactly as that of the first. $\langle \nabla_T V_i, V_i\rangle_p$, $\langle B, W\rangle_p$ and $\langle \nabla_T V, V\rangle_p$ are all zero because $\nabla_T V_i$ and $\nabla_T V$ are zero at p. □

We can now state that geodesics minimize locally up to the first conjugate point among curves with the same end points. More precisely, if $\gamma(t_0)$ is the first conjugate point of $\gamma(0)$ along γ, then for vector fields vanishing at $\gamma(0), \gamma(t)$ with $t < t_0$, the second variation is positive; that is, the index form is positive definite. This follows by taking $W(q) = 0$ in Lemma (1.24). Moreover, we have the following important converse.

COROLLARY 1.29. *Let $\gamma : [0, \infty) \to M$ be a geodesic, and let $\gamma(t_0)$ be conjugate to $\gamma(0)$. Then $\gamma|[0, t]$ is not minimal for $t > t_0$.*

PROOF. We can assume that $\gamma(t_0)$ is the first point conjugate to $\gamma(0)$. Let J be a nonzero Jacobi field along $\gamma[0, t_0]$ such that $J(0) = J(t_0) = 0$. Extend J to a vector field X on all of γ by declaring $X(t) = 0$ for $t > t_0$. Clearly $I(X, X) = 0$ on $[0, t]$, but X is not smooth at t_0.

Fix δ small enough so that there are no conjugate pairs on $\gamma|[t_0 - \delta, t_0 + \delta]$, and define a vector field V by:

$V = J$ on $[0, t_0 - \delta]$,

$V =$ Jacobi field W on $[t_0 - \delta, t_0 + \delta]$ such that

$$W(t_0 - \delta) = J(t_0 - \delta), \qquad W(t_0 + \delta) = 0,$$

$V = 0$ on $[t_0 + \delta, t]$ (See Fig. 1.3)

Since X is not smooth on $[t_0 - \delta, t_0 + \delta]$, it is definitely not a Jacobi field. Hence on $[t_0 - \delta, t_0 + \delta]$ we have $I(V, V) < I(X, X)$. Since $X = V$ outside this interval, in fact on $[0, t]$ we have $I(V, V) < I(X, X) = 0$. Since any V arises from a variation, it follows that there is a variation which keeps the endpoints fixed and decreases the length of γ.

□

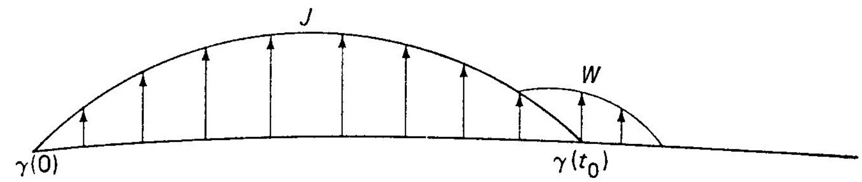

Fig. 1.3.

10. Ricci curvature and Myers' and Bonnet's Theorems

As an application of the First Index Lemma (1.22), we now prove the theorems of Myers and Bonnet. We define the *diameter* $d(M)$ of M to be the supremum of $\rho(p,q)$ for $p,q \in M$.

DEFINITION 1.30. The *Ricci curvature* is a symmetric bilinear form on each M_p of $T(M)$ defined to be the trace of the linear tranformation $z \to R(z,x)y$. Hence

$$\mathrm{Ric}(x,y) = \sum_i \langle R(e_i,x)y, e_i\rangle,$$

where $\{e_i\}$ is an orthonormal basis of M_p.

THEOREM 1.31 (Myers and Bonnet). *Let M^n be a complete Riemannian manifold. If*

(1) (Myers) *for all unit vectors x,* $\mathrm{Ric}(x,x) \geq (n-1)H$, *or*

(2) (Bonnet) $K_M \geq H$,

then every geodesic of length $\geq \pi/\sqrt{H}$ has conjugate points. Hence the diameter of M satisfies $d(M) \leq \pi/\sqrt{H}$.

PROOF. Fix a normal geodesic $\gamma : [0,l] \to M$, and let $\{E_i\}$ be an orthonormal basis of parallel fields along γ such that $E_n = \gamma' = T$. Let $W_i = \sin(\frac{\pi t}{l})E_i(t)$ be vector fields on γ. Then

$$\begin{aligned} I(W_i,W_i) &= -\int_0^l \langle W_i, \nabla_T^2 W_i + R(W_i,T)T\rangle \mathrm{d}t \\ &= \int_0^l (\sin(\frac{\pi t}{l}))^2(\frac{\pi^2}{l^2} - \langle R(E_i,T)T,E_i\rangle)\mathrm{d}t. \end{aligned}$$

Thus if, for any i, $\langle R(E_i,T)T,E_i\rangle \geq H$ and $l \geq \pi/\sqrt{H}$, then $I(W_i,W_i) \leq 0$. Also,

$$\sum_{i=1}^{n-1} I(W_i,W_i) = \int_0^l (\sin(\frac{\pi t}{l}))^2((n-1)\frac{\pi^2}{l^2} - \mathrm{Ric}(T,T))\mathrm{d}t.$$

So if $\mathrm{Ric}(T,T) \geq (n-1)H$ and $l \geq \frac{\pi}{\sqrt{H}}$, then the sum and therefore at least one summand must be nonpositive. But if γ had no conjugate points, the First Index Lemma 1.24 would imply that there is a Jacobi field J such that $I(J,J) < 0$ and J would vanish at $\gamma(0)$ and $\gamma(l)$. This is impossible, so γ

has conjugate points. By Corollary 1.29, γ is not minimal. More directly, since some $I(W_i, W_i)$ is negative, the second variation of arc length is in some direction negative, so γ cannot be minimal. The theorem follows. □

The following is our first result which illustrates the influence of the sign of the curvature on the topology of M.

COROLLARY 1.32. *Let M be complete. If there exists a constant $H > 0$ such that for all unit vectors x we have* $\mathrm{Ric}(x, x) > (n-1)H > 0$*, then M is compact and has finite fundamental group.*

To prove this corollary, we use the notion of local isometry.

Let M and N be Riemannian manifolds with metrics $\langle\langle\ ,\ \rangle\rangle$ and $\langle\ ,\ \rangle$, and let $\phi : M \to N$ be a smooth map. We say ϕ is a *local isometry* if for all $p \in M$ and $v, w \in M_p$,

$$\langle\langle v, w\rangle\rangle = \langle \mathrm{d}\phi(v), \mathrm{d}\phi(w)\rangle.$$

ϕ is an *isometry* if ϕ is local isometry and also a diffeomorphism.

This notion of isometry is equivalent to the usual notion of isometry of M and N as metric spaces. The implication in one direction is easy, but it is not so obvious that every metric space isometry is smooth. This is an essential step in the Theorem of Myers and Steenrod quoted in Chapter 3. It is easy to check that a local isometry respects Riemannian connections and maps geodesics to geodesics.

Note also that if ϕ is a local isometry, then $\mathrm{d}\phi$ must be everywhere nonsingular; for if $\mathrm{d}\phi(v) = 0$, then

$$\langle v, v\rangle = \langle \mathrm{d}\phi(v), \mathrm{d}\phi(v)\rangle = 0,$$

so v is the zero vector. Conversely, if $d\phi$ is everywhere nonsingular, then we can use the Riemannian metric of N to induce one on M. For $v, w \in M_p$, we define $\langle\langle v, w\rangle\rangle$ to be $\langle \mathrm{d}\phi(v), \mathrm{d}\phi(w)\rangle$. It is easy to check that $\langle\langle\ ,\ \rangle\rangle$ is a Riemannian metric and that ϕ is a local isometry with respect to the metrics $\langle\langle\ ,\ \rangle\rangle$ on M and $\langle\ ,\ \rangle$ on N.

PROOF OF COROLLARY. Let M_c be the universal covering space of M. Since $\pi : M_c \to M$ is a local diffeomorphism, it induces a Riemannian structure on M_c, and the curvature tensor R_c at $p_c \in M_c$ is isomorphic to R at $\pi(p_c) \in M$. Therefore by Myers' theorem, $d(M) \leq \pi H^{-\frac{1}{2}}$, so M_c is compact. Hence the set $\pi^{-1}(p)$ must have finite cardinality and the corollary follows. □

11. Rauch Comparison Theorem

We will often study a Riemannian manifold when our only given infirmation consists of bounds on its sectional curvature. The following theorems allow us to draw geometrical conclusions from such information by comparing lengths in M to corresponding lengths in a manifold M_0 whose curvature is suitably related to that of M. We usually take M_0 to have constant curvature, in which case the lengths in M_0 can be calculated explicitly.

THEOREM 1.33 (First Theorem (Rauch)). *Let M^n, M_0^{n+k}, be Riemannian manifolds* ($\dim M_0 \geq \dim M$)*, and let $\gamma, \gamma_0 : [0,l] \to M, M_0$ be normal geodesics, and set $\gamma' = T$, $\gamma_0' = T_0$. Assume that for each $t \in [0,l]$ and any $X \in M_{\gamma(t)}$, $X_0 \in (M_0)_{\gamma_0(t)}$, the sectional curvatures of the sections σ, σ_0 spanned by X, T and X_0, T_0 satisfy $K(\sigma_0) \geq K(\sigma)$. Assume further that for no $t \in [0,l]$ is $\gamma_0(t)$ conjugate to $\gamma_0(0)$ along γ_0. Let V, V_0 be Jacobi fields along γ, γ_0 such that $V(0), V_0(0)$ are tangent to γ, γ_0,*

$$||V(0)|| = ||V_0(0)||, \quad \langle T, V'(0)\rangle = \langle T_0, V_0'(0)\rangle, \quad \textit{and} \quad ||V'(0)|| = ||V_0'(0)||.$$

Then for all $t \in [0,l]$,

$$||V(t)|| \geq ||V_0(t)||.$$

THEOREM 1.34 (Second Theorem (Berger)). *Let the notation be as above, and assume that for all $t \in [0,l]$ and plane sections σ, σ_0 as above, $K(\sigma_0) \geq K(\sigma)$. Assume further that for no $t \in [0,l]$ is $\gamma(t)$ a focal point of the geodesic submanifold N_0 defined by T_0. Let V, V_0 be Jacobi fields along γ, γ_0 satisfying $V'(0), V_0'(0)$ are tangent to γ, γ_0, and $||V'(0)|| = ||V_0'(0)||$, $\langle T, V(0)\rangle = \langle T_0, V_0(0)\rangle$, $||V(0)|| = ||V_0(0)||$. Then for all $t \in [0,l]$,*

$$||V(t)|| \geq ||V_0(t)||.$$

The theorems will be referred to in the sequel as Rauch I and Rauch II. Before proving them, we will investigate a few of their consequences.

Notice first of all that by continuity it is only necessary to assume that $[0,l)$ (rather than $[0,l]$) is free of conjugate (respectively focal) points. Also, suppose that the hypothesis $K(\sigma_0) \geq K(\sigma)$ holds for all $t \in [0,\infty)$. It then follows that the first conjugate (respectively focal) point along γ occurs no sooner than the first conjugate (respectively focal) point along γ_0. In fact, if γ_0 has no conjugate point on $[0,l)$ then by the theorem, for any Jacobi field as above, $||V(t)|| \geq ||V_0(t)||$. But $||V_0(t)|| > 0$ by assumption, so $||V(t)|| > 0$. *In particular, if the sectional curvature of M satisfies $K_M \leq K$, and $B_r(0)$ denotes the ball of radius r in the tangent space at $m \in M$, then* $\exp_m|B_r(0)$ *is nonsingular for $r < \pi/\sqrt{K}$.*

COROLLARY 1.35 (Corollary of Rauch I). *Let M, M_0 be Riemannian manifolds with* $\dim M_0 \geq \dim M$*, and let $m, m_0 \in M, M_0$. Assume $K_{M_0} \geq K_M$, i.e., for all plane sections $\sigma, \sigma_0 \in M, M_0$, $K(\sigma_0) \geq K(\sigma)$. Let r be chosen such that* $\exp_m|B_r(0)$ *is an imbedding and* $\exp_{m_0}|B_r(0)$ *is nonsingular. Let $I : M_m \to M_{m_0}$ be a linear injection preserving inner products. Then for any curve $c : [0,1] \to \exp_m(B_r(0))$, we have*

$$L[c] \geq L[\exp_{m_0} \circ I \circ \exp_m^{-1}(c)] = L[c_0(t)].$$

PROOF OF COROLLARY. Let $\tilde{c} : [0,1] \to B_r(0)$ be the unique curve in $B_r(0)$ such that $\exp_m \tilde{c}(s) = c(s)$. Consider the rectangle $\alpha(t,s) \to$

$\exp_m(t\tilde{c}(s))$. For fixed s, the associated variation field V_s is a Jacobi field along the geodesic $\gamma_s = t \to \exp_m(t\tilde{c}(s))$ with $V_s(1) = c'(s)$. Then

$$V_s = d\exp_m(t\tilde{c}'(s)) = t d\exp_m(\tilde{c}'(s)).$$

Therefore $\nabla_T V_s = \tilde{c}'(s)$. Similarly, associated to the rectangle $\alpha_0(t,s) \to \exp_{m_0} \circ I(t\tilde{c}(s))$ there is a Jacobi field V_{0s} with

$$V_{0s}(1) = c_0'(s), \qquad \nabla_T V_{0s} = (I \circ (\tilde{c}'(s)))' = I \circ (\tilde{c}'(s)).$$

Since I preserves lengths,

$$||c_0'(s)|| = ||I(c'(s))||.$$

By Rauch I,

$$||c'(s)|| = ||V_s(1)|| \geq ||V_{0s}(1)|| = ||c_0'(s)||,$$

and by integrating this inequality we are done. □

Corollary 1.36 (Corollary of Rauch II). *Let γ, γ_0 be geodesics on M and M_0 parameterized on $[0,l]$, with tangent vectors T and T_0. Let E and E_0 be parallel unit vectors along γ and γ_0 which are everywhere perpendicular to T and T_0. Let $c : [0,l] \to M$ be a smooth curve defined by*

$$c(t) = \exp(f(t)E(t)),$$

where $f : [0,l] \to \mathbb{R}$ is a smooth function, and let $c_0 : [0,l] \to M_0$ be defined by

$$c_0(t) = \exp(f(t)E_0(t)).$$

Assume that $K_{M_0} \geq K_M$, and assume that for each t the geodesic $\eta_0 : [0,l] \to M_0$ defined by

$$\eta_0(s) = \exp(sf(t)E_0(t))$$

contains no focal points of the geodesic submanifold defined by $\eta_0'(0)$. Then

$$L[c] \geq L[c_0].$$

Proof. Since c and c_0 are both parameterized from 0 to l, it suffices to compare the lengths of their tangent vectors.

Fix $t_1 \in [0,l]$. Let η be the geodesic

$$\eta(s) = \exp(sf(t_1)E(t))$$

and let

$$h(t) = \exp(f(t_1)E(t)), \qquad h_0(t) = \exp(f(t_1)E_0(t)).$$

Then

$$c'(t_1) = h'(t_1) + f'(t_1)\eta'(1),$$

while

$$c_0'(t_1) = h_0'(t_1) + f_0'(t_1)\eta_0'(1).$$

By the Gauss Lemma, these sums decompose $c'(t_1)$ and $c_0'(t_1)$ into pairs of perpendicular vectors. Since $E(t_1)$ and $E_0(t_1)$ are both unit vectors,

$$||f'(t_1)\eta'(1)|| = ||f'(t_1)\eta_0'(1)||.$$

Therefore we need only compare h' and h'_0. but h' and h'_0 are tangents to the families of geodesics $\tau_t(s) = \exp(sE(t))$ and $\tau_{0t}(s) = \exp(sE_0(t))$. Therefore h' and h'_0 can be extended to Jacobi fields V and V_0 along η and η_0, and since E and E_0 are parallel, these fields satisfy the hypotheses of the Second Rauch Theorem, that is

$$\nabla_{\eta'} V = \nabla_T E = 0, \qquad \nabla_{\eta'_0} V_0 = \nabla_{T_0} E_0 = 0.$$

The corollary follows.[1] □

Proof of the First Rauch Theorem 1.33. First assume that V, V_0 are perpendicular to T, T_0 or that

$$||V(0)|| = \langle T, V'(0)\rangle = 0, \qquad ||V_0(0)|| = \langle T_0, V'_0(0)\rangle = 0.$$

Consider the ratio $\frac{||V||^2}{||V_0||^2}$, as a function of the parameter t of the geodesics. Since $V_0 = 0$ only at $\gamma_0(0)$, this is well defined everywhere except at $t = 0$. By L'Hôpital's rule (taking two derivatives),

$$\lim_{t\to 0}\frac{||V||^2}{||V_0||^2} = \frac{\langle V', V'\rangle}{\langle V'_0, V'_0\rangle} = 1.$$

Therefore to show that $||V|| \geq ||V_0||$ it suffices to show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{||V||^2}{||V_0||^2}\Big) \geq 0.$$

Equivalently, for $t > 0$,

$$\frac{\langle V', V\rangle}{\langle V, V\rangle} \geq \frac{\langle V'_0, V_0\rangle}{\langle V_0, V_0\rangle}.$$

Fix $t_1 \in [0, l)$ and define vector fields

$$W_{t_1} = \frac{V(t)}{||V(t_1)||} \qquad W_{0t_1} = \frac{V_0(t)}{||V_0(t_1)||}.$$

Then $||W(t_1)|| = ||W_0(t_1)|| = 1$, and since W_{t_1} is a constant multiple of V,

$$\frac{\langle V, V'\rangle}{\langle V, V\rangle} = \frac{\langle W_{t_1}, W'_{t_1}\rangle}{\langle W_{t_1}, W_{t_1}\rangle}, \qquad \frac{\langle V_0, V'_0\rangle}{\langle V_0, V_0\rangle} = \frac{\langle W_{0t_1}, W'_{0t_1}\rangle}{\langle W_{0t_1}, W_{0t_1}\rangle}.$$

In particular,

$$\frac{\langle V, V'\rangle}{\langle V, V\rangle}\Big|_{t_1} = \langle W'_{t_1}, W_{t_1}\rangle|_{t_1}, \qquad \frac{\langle V_0, V'_0\rangle}{\langle V_0, V_0\rangle}\Big|_{t_1} = \langle W'_{0t_1}, W_{0t_1}\rangle|_{t_1}.$$

[1] One might conjecture that Corollary 1.36 could be strengthened so as to compare lengths of any two curves which have the same expression in Fermi coordinates (see Hicks [1965], p. 133, for definition). Somewhat surprisingly this turns out not to be possible.

Then

$$\begin{aligned}\frac{\langle V, V'\rangle}{\langle V, V\rangle}\Big|_{t_1} &= \langle W'_{t_1}, W_{t_1}\rangle|_{t_1}\\ &= \int_0^{t_1} \langle W'_{t_1}, W_{t_1}\rangle'\\ &= \int_0^{t_1} \langle W'_{t_1}, W'_{t_1}\rangle + \langle W''_{t_1}, W_{t_1}\rangle\\ &= \int_0^{t_1} \langle W'_{t_1}, W'_{t_1}\rangle - \langle R(W_{t_1}, T)T, W_{t_1}\rangle\\ &= \int_0^{t_1} \langle W'_{t_1}, W'_{t_1}\rangle - K(\sigma)||W_{t_1}||^2,\end{aligned}$$

where σ is the plane section spanned by T and W_{t_1}.

Let $P_{-\gamma}$ denote parallel translation along γ with the opposite parameterization and P_{γ_0} parallel translation along γ_0. Let $I : M_{\gamma(0)} \to (M_0)_{\gamma_0(0)}$ be an inner product-preserving injection and define $I_t : M_{\gamma(t)} \to (M_0)_{\gamma_0(t)}$ by

$$I_t(X) = P_{\gamma_0} \circ I \circ P_{-\gamma}(X).$$

Assume that I has been chosen so that $I_t(T) = T_0$ and $I_{t_1}(W_{t_1}) = W_{0t_1}$. Define a field $\widehat{W}_{0t_1}$ by

$$\widehat{W}_{0t_1} = I_t(W_{t_1}(t)).$$

Since $\widehat{W}_{0t_1}$ has the same expression in terms of parallel frames as W_{t_1}, clearly

$$\begin{aligned}\langle W_{t_1}(t), W_{t_1}(t)\rangle &= \langle \widehat{W}_{0t_1}(t), \widehat{W}_{0t_1}(t)\rangle,\\ \langle W'_{t_1}(t), W_{t_1}(t)\rangle &= \langle \widehat{W}'_{0t_1}(t), \widehat{W}_{0t_1}(t)\rangle.\end{aligned}$$

Using this and our assumption on the curvatures gives

$$\begin{aligned}&\int_0^{t_1} \langle W'_{t_1}, W'_{t_1}\rangle - \langle R(W_{t_1}, T)T, W_{t_1}\rangle\\ &\qquad\geq \int_0^{t_1} \langle \widehat{W}'_{0t_1}, \widehat{W}'_{0t_1}\rangle - \langle R_0(\widehat{W}_{0t_1}, T_0)T_0, \widehat{W}_{0t_1}\rangle\\ &\qquad\geq \int_0^{t_1} \langle W'_{0t_1}, W'_{0t_1}\rangle - \langle R_0(W_{0t_1}, T)T, W_{0t_1}\rangle,\end{aligned}$$

where the last inequality is just the First Index Lemma 1.24. But reversing the first part of our argument shows that this last expression is equal to

$$\frac{\langle V_0, V_0'\rangle|_{t_1}}{\langle V_0, V_0\rangle}.$$

Thus for arbitrary t_1,

$$\frac{\langle V', V\rangle}{\langle V, V\rangle}\Big|_{t_1} \geq \frac{\langle V_0', V_0\rangle}{\langle V_0, V_0\rangle}\Big|_{t_1}.$$

and we are done.

In the general case let

$$V = \widehat{V} + \langle T, V\rangle T, \qquad V_0 = \widehat{V}_0 + \langle T_0, V_0\rangle T_0.$$

Then $||\widehat{V}(t)|| \geq ||\widehat{V}_0(t)||$ as above. Also,

$$\langle T, V\rangle = \langle T, V(0)\rangle + \langle T, V'(0)\rangle t = \langle T_0, V_0\rangle,$$

so $||V(t)|| \geq ||V_0(t)||$. □

Note that the equality of the index form and the term $\langle W'_{t_0}, W_{t_0}\rangle$ may be seen geometrically as follows. Let $\tilde{c}(s)$ be a curve in $M_{\gamma(0)}$ such that $||\tilde{c}(s)|| = 1$ and

$$\mathrm{d}\exp(t_1\tilde{c}'(s))|_{s=0} = W_{t_1}(t_1).$$

Consider the rectangle $\alpha : (t, s) \to \exp_{\gamma(0)}(t\tilde{c}(s))$. Then $\mathrm{d}\alpha(\frac{\partial}{\partial s})$ is a vector field along α extending W_{t_1}. We will denote this field by W_{t_1} also. Then

$$\begin{aligned}\langle W'_{t_1}, W_{t_1}\rangle &= \langle \nabla_{W_{t_1}} T, W_{t_1}\rangle = W_{t_1}\langle T, W_{t_1}\rangle - \langle T, \nabla_{W_{t_1}} W_{t_1}\rangle \\ &= -\langle T, \nabla_{W_{t_1}} W_{t_1}\rangle\end{aligned}$$

The curves $t \to \exp_{\gamma(0)}(t\tilde{c}(s))$ are all geodesics of length t_1. Therefore the second variation formula for the variation α gives

$$\langle \nabla_{W_{t_1}} W_{t_1}, T\rangle + I(W_{t_1}, W_{t_1}) = 0,$$

where I is the index form. This, together with the above, yields

$$\langle W'_{t_1}, W_{t_1}\rangle = I(W_{t_1}, W_{t_1}).$$

Proof of the Second Rauch Theorem 1.34. In this case $||V||_0 = ||V_0||_0$ and V_0 is never zero since γ_0 contains no focal point. Therefore we need only show

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{||V||^2}{||V_0||^2}\Big) \geq 0.$$

This follows in the same way as in the previous proof by the use of the Second Index Lemma 1.25. □

Remark 1.37. It is interesting to study the case $||V(t)|| = ||V_0(t)||$ in Theorems 1.33 and 1.34. One finds immediately that all the inequalities of the proofs must be equalities. Also, by lemma 1.24,

$$\widehat{W}_{0l} = W_{0l}.$$

Thus $\langle R(V, T)T, V\rangle = \langle R(V_0, T_0)T_0, V_0\rangle$.

Warner [1966] gives technical generalizations of the Rauch Theorems.

12. The Cartan-Hadamard Theorem

An easy application of Rauch I gives us some information about manifolds of nonpositive sectional curvature. We begin with a lemma.

LEMMA 1.38. *If $\phi : M^n \to N^n$ is a local isometry and M is complete, then ϕ is a covering map; i.e. for all $p \in N$ there exists a neighborhood U of p such that $\phi^{-1}(U) = \underset{\alpha}{\cup} U_\alpha$, a union of disjoint subsets of M, and for each α, $\phi|U_\alpha : U_\alpha \to U$ is a diffeomorphism.*

PROOF. Fix $p \in N$ and assume r is small enough so that a ball about p of radius r is properly contained in a normal coordinate neighborhood of p. Let U be the ball of radius r about p. Let $\{q_\alpha\} = \phi^{-1}(p)$, and let U_α be the ball of radius r about q_α. We will show that $\phi^{-1}(U)$ is the disjoint union $\underset{\alpha}{\cup} U_\alpha$, and that $\phi : U_\alpha \to U$ is a diffeomorphism for each α.

First let $B_r(0), B_r^\alpha(0)$ be the balls about zero of radius r in N_p, M_{q_α}, respectively. Since ϕ is a local isometry, the diagram

$$\begin{array}{ccc} M_{q_\alpha} & \xrightarrow{\mathrm{d}\phi} & N_p \\ \big\downarrow{\scriptstyle \exp_{q_\alpha}} & & \big\downarrow{\scriptstyle \exp_p} \\ M & \longrightarrow & N \end{array}$$

commutes. It restricts to the diagram

$$\begin{array}{ccc} B_r^\alpha(0) & \longrightarrow & B_r(0) \\ \big\downarrow & & \big\downarrow \\ U_\alpha & \longrightarrow & U \end{array}$$

and since $\exp_p \circ \, \mathrm{d}\phi : B_r^\alpha(0) \to U_\alpha$ is a diffeomorphism, so is $\phi : U_\alpha \to U$.

It is clear that $\cup U_\alpha \subseteq \phi^{-1}(U)$. We shall show the opposite inclusion. Given $\bar{q} \in \phi^{-1}(U)$, let $q = \phi(\bar{q})$, and let γ be the normal minimal geodesic from q to p. Let $v = \mathrm{d}\phi^{-1}(\gamma'(q))$, and let $\bar{\gamma}$ be the geodesic from $\bar{q}$ in direction v. Since M is complete, $\bar{\gamma}$ may be extended arbitrary far. Let $t_0 = \rho(p, q)$ and set $q = \bar{\gamma}(t_0)$.

Since $\phi \circ \bar{\gamma} = \gamma$, $\phi(\bar{p}) = p$. Also $\rho(\bar{p}, \bar{q}) = \rho(p, q) < r$. Hence $q \in \underset{\alpha}{\cup} U_\alpha$.

It remains to show that $U_\alpha \cap U_\beta$ is empty if $\alpha \neq \beta$. For this it is clearly sufficient to show that if $\bar{p}_\alpha, \bar{p}_\beta \in \phi^{-1}(p)$, then $\rho(\bar{p}_\alpha, \bar{p}_\beta) > 2r$. Let γ be a minimal geodesic from $\bar{p}_\alpha$ to $\bar{p}_\beta$. Then $\gamma = \phi \circ \bar{\gamma}$ is a closed geodesic on p. Therefore, since p is contained in a normal coordinate neighborhood of radius $> r$, γ must have length more than $2r$. Therefore $\rho(\bar{p}_\alpha, \bar{p}_\beta) > 2r$. The lemma follows. □

THEOREM 1.39 (Cartan-Hadamard). *Let M be complete and $K_M \leq 0$. Then for any $p \in M$, $\exp_p : M_p \to M$ is a covering map. Hence the universal covering space of M is diffeomorphic to $\mathbb{R}^n$. Hence the homotopy groups $\pi_i(M)$ vanish for $i > 1$.*

PROOF. Using Rauch I and comparing with Euclidean space, we find that M has no conjugate points. Hence $\exp_p : M_p \to M$ has nonsingular differential. Therefore we can let $\exp_p$ induce a metric $\langle\langle\ ,\ \rangle\rangle$ on M_p which makes it a local isometry. The lines through the origin of M_p are geodesics in this metric because they are mapped by $\exp_p$ into geodesics on M. Hence by the Hopf-Rinow Theorem 1.10, M_p is complete in the metric $\langle\langle\ ,\ \rangle\rangle$. Therefore by Lemma 1.38, $\exp_p$ is a covering map. It is a standard fact from the theory of covering spaces that if $\bar{M}$ covers M,

$$\pi_i(\bar{M}) = \pi_i(M) \qquad \text{for } i \geq 2.$$

Hence

$$\pi_i(M) = \pi_i(\mathbb{R}^n) = 0.$$

□

COROLLARY 1.40. *Let M^n be complete, simply connected and have nonpositive curvature. Then M is diffeomorphic to $\mathbb{R}^n$.*

PROOF. A covering map onto a simply connected space must be a homeomorphism. $\exp_p$ is smooth and nonsingular, so it is a diffeomorphism. □

13. The Cartan-Ambrose-Hicks Theorem

We shall prove a theorem which tells how under suitable conditions relating the curvatures of M^n and $\bar{M}^n$, we can construct an isometry between them.

We begin with a local result. Fix $p \in M^n$, $\bar{p} \in \bar{M}^n$ and let $I : M^n_p \to \bar{M}^n_{\bar{p}}$ be a linear isometry. Let $B_r(p)$ be a normal coordinate neighborhood of p. Define $\phi : B_r(p) \to B_r(\bar{p})$ by $\phi = \exp_{\bar{p}} \circ I \circ \exp_p^{-1}$. Then if r is sufficiently small, $B_r(\bar{p})$ is a normal coordinate about $\bar{p}$ and ϕ is a diffeomorphism. Let P_γ denote parallel translation along a geodesic γ. Let $R, \bar{R}$ denote the curvature tensors in $M, \bar{M}$ and $\bar{\gamma} = \phi(\gamma)$. Set $I_\gamma = P_{\bar{\gamma}} \circ I \circ P_{-\gamma}$. The following lemma is the local version of the fact that behavior of the curvature tensor under parallel translation determines the metric.

LEMMA 1.41. *In the above situation suppose that for all geodesics γ emanating from p we have*

$$I_\gamma(R(x, y)z) = \bar{R}(I_\gamma(x), I_\gamma(y))I_\gamma(z),$$

Then ϕ is an isometry and $\mathrm{d}\phi = I_\gamma$.

PROOF. Given $x \in M_q$, let γ be the geodesic from p to $q = \gamma(t^*)$ lying in $B_r(p)$ and let J be the Jacobi field along γ such that $J(0) = 0$ and $J(t^*) = x$. Let $\gamma_t = \gamma|[0, t]$, and define $\bar{J}$ along $\bar{\gamma}$ by $\bar{J}(t) = I_{\gamma_t}(J(t))$. It follows immediately from the hypothesis that $\bar{J}(t)$ is a Jacobi field along $\bar{\gamma}$. Moreover, clearly

$$||J(t)|| = ||\bar{J}(t)||.$$

To complete the proof it will suffice to show that $\bar{J}(t) = \mathrm{d}\phi(J(t))$. From the relation

$$\bar{J}(t) = P_{\gamma_t} \circ I \circ P_{-\gamma_t}(J(t))$$

it follows that $I(J'(0)) = \bar{J}'(0)$. Since $J, \bar{J}$ are Jacobi fields vanishing at $t = 0$, we have, as in Section 6.

$$J(t) = \mathrm{d}\exp_{\gamma(0)} tJ'(0)|_{t\gamma'(0)}, \qquad \bar{J}(t) = \mathrm{d}\exp_{\bar{\gamma}(0)} t\bar{J}'(0)|_{t\bar{\gamma}'(0)}.$$

Then

$$\begin{aligned}\bar{J}(t) &= \mathrm{d}\exp_{\bar{\gamma}(0)} I(t\bar{J}'(0))|_{t\bar{\gamma}'(0)} \\ &= \mathrm{d}\exp_{\bar{\gamma}(0)} \circ \mathrm{d}I \circ \mathrm{d}\exp^{-1}_{\gamma(0)}(J(t)) = \mathrm{d}\phi(J(t)),\end{aligned}$$

which completes the proof. □

Now let M be complete. We proceed to a global version of the ablove lemma.

A broken geodesic is a continuous curve $\gamma : [0, l] \to M$ such that there exists $0 < t_0 < t_1 < \cdots < t_n < l$ and $\gamma|[t_i, t_{i+1}]$ is a smooth geodesic. Set

$$i\gamma = \gamma|[0, t_i],$$

and define v_i by

$$\gamma|[t_i, t_{i+1}] = t \to \exp_{\gamma(t_i)}((t - t_i)v_i).$$

If $I : M_p \to \bar{M}_{\bar{p}}$, we define a correspondence between broken geodesics emanating from p, $\bar{p}$ as follows: Set

$$\bar{\gamma}_1(t) = \exp_{\gamma(0)}(tI(v_0)).$$

Assume $\bar{\gamma}_i$ is already defined.

Set

$$\bar{\gamma}_{i+1}(t) = \begin{cases} \bar{\gamma}_i(t), & 0 \le t \le t_i \\ \exp_{\bar{\gamma}_i(t_i)}(t(P_{\bar{\gamma}_i} \circ I \circ P_{-\bar{\gamma}_i}(v_i))), & t_i \le t \le t_{i+1}. \end{cases}$$

Note that this is consistent with the definition of $\bar{\gamma}$ preceeding Lemma 1.41.

THEOREM 1.42 (Cartan, Ambrose, Hicks). *Let M^n, $\bar{M}^n$ be complete, M^n simply connected and $I : M^n_p \to \bar{M}^n_{\bar{p}}$ be a linear isometry. Suppose that for all broken geodesics γ,*

$$I_\gamma(R(x, y)z) = \bar{R}(I_\gamma(x), I_\gamma(y))I_\gamma(z).$$

Then for all broken geodesics γ_0, γ_1 from p such that $\gamma_0(l_0) = \gamma_1(l_1)$ we have

$$\bar{\gamma}_0(l_0) = \bar{\gamma}_1(l_1).$$

Thus there is a map $\Phi : M^n \to \bar{M}^n$ defined by $\gamma(l) \to \bar{\gamma}(\bar{l})$. Moreover, Φ is a local isometry and hence a covering map.

PROOF. (1) First assume that $\gamma_0(l_0) = \gamma_1(l_1)$, and that $\gamma_0, \gamma_1, \bar{\gamma}_0, \bar{\gamma}_1$ are contained in normal coordinate balls $B_r(p), B_r(\bar{p})$, respectively. Then Lemma 1.41 implies that the map $\phi = \exp_{\bar{p}} \circ I \circ \exp_p^{-1}|B_r(p)$ is an isometry. It follows that $\phi(\gamma_i) = \bar{\gamma}_i$. Therefore $\bar{\gamma}_0(l_0) = \bar{\gamma}_1(l_1)$ and $\mathrm{d}\phi = I_{\gamma_0} = I_{\gamma_1}$.

(2) Now, for convenience and without real loss of generality, we will suppose that γ_0, γ_1 both have n breaks at the points $t_1 < \cdots < t_n$, and $l_0 = l_1 = l$. Assume that for all i, $\gamma_1(t_{i+2}), \gamma_1(t_{i+1}), \gamma_0(t_{i+1})$ and the minimal segments between them lie in a normal coordinate ball about $\gamma_0(t_i)$ and that the same is true for $\bar{\gamma}_1(t_{i+2}), \bar{\gamma}_1(t_{i+1}), \bar{\gamma}_0(t_{i+1}), \bar{\gamma}_0(t_i)$. Let $\tau : (t_{n-1}, t_n] \to M$ be the minimal geodesic from $\gamma_0(t_{n-1})$ to $\gamma_1(t_n)$. By induction we may assume that

$$\overline{{}_{n-1}\gamma_0 \cup \tau(t_n)} = \overline{{}_n\gamma_1(t_n)}$$

and that

$$I_{n-1\gamma_0 \cup \tau} = I_{n\gamma_1}. \qquad (*)$$

Also the isometry $I_{n-1\gamma_0} : M_{\gamma_0(t_{n-1})} \to \bar{M}_{\bar{\gamma}_0(t_{n-1})}$ induces a correpondence (which we denote with a double bar) between geodesics emanating from $\gamma_0(t_{n-1})$ and $\bar{\gamma}_0(t_{n-1})$. Set

$$\sigma_0 = \gamma_0|[t_{n-1}, l], \qquad \theta_1 = \gamma 1|[t_n, l].$$

Applying step (1) gives

$$\overline{\overline{\sigma_0}}(l) = \overline{\overline{\tau \cup \theta_1}}(l), \qquad I_{\sigma_0} = I_{\tau \cup \theta_1}.$$

But this is equivalent to

$$\bar{\gamma}_0(l) = \overline{{}_{n-1}\gamma_0 \cup \tau \cup \theta_1}(l), \qquad I_{\gamma_0} = I_{n-1\gamma_0 \cup \tau \cup \theta_1}.$$

Using $(*)$, the isometry on the right can be rewritten as

$$P_{\bar{\theta}_1} \circ I_{n-1\gamma_0 \cup \tau} \circ P_{-\theta_1} = P_{\bar{\theta}_1} \circ I_{n\gamma_1} \circ P_{-\theta_1} = I_{\gamma_1}.$$

In particular, $\bar{\gamma}_0(l) = \bar{\gamma}_1(l)$.

(3) Now let γ_0 and γ_1 be any two broken geodesics such that $\gamma_0(l) = \gamma_1(l)$. Since M is simply connected, there is a homotopy h_s from γ_0 to γ_1. By uniform continuity of h_s we may choose subdivisions $0 < s_1 < \cdots < s_m < 1$ and $0 < t_1 < \cdots < t_n < l$ such that for all i, j, the points $h_{s_{j+1}}(t_{i+2})$, $h_{s_{j+1}}(t_{i+1})$, $h_{s_j}(t_{i+1})$ and the geodesics between them lie in a normal coordinate neighborhood of $h_{s_j}(t_i)$. By inserting further breakpoints we can assume that $\{t_i\}$ is exactly the set of breakpoints of γ_0, γ_1. Let γ_{s_j} denote the broken geodesic formed by minimal segments from $h_{s_j}(0)$ to $h_{s_j}(t_1)$, $h_{s_j}(t_1)$ to $h_{s_j}(t_2)$ .... Since from the theory of ordinary differential equations the correspondence $\gamma \to \bar{\gamma}$ is continuous in an obvious sense, we may assume that for all j, $\bar{\gamma}_{s_j}$ and $\bar{\gamma}_{s_{j+1}}$ are sufficiently close, so that for all i, $\bar{\gamma}_{s_{j+1}}(t_{i+2}), \bar{\gamma}_{s_{j+1}}(t_{i+1}), \bar{\gamma}_{s_j}(t_{i+1})$ lie in a normal coordinate neighborhood of $\bar{\gamma}_{s_j}(t_i)$. Therefore, each pair $\bar{\gamma}_{s_j}, \bar{\gamma}_{s_{j+1}}$ satisfies the hypothesis of step (2). Therefore

$$\bar{\gamma}_0(l) = \bar{\gamma}_{s_1}(l) = \cdots = \bar{\gamma}_1(l).$$

Finally, let $q \in M$ be arbitrary, and let γ be a geodesic such that $\gamma(l) = q$. Then by Lemma 1.41 the map

$$\phi = \exp_{\bar{\gamma}(l)} \circ I_\gamma \circ \exp^{-1}_{\gamma(l)}$$

is an isometry from a neighborhood of $B_r(\gamma(l))$ to $B_r(\bar{\gamma}(l))$. But referring to the definition of the correspondence $\gamma \to \bar{\gamma}$ and of Φ, one sees that $\phi = \Phi | B_r(q)$. Therefore Φ is a local isometry. □

14. Spaces of constant curvature

The simplest examples of Riemannian manifolds are those whose sectional curvature is a constant K. The complete ones are called *space forms*. We will show that for each K all simply connected spaces forms with curvature K are isometric. They may be described as follows:

(a) $K = 0$. Let $M^n = \mathbb{R}^n$ with the usual metric.

(b) $K > 0$. Let $M^n = S^n_{\frac{1}{\sqrt{K}}}$, the sphere in $\mathbb{R}^{n+1}$, with the induced metric.[2]

(c) $K < 0$. Let M^n be the open set in $\mathbb{R}^n$ defined by

$$M = \Big\{ x \in \mathbb{R}^n \Big| \ ||x||^2 < -\frac{4}{K} \Big\}.$$

Using the standard coordinates in $\mathbb{R}^n$, define the metric by

$$\langle v, w \rangle = \frac{\sum_{i=1}^{n} v_i w_i}{1 + \frac{1}{4} K \sum_{i=1}^{n} (x_i)^2},$$

where $v, w \in M_x$.

Recall from Section 6 that if M has constant curvature K, then

$$R(x,y)z = K(\langle z, y \rangle x - \langle z, x \rangle y).$$

THEOREM 1.43. *Let M^n and $\bar{M}^n$ be complete simply connected manifolds with constant curvature K. Then M^n and $\bar{M}^n$ are isometric.*

In fact, given any $p \in M^n$, $\bar{p} \in \bar{M}^n$ and an isometry $I : M^n_p \to \bar{M}^n_{\bar{p}}$, there exists an isometry $\Phi : M^n \to \bar{M}^n$ such that $\Phi(p) = \bar{p}$ and $\mathrm{d}\Phi_p = I$.

PROOF. This is immediate from the Cartan-Ambrose-Hicks Theorem 1.42 and the formula above for $R(x,y)z$. □

Theorem 1.42 (or Lemma 1.41) shows in particular that the vanishing of the sectional curvature is a necessary and sufficient condition for M to be locally isometric to Euclidean space.

[2] As a matter of notation, we will denote by S^n_r the n-sphere of radius r. We denote by S^n the unit sphere and by S^n_p the tangent space at $p \in S^n$.