Notes of Riemannian Geometry, A.A. 24/25

Diego Conti

December 11, 2025

These notes cover a subset of [4], with some proofs taken from [7, Chapter 6] and [3] instead. These notes are meant as a guide to reading the books.

Please let me know of any errors.

I will correct any errors of which I become aware, but other than that, I do not expect to modify this text.

1 Contents

The course will focus on comparison theorems in Riemannian geometry, namely on results that put constraints on the topology of Riemannian manifolds that satisfy various bounds on the sectional and Ricci curvature. Most of the results are contained in the classical book of Cheeger and Ebin [4], but I will also refer to the more recent book by Petersen [7].

I will begin with some explicit curvature computations, particularly for the Berger sphere and the Fubini-Study metric. I will then introduce variations, and in particular geodesic variations, i.e. variations of a geodesic through geodesics. At the infinitesimal level, this leads to the notion of Jacobi vector fields, which will be a central topic in the course. A first application of Jacobi fields will be the classification of manifolds of constant curvature through the Cartan-Ambrose-Hicks theorem.

On the other hand, Jacobi fields allow one to characterize the cut locus (roughly, the locus in the tangent space T_pM at which $\exp_p$ ceases to be a diffeomorphism). In particular, we will see that for manifolds with sectional curvature $\sec \leq 0$, the exponential map has no critical points, leading to the Cartan-Hadamard theorem, stating that any complete manifold with $\sec \leq 0$ is topologically a quotient of $\mathbb{R}^n$. We will also see two further restrictions on the topology (more precisely, the fundamental group) in this context, namely the theorems of Cartan and Preissmann.

In the positively curved setting, I will illustrate Myers' theorem for complete metrics on which the Ricci curvature Ric is positively bounded below, stating in particular that such manifolds are compact, and the Cheeger-Gromoll splitting theorem, which holds more generally for $\mathrm{Ric} \geq 0$.

Turning to the more restrictive case where $\sec > 0$, I will prove Synge's theorem, based on the energy variation formula, and two results where sec is additionally required to be bounded above, namely Klingerberg's estimate on the injectivity radius and Berger's sphere theorem (at least in even dimensions, following [**?**]).

I will conclude with the soul theorem of Cheeger and Gromoll, which realizes any noncompact complete manifold with sec ≥ 0 as the normal bundle of a totally geodesic compact submanifold.

2 Connections

This introductory section establishes notation and recalls some facts about the Levi-Civita connection, connection forms, and curvature.

A Riemannian metric g on a manifold M is a tensor that defines a positive definite scalar product at each point, $g_x \colon T_xM \times T_xM \to \mathbb{R}$. One than says that (M, g) (or simply M) is a Riemannian manifold.

We will say that $\{e_1, \dots, e_n\}$ is a *frame* if the e_i are vector fields on some $U \subset M$ such that at each $x \in U$, $\{(e_1)_x, \dots, (e_n)_x\}$ is a basis of T_xM. We will denote by $\{e^1, \dots, e^n\}$ the dual coframe, i.e. at each $x \in M$ $\{(e_1)_x, \dots, (e_1)_x\}$ and $\{(e^1)_x, \dots, (e^1)_x\}$ are dual bases of T_xM, T_x^*M respectively. The frame is *orthonormal* if at each point it determines an orthonormal basis of T_xM; in other words, if the metric takes the form $\sum e^i \otimes e^i$.

If E is a vector bundle over M, we denote by $\Gamma(M, E)$ the space of smooth sections of E. Denoting by $C^\infty(M)$ the algebra of smooth functions $M \to \mathbb{R}$, we see that $\Gamma(M, E)$ is a $C^\infty(M)$-module.

Given a tensor F of type $\binom{k}{l}$, a $C^\infty(M)$-multilinear map is defined as

$$\tau_F \colon \overbrace{\mathfrak{X}(M) \times \cdots \times \mathfrak{X}(M)}^{k} \times \underbrace{\Omega^1(M) \times \cdots \times \Omega^1(M)}_{l} \to C^\infty(M),$$

$$\tau_F(X_1, \dots, X_k, \omega_1, \dots, \omega_l)(x) = F_x((X_1)_x, \dots, (X_k)_x, (\omega_1)x, \dots, (\omega_l)x).$$

Example 2.1. If F is the tensor $g_{ij}dx^i \otimes dx^j$, then the associated map is

$$\tau_F(\tfrac{\partial}{\partial x_i}, \tfrac{\partial}{\partial x_j}) = g_{ij}.$$

Proposition 2.2. *A map*

$$\tau \colon \underbrace{\Omega^1(M) \times \cdots \times \Omega^1(M)}_{l} \times \overbrace{\mathfrak{X}(M) \times \cdots \times \mathfrak{X}(M)}^{k} \to C^\infty(M)$$

is induced by a tensor if and only if it is $C^\infty(M)$-multilinear.

Proof. Given a $C^\infty(M)$-multilinear map τ and $x \in M$, we first show that

$$\tau(X_1, \dots, X_k, \omega_1, \dots, \omega_l)(x)$$

depends only on $(X_1)_x, \dots, (X_k)_x, (\omega_1)_x, \dots, (\omega_l)_x$. Take a local frame $e_1, \dots, e_n$ defined in a neighborhood U of x. Let ρ be a bump function with support in U and $\rho \equiv 1$ in a neighborhood of x, so that $\rho e_i, \rho e^i$ are globally defined. Let X_1 be a vector field that vanishes at x, and suppose $X_1 = a^i e_i$ on U. Then

$$X_1 = (\rho a^i)(\rho e_i) + (1 - \rho^2)X_1$$

Since ρa^i and ρe_i are globally defined, we can write

$$\tau(X_1, \dots, X_k, \omega_1, \dots, \omega_l) = \rho a^i \tau(\rho e_i, X_2, \dots, X_k, \omega_1, \dots, \omega_l) + (1-\rho^2)\tau(X_1, \dots, X_k, \omega_1, \dots, \omega_l);$$

evaluating at x, we find that if X_1 vanishes at x, then $\tau(X_1,\dots,X_k,\omega_1,\dots,\omega_l)$ also vanishes at x. This proves that

$$\tau(X_1,\dots,X_k,\omega_1,\dots,\omega_l)(x)$$

depends only on $(X_1)_x,\dots,(X_k)_x,(\omega_1)_x,\dots,(\omega_l)_x$. At this point, on the neighbourhood of x where $\rho=1$ we can define

$$F^x = F^{j_1,\dots,j_l}_{i_1,\dots,i_k} e^{i_1}\otimes\dots\otimes e^{i_k}\otimes e_{j_1}\otimes\dots e_{j_l}, \quad F^{j_1,\dots,j_l}_{i_1,\dots,i_k}(y) = \tau(\rho e_{i_1},\dots,\rho e_{i_k},\rho e^{j_1},\dots,\rho e^{j_l})(y);$$

this clearly defines a tensor F^x on a neighbourhood of x. Since F^x is uniquely determined in this neighbourhood, the tensors F^x glue together to give a global tensor. □

Remark 2.3. A similar statement can be made from maps taking values in $\Omega^1(M)$ or $\mathcal{X}(M)$.

A connection on a vector bundle E is a map

$$\nabla\colon \mathcal{X}(M)\times\Gamma(M,E)\to\Gamma(M,E), \quad (X,s)\mapsto\nabla_X s,$$

satisfying the following conditions:

1. ∇ is $C^\infty(M)$-linear in X;
2. ∇ is $\mathbb{R}$-linear in s;
3. $\nabla_X(fs)=(Xf)s+f\nabla_X s$ for any $f\in C^\infty(M)$.

Example 2.4. Given any conection on TM, its *torsion* is the tensor corresponding to the $C^\infty(M)$-bilinear map

$$(X,Y)\mapsto\nabla_X Y-\nabla_Y X-[X,Y].$$

Proposition 2.5. *If ∇ is a connecton on TM, there is a unique connection ∇ on each $(T^*M)^{\otimes^r}\otimes TM^{\otimes s}$ such that*

1. *$\nabla f = df$ for $r=s=0$;*
2. *∇ coincides with the given connection for $s=1$, $r=0$;*
3. *$\nabla_X(\alpha\otimes\beta)=\nabla_X\alpha\otimes\beta+\alpha\otimes\nabla_X\beta$;*
4. *∇_X commutes with contractions.*

Proof. Let c be the contraction from $T^*M\otimes TM\to\Lambda^0 M$, i.e. $c(\alpha\otimes Y)=\alpha(Y)$. If $T^*M\otimes M$ is identified with $\operatorname{End}TM$, then c is the pointwise trace.

If ∇ is to satisfy the conditions in the statement, we must have

$$\nabla_X(\alpha\otimes Y)=\nabla_X(\alpha)\otimes Y+\alpha\otimes\nabla_Y;$$

taking the contraction, the left hand side gives

$$c(\nabla_X(\alpha\otimes Y))=\nabla_X(c(\alpha\otimes Y))=\nabla_X(\alpha(Y))=X\alpha(Y),$$

whilst the right hand side gives

$$c(\nabla_X(\alpha)\otimes Y+\alpha\otimes\nabla_X Y)=(\nabla_X\alpha)(Y)+\alpha(\nabla_X Y),$$

so that we obtain

$$(\nabla\alpha)(Y) = d(\alpha(Y)) - \alpha(\nabla Y).$$

One can easily check that this defines a connection on T^*M. At this point, there is only one way of extending a connection on all tensor products so that 3. is satisfied. $\square$

The Levi-Civita connection is the only torsion-free connection compatible with the metric, i.e. the only connection on TM such that

$$dg(X,Y) = g(\nabla X, Y) + g(X, \nabla Y), \quad \nabla_X Y - \nabla_Y X - [X,Y] = 0.$$

The *connection form* is the matrix of one-forms (ω^i_j), where

$$\omega^i_j(X) = e^i(\nabla_X e_j).$$

The connection form is not a tensor, because ∇ is not $C^\infty(M)$-linear, but $\omega(X)$ maps sections to sections, as a section of $\operatorname{End}(TM)$ would; indeed, index placement is consistent with the fact that for $A \in \Gamma(\operatorname{End} TM) = \Gamma(T^*M \otimes TM)$, one can write

$$A = A^i_j e^j \otimes e_i, \quad A(e_j) = A^i_j e_i,$$

using the Einstein convention. Similarly, we can write

$$\nabla e_j = \omega^i_j \otimes e_i.$$

Remark 2.6. The compatibility between ∇ and g can be expresssed as $\nabla g = 0$.

If ω^i_j is the Levi-Civita connection form relative to a frame $\{e_1, \dots, e_n\}$, then

$$\nabla e^i = -\omega^i_j \otimes e^j.$$

If the frame is orthonormal, ω^i_j is skew-symmetric, and one can also write $\nabla e^j = \sum_i \omega^i_j \otimes e^i$.

This reflects the fact that covariant derivatives commutes with the musical isomorphisms

$$\flat\colon a_i e_i \mapsto a_i e^i, \quad \sharp\colon a_i e^i \mapsto a_i e_i.$$

Indeed,

$$\begin{aligned} \nabla_X a_j e_j &= (X a_j) e_j + a_j \omega_{ij}(X) e_i, \\ \nabla_X a_j e^j &= (X a_j) e^j + a_j \omega_{ij}(X) e^i, \end{aligned}$$

i.e. $\nabla_X(Y^\flat) = (\nabla_X Y)^\flat$.

Remark 2.7. The fact that ∇ is torsion-free can be rewritten as

$$de^i + \omega^i_j \wedge e^j = 0. \tag{1}$$

This follows from the general formula

$$d\alpha(X,Y) = X\alpha(Y) - Y\alpha(X) - \alpha([X,Y]), \quad \alpha \in \Omega^1(M), X, Y \in \mathcal{X}(M). \tag{2}$$

Indeed, fixing a frame $\{e_i\}$,

$$\begin{aligned} de^i(e_h, e_k) = -e^i([e_h, e_k]) = -e^i(\nabla_{e_h} e_k - \nabla_{e_k} e_h) = -\omega^i_k(e_h) + \omega^i_h(e_k) \\ = (-\omega^i_j \wedge e^j)(e_h, e_k). \end{aligned}$$

Relative to a coordinate frame $\{\frac{\partial}{\partial x_i}\}$, the connection form is determined by the Christoffel symbols

$$\nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_j} = \Gamma_{ij}^k \frac{\partial}{\partial x_k};$$

in these terms, the torsion-free condition reads

$$\Gamma_{ij}^k = \Gamma_{ji}^k.$$

The Levi-Civita connection can be computed using Koszul's formula

$$\begin{aligned} 2g(\nabla_X Y, Z) = Xg(Y,Z) + Yg(X,Z) - Zg(X,Y) \\ + g([X,Y],Z) - g([X,Z],Y) - g([Y,Z],X), \end{aligned} \quad (3)$$

which in terms of Christoffel symbols reads

$$2\Gamma_{ij}^k = g^{hk}(\tfrac{\partial}{\partial x_i} g_{jh} + \tfrac{\partial}{\partial x_j} g_{ih} - \tfrac{\partial}{\partial x_h} g_{ji}).$$

The curvature form is the section R of $\Lambda^2 M \otimes \operatorname{End} TM$ given locally by

$$R_j^i = d\omega_j^i + \omega_k^i \wedge \omega_j^k. \quad (4)$$

Proposition 2.8. *If X, Y, Z are vector fields,*

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

Proof. Let $\{e_i\}$ be an orthonormal frame. Then

$$\nabla_X \nabla_Y e_j = \nabla_X(\omega_j^i(Y) e_i) = X(\omega_j^i(Y)) e_i + \omega_j^k(Y)\omega_k^i(X) e_i.$$

Thus,

$$\begin{aligned} &\nabla_X \nabla_Y e_j - \nabla_Y \nabla_X e_j - \nabla_{[X,Y]} e_j \\ &= \big(X(\omega_j^i(Y)) - Y(\omega_j^i(X) - \omega_j^i([X,Y]\big) e_i + \big(\omega_j^k(Y)\omega_k^i(X) - \omega_j^k(X)\omega_k^i(Y)\big) e_i \\ &\qquad = d\omega_j^i(X,Y) + \omega_k^i(X)\omega_j^k(Y) - \omega_k^i(Y)\omega_j^k(X) = R_j^i(X,Y) e_i. \end{aligned}$$

since $R_j^i(X,Y)$ is the i-th component of $R(X,Y)e_j$, this completes the proof when Z belongs to the frame.

In general, one writes $Z = Z^i e_i$ and shows that both sides are $C^\infty(M)$-bilinear in Z. □

Proposition 2.8 implies that R is globally defined and independent of the frame. Taking d of (1) gives

$$R_j^i \wedge e^j = 0,$$

which evaluated on X, Y, Z yields the first Bianchi identity,

$$R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0.$$

We will also use the Riemann tensor, defined by

$$\operatorname{Rm}(X,Y,Z;W) = g\left(R(X,Y)Z, W\right),$$

which satisfies the identities

$$\mathrm{Rm}(X,Y,Z,W)=\mathrm{Rm}(Z,W,X,Y)=-\mathrm{Rm}(Y,X,Z,W)=-\mathrm{Rm}(X,Y,W,Z).$$

One then defines the *sectional curvature*

$$\sec(v,w)=\frac{1}{|v\wedge w|^2}\mathrm{Rm}(v,w,w,v);$$

by construction, sectional curvature only depends on the plane spanned by v and w in T_pM; it is a consequence of Gauss' equation that $\sec(v,w)$ it measures the curvature at p of the surface $\exp_p(\mathrm{Span}\,\{v,w\})$.

Remark 2.9. The identity

$$6\,\mathrm{Rm}(x,y,z,t)=$$
$$\frac{\partial^2}{\partial\alpha\partial\beta}|_{\alpha=0,\beta=0}\big(\mathrm{Rm}(x+\alpha z,y+\beta t,x+\alpha z,y+\beta t)-\mathrm{Rm}(x+\alpha t,y+\beta z,x+\alpha t,y+\beta z)$$

implies that R is determined by sectional curvatures.

Sectional curvature is related to the *Ricci curvature*, which comes in two flavours: the Ricci tensor

$$\mathrm{ric}(v,w)=\mathrm{tr}(R(\cdot,v)w),$$

and the Ricci operator

$$\mathrm{Ric}\colon T_pM\to T_pM,\quad g(\mathrm{Ric}(v),w)=\mathrm{ric}(v,w).$$

Due to the symmetries of the Riemann tensor, ric is a symmetric bilinear form, and Ric is self-adjoint relative to g. In terms of an orthonormal frame $\{e_i\}$, we have

$$\mathrm{ric}(v,w)=\sum_i \mathrm{Rm}(e_i,v,w,e_i).$$

By construction,

$$\mathrm{ric}(e_1,e_1)=\sum_{i=2}^n \sec(e_1,e_i). \tag{5}$$

Remark 2.10. In dimension 2, the Riemann tensor has only one component, i.e. $R=ke^1\wedge e^2\otimes(e^2\otimes e_1-e^1\otimes e_2)$. We can identify k with either the sectional curvature or scalar curvature. In particular, the Ricci tensor is $\mathrm{Ric}=kg$.

Remark 2.11. In dimension 3, the Ricci tensor completely determines the Riemann tensor. This is because at a point x, the Riemann tensor lies in $S^2(\Lambda^2\mathbb{R}^3)$, and the contraction map

$$S^2(\Lambda^2\mathbb{R}^3)\to S^2\mathbb{R}^3$$

is an isomorphism.

Remark 2.12. Consider the space $(\mathbb{R}^n)^*\otimes(\mathbb{R}^n)^*\otimes(\mathbb{R}^n)^*\otimes(\mathbb{R}^n)^*$ of multilinear maps from $\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^n$ to $\mathbb{R}$.

An element R of $(\mathbb{R}^n)^*\otimes(\mathbb{R}^n)^*\otimes(\mathbb{R}^n)^*\otimes(\mathbb{R}^n)^*$ determines another element of the same space, call it $\mathfrak{S}R\in(\mathbb{R}^n)^{\otimes^4}$, satisfying

$$\mathfrak{S}R(X,Y,Z,W)=R(X,Y,Z,W)+R(Y,Z,X,W)+R(Z,X,Y,W).$$

The Riemann tensor at a point defines an element R such that $\mathfrak{S}R = 0$; in addition, R lies in the subspace $S^2(\Lambda^2(\mathbb{R}^n)^*))$.

If R is an element of $S^2(\Lambda^2(\mathbb{R}^n)^*)$, then $\mathfrak{S}R = 0$ if and only if R is in the kernel of

$$S^2(\Lambda^2(\mathbb{R}^n)^*) \to \Lambda^2(\mathbb{R}^n)^* \otimes \Lambda^2(\mathbb{R}^n)^* \to \Lambda^4(\mathbb{R}^n)^*,$$

where the first map is inclusion and the second map is the wedge product. Indeed, if α, β lie in $\Lambda^2(\mathbb{R}^n)^*$, we compute

$$\begin{aligned}(\alpha \wedge \beta)(X, Y, Z, W) &= \alpha(X, Y)\beta(Z, W) + \alpha(Z, X)\beta(Y, W) + \alpha(X, W)\beta(Y, Z)\\ &\quad + \alpha(Z, W)\beta(X, Y) + \alpha(Y, W)\beta(Z, X) + \alpha(Y, Z)\beta(X, W)\\ &= \mathfrak{S}(\alpha \otimes \beta + \beta \otimes \alpha).\end{aligned}$$

Since an element of $S^2(\Lambda^2(\mathbb{R}^n)^*)$ is the sum of elements of the form $\alpha\otimes\beta+\beta\otimes\alpha$, the claim follows.

More generally, we can consider connections on a general vector bundle E; the notion of torsion is not generally defined, but we can still define the connection form as a local object taking the shape of a one-form taking values in $\operatorname{End} E$ (though not a tensor), and the curvature defined locally as in (4) satisfies

$$R(X, Y)\alpha = \nabla_X \nabla_Y \alpha - \nabla_Y \nabla_X \alpha - \nabla_{[X,Y]}\alpha, \tag{6}$$

for any section α.

In particular, the connection induced on any bundle of tensors by the Levi-Civita connection satisfies (6).

Notation. Throughout these notes, the symbols $M, g, \nabla, \omega, \Gamma^k_{ij}, \omega^i_j$ will have the meaning introduced above. In particular, I will not repeat "let (M, g) be a Riemannian manifold" at the beginning of every theorem.

3 The Berger sphere ([7, §4.4.3])

Consider the group of matrices

$$\mathrm{SU}(2) = \begin{pmatrix} a & -\bar{b} \\ b & \bar{a} \end{pmatrix}, (a, b) \in S^3 \subset \mathbb{C}^2.$$

At the identity, we have

$$\mathfrak{su}(2) = T_e\mathrm{SU}(2) = \mathrm{Span}\,\{\underline{\mathrm{i}}, \underline{\mathrm{j}}, \underline{\mathrm{k}}\}$$

where

$$\underline{\mathrm{i}} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \underline{\mathrm{j}} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \underline{\mathrm{k}} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

satisfy the quaternionic identities $\underline{\mathrm{i}}\underline{\mathrm{j}} = \underline{\mathrm{k}}$, $\underline{\mathrm{i}}^2 = -I = \underline{\mathrm{j}}^2 = \underline{\mathrm{k}}^2$. We can induce left-invariant vector fields using left multiplication, which we also denote by $\underline{\mathrm{i}}, \underline{\mathrm{j}}, \underline{\mathrm{k}}$, i.e.

$$\underline{\mathrm{i}} = \begin{pmatrix} ia & i\bar{b} \\ ib & -i\bar{a} \end{pmatrix}, \quad \underline{\mathrm{j}} = \begin{pmatrix} \bar{b} & a \\ -\bar{a} & b \end{pmatrix}, \quad \underline{\mathrm{k}} = \begin{pmatrix} -i\bar{b} & ia \\ i\bar{a} & ib \end{pmatrix}$$

In the coordinates $a = x_0 + ix_1, b = x_2 + ix_3 \in \mathbb{C}^2$, we can write

$$\underline{\mathrm{i}} = x_0\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_0} + x_2\frac{\partial}{\partial x_3} - x_3\frac{\partial}{\partial x_2}, \quad \underline{\mathrm{j}} = x_2\frac{\partial}{\partial x_0} - x_3\frac{\partial}{\partial x_1} - x_0\frac{\partial}{\partial x_2} + x_1\frac{\partial}{\partial x_3},$$
$$\underline{\mathrm{k}} = -x_3\frac{\partial}{\partial x_0} - x_2\frac{\partial}{\partial x_1} + x_0\frac{\partial}{\partial x_3} + x_1\frac{\partial}{\partial x_2},$$

from which one can check that Lie brackets coincide with commutators, i.e.

$$[\underline{\mathrm{i}}, \underline{\mathrm{j}}] = 2\underline{\mathrm{k}}, \quad [\underline{\mathrm{j}}, \underline{\mathrm{k}}] = 2\underline{\mathrm{i}}, \quad [\underline{\mathrm{j}}, \underline{\mathrm{k}}] = 2\underline{\mathrm{i}}.$$

Consider the left-invariant metric for which

$$e_1 = \frac{1}{a}\underline{\mathrm{i}}, e_2 = \underline{\mathrm{j}}, e_3 = \underline{\mathrm{k}}$$

is an orthonormal frame, so that

$$[e_1, e_2] = \frac{2}{a}e_3, \quad [e_3, e_1] = \frac{2}{a}e_2, \quad [e_2, e_3] = 2ae_1.$$

For $a = 1$, we obtain the standard metric on the sphere. By (2), the dual coframe satisfies

$$de^1 = -2ae^2 \wedge e^3, \quad de^2 = -\frac{2}{a}e^3 \wedge e^1, \quad de^3 = -\frac{2}{a}e^1 \wedge e^2.$$

This is known as the *Berger sphere* metric. The Koszul formula for vector fields X, Y, Z in an orthonormal frame can be written as

$$2g(\nabla_X Y, Z) = g([X, Y], Z) + g([Z, X], Y) - g([Y, Z], X),$$

where the nonzero summands can be arranged in the following table:

X	Y	Z	$g([X,Y],Z)$	$g([Z,X],Y)$	$g([Y,Z],X)$
e_1	e_2	e_3	a^{-1}	a^{-1}	a
e_2	e_3	e_1	a	a^{-1}	a^{-1}
e_3	e_1	e_2	a^{-1}	a	a^{-1}
e_2	e_1	e_3	$-a^{-1}$	$-a$	$-a^{-1}$
e_1	e_3	e_2	$-a^{-1}$	$-a^{-1}$	$-a$
e_3	e_2	e_1	$-a$	$-a^{-1}$	$-a^{-1}$

Thus

$$\nabla_{e_1}e_2 = (2a^{-1} - a)e_3, \quad \nabla_{e_2}e_3 = ae_1, \quad \nabla_{e_3}e_1 = ae_2$$
$$\nabla_{e_2}e_1 = -ae_3, \quad \nabla_{e_1}e_3 = (-2a^{-1} + a)e_2, \quad \nabla_{e_3}e_2 = -ae_1.$$

The connection form is therefore

$$\omega = \begin{pmatrix} 0 & -ae^3 & ae^2 \\ ae^3 & 0 & (-2a^{-1}+a)e^1 \\ -ae^2 & (2a^{-1}-a)e^1 & 0 \end{pmatrix}.$$

Thus

$$R = d\omega + \omega \wedge \omega = \begin{pmatrix} 0 & a^2e^1 \wedge e^2 & -a^2e^3 \wedge e^1 \\ * & 0 & (4-3a^2)e^2 \wedge e^3 \\ * & * & 0 \end{pmatrix}$$

i.e.

$$R(e_1,e_2)=a^2(e^2\otimes e_1-e^1\otimes e_2),\quad R(e_1,e_3)=a^2(e^3\otimes e_1-e^1\otimes e_3),$$
$$R(e_2,e_3)=(4-3a^2)(e^3\otimes e_2-e^2\otimes e_3).$$

Thus,

$$\sec(e_1,e_2)=a^2=\sec(e_1,e_3),\quad \sec(e_2,e_3)=4-3a^2;$$

If we take an arbitrary plane Span $\{\cos\theta e_1+\sin\theta e_2,e_3\}$, we see that

$$\sec(\cos\theta e_1+\sin\theta e_2,e_3)=\mathrm{Rm}(\cos\theta e_1+\sin\theta e_2,e_3,e_3,\cos\theta e_1+\sin\theta e_2))$$
$$=\cos^2\theta\,\mathrm{Rm}(e_1,e_3,e_3,e_1)+\sin^2\theta\,\mathrm{Rm}(e_2,e_3,e_3,e_2))=a^2\cos^2\theta+(4-3a^2)\sin^2\theta;$$

we conclude that sectional curvature lies in the interval

$$a^2\leq\sec\leq 4-3a^2.$$

The Ricci operator is

$$\mathrm{Ric}=2a^2e^1\otimes e_1+(4-2a^2)(e^2\otimes e_2+e^3\otimes e_3).$$

4 Complex projective space [7, §4.5.3]

Given two Riemannian manifolds (M,g), (N,h), a Riemannian submersion $f\colon M\to N$ is a submersion such that $h(df_x(v),df_x(w))=g(v,w)$ whenever v,w are orthogonal to $\ker df_x$.

Let $S^1=\{\lambda\in\mathbb{C}\mid|\lambda|=1\}$ act on $S^{2n+1}\subset\mathbb{C}^{n+1}$ diagonally. Then the quotient S^{2n+1}/S^1 is $\mathbb{CP}^n$. The projection $S^{2n+1}\to\mathbb{CP}^n$ is called Hopf fibration. The aim is to obtain a metric on $\mathbb{CP}^n$ such that the projection is a Riemannian submersion.

Write the generic point of $\mathbb{C}^{n+1}$ as $(\rho_1e^{i\gamma_1},\dots,\rho_{n+1}e^{i\gamma_{n+1}})$; the Euclidean metric reads $\sum d\rho_i^2+\rho_i^2d\gamma_i^2$. Write

$$H^{n+1}=\tfrac{\partial}{\partial\gamma_1}+\dots+\tfrac{\partial}{\partial\gamma_{n+1}}.$$

Then $g(H^{n+1},H^{n+1})=\rho_1^2+\dots+\rho_n^2$. The isomorphism between tangent and cotangent spaces induced by the metric $\sum d\rho_i^2+\rho_i^2d\gamma_i^2$ determines a one-form $(H^{n+1})^\flat$; explicitly,

$$(H^{n+1})^\flat=\rho_1^2d\gamma_1+\dots+\rho_{n+1}^2d\gamma_{n+1}.$$

Notice that H^{n+1} is tangent to S^{2n+1}; its restriction is a unit vector field, and its integral curves are the orbits of the S^1 action.

Now parametrize the sphere $S^{2n+1}\subset\mathbb{R}^{2n}\times\mathbb{C}$ by

$$\phi\colon(0,\pi/2)\times S^{2n-1}\times S^1\to S^{2n+1}\subset\mathbb{C}^n\times\mathbb{C},\quad(t,x,e^{i\theta})\mapsto\big((\sin t)x,(\cos t)e^{i\theta}\big).\tag{7}$$

At the infinitesimal level,

$$d\phi|_{t,x,e^{i\theta}}(\tfrac{\partial}{\partial t})=((\cos t)x,-\sin te^{i\theta}),\quad d\phi_{t,x,e^{i\theta}}(H^n)=(H^n_{(\sin t)x},0),$$
$$d\phi_{t,x,e^{i\theta}}(\tfrac{\partial}{\partial\theta})=(0,\tfrac{\partial}{\partial\theta}|_{(\cos t)e^{i\theta}}).$$

In fact, (7) can be defined on all of $[0,\pi/2]\times S^{2n-1}\times S^1$, making it surjective, but we will work on the open set where it defines a parametrization. The round metric pulls back to

$$\phi^* g_{S^{2n+1}} = dt^2 + \sin^2 t g_{S^{2n-1}} + \cos^2 t d\theta^2.$$

Assume $n>1$. At each point of S^{2n-1}, we can decompose the tangent space as $(H^n)^\perp \oplus \operatorname{Span}\{H^n\}$; accordingly, the round metric decomposes as

$$g_{S^{2n-1}} = g^\perp + (H^n)^\flat \otimes (H^n)^\flat.$$

Thus, the round sphere on S^{2n+1} reads

$$\phi^* g_{S^{2n+1}} = dt^2 + \sin^2 t g^\perp + \sin^2 t (H^n)^\flat \otimes (H^n)^\flat + \cos^2 t d\theta^2. \tag{8}$$

Identifying tangent vectors through (7), we can write

$$H^{n+1} = H^n + \tfrac{\partial}{\partial\theta}.$$

Notice that $\{H^{n+1}, \frac{\cos t}{\sin t}H^n - \frac{\sin t}{\cos t}\frac{\partial}{\partial\theta}\}$ is an orthonormal basis of $\operatorname{Span}\left\{H^{n+1}, \frac{\partial}{\partial\theta}\right\}$, and the dual basis is $\{(H^{n+1})^\flat, \sin t\cos t((H^n)^\flat - d\theta))\}$

Thus, we can rewrite (8) as

$$\phi^* g_{S^{2n+1}} = dt^2 + \sin^2 t g^\perp + (H^{n+1})^\flat \otimes (H^{n+1})^\flat + \sin^2 t\cos^2 t((H^n)^\flat - d\theta)^2.$$

Projecting (8) on $(H^{n+1})^\perp$ yields

$$dt^2 + \sin^2 t g^\perp + \sin^2 t\cos^2 t((H^n)^\flat - d\theta)^2.$$

Now consider the diagram

$$\begin{tikzcd}
(t,x,e^{i\theta}) \arrow[d] & \in & (0,\pi/2)\times S^{2n-1}\times S^1 \arrow[r] \arrow[d] & \mathbb{CP}^n \\
(t,e^{-i\theta}x) & \in & (0,\pi/2)\times S^{2n-1} \arrow[ur] &
\end{tikzcd}$$

Then the metric on $(0,\pi/2)\times S^{2n-1}$ is

$$dt^2 + \sin^2 t g^\perp + \sin^2 t\cos^2 t((H^n)^\flat \otimes (H^n)^\flat).$$

If $n=1$, $g^\perp$ does not appear, and we obtain

$$dt^2 + \sin^2 t\cos^2 t d\gamma^2 = \frac{1}{4}(d\rho^2 + \sin^2\rho d\gamma^2), \quad \rho = 2t.$$

This is the local form of the metric on the sphere of radius 1/2 (see also Example 5.7).

Now fix $n=2$. Then $S^3 = \mathrm{SU}(2)$ and $H^2 = \underline{\mathrm{i}}$, so the Fubini-Study metric on $\mathbb{CP}^2$ has an orthonormal frame of the form

$$e_0, e_1, e_2, e_3 = \tfrac{\partial}{\partial t}, \frac{1}{\sin t\cos t}\underline{\mathrm{i}}, \frac{1}{\sin t}\underline{\mathrm{j}}, \frac{1}{\sin t}\underline{\mathrm{k}}.$$

The nonzero Lie brackets are

$$
\begin{aligned}
[e_0, e_1] &= -\frac{4\cos 2t}{\sin^2 2t}\mathrm{i} = -2(\cot 2t)e_1, \\
[e_0, e_2] &= -\frac{\cos t}{\sin^2 t}\mathrm{i} = -(\cot t)e_2, \\
[e_0, e_3] &= -(\cot t)e_3, \quad [e_1, e_2] = \frac{4}{\sin 2t}e_3, \\
[e_2, e_3] &= 2(\cot t)e_1, \quad [e_3, e_1] = \frac{4}{\sin 2t}e_2,
\end{aligned}
$$

and the dual coframe satisfies

$$de^0 = 0, \quad de^1 = 2\cot 2te^{01} - 2\cot te^{23}, de^2 = \cot te^{02} + 4/\sin 2te^{13}, de^3 = \cot te^{03} - 4/\sin 2te^{12}.$$

The Koszul formula for vector fields X, Y, Z in an orthonormal frame gives

$$2g(\nabla_X Y, Z) = g([X,Y],Z) + g([Z,X],Y) - g([Y,Z],X),$$

showing that $\nabla_{e_i} e_j$ has a component along e_k only if up to permutation $\{i,j,k\}$ is one of $\{0,1,1\}$, $\{0,2,2\}$, $\{0,3,3\}$, $\{1,2,3\}$. Thus, the connection form is

$$\omega = \begin{pmatrix} 0 & -2\cot 2te^1 & -\cot te^2 & -\cot te^3 \\ 2\cot 2te^1 & 0 & -\cot te^3 & \cot te^2 \\ \cot te^2 & \cot te^3 & 0 & (\cot t - 4/\sin 2t)e^1 \\ \cot te^3 & -\cot te^2 & (-\cot t + 4/\sin 2t)e^1 & 0 \end{pmatrix}$$

It is sufficient to compute curvature at $t = \pi/4$ because the group SU(3) acts on $\mathbb{CP}^2$ transitively by isometries. Thus, at $t = \pi/4$, and using the short-hand notation e^{ij} for $e^i \wedge e^j$, we obtain

$$d\omega = \begin{pmatrix} 0 & 4e^{01} & e^{02}-4e^{13} & e^{03}+4e^{12} \\ * & 0 & e^{03}+4e^{12} & -e^{02}+4e^{13} \\ * & * & 0 & -2e^0\wedge e^1+6e^{23} \\ * & * & * & 0 \end{pmatrix}, \quad \omega\wedge\omega = \begin{pmatrix} 0 & -2e^{23} & 3e^{13} & -3e^{12} \\ & 0 & -3e^{12} & -3e^{13} \\ * & * & 0 & -2e^{23} \\ & * & * & 0 \end{pmatrix},$$

$$R = \begin{pmatrix} 0 & 4e^0\wedge e^1-2e^2\wedge e^3 & e^0\wedge e^2-e^1\wedge e^3 & e^0\wedge e^3+e^1\wedge e^2 \\ -4e^0\wedge e^1+2e^2\wedge e^3 & 0 & e^0\wedge e^3+e^1\wedge e^2 & -e^0\wedge e^2+e^1\wedge e^3 \\ -e^0\wedge e^2+e^1\wedge e^3 & -e^0\wedge e^3-e^1\wedge e^2 & 0 & -2e^0\wedge e^1+4e^{23} \\ -e^0\wedge e^3-e^1\wedge e^2 & e^0\wedge e^2-e^1\wedge e^3 & 2e^0\wedge e^1-4e^2\wedge e^3 & 0 \end{pmatrix}.$$

In particular, we see that Ricci curvature is

$$\mathrm{ric} = 6(e^0\otimes e^0 + e^1\otimes e^1 + e^2\otimes e^2 + e^3\otimes e^3),$$

and the metric is Einstein.

To compute sectional curvature, observe that the curvature at a point is invariant under the action of the stabilizer $U(2)$. Since $U(2)$ acts transitively on $\mathbb{R}^4$, up to the action of $U(2)$, any plane is contained in the orthogonal of e_0. Thus, in order to compute the sectional curvature, it suffices to do so at a point, and take planes orthogonal to e_0.

The restriction of the Riemann tensor restricted to Span $\{e_1, e_2, e_3\}$ is given by

$$-e^1\wedge e^2\otimes(e^1\otimes e_2 - e^2\otimes e_1) - e^{13}\otimes(e^1\otimes e_3 - e^3\otimes e_1) - 4e^{23}\otimes(e^2\otimes e_3 - e^3\otimes e_2).$$

This shows that the sectional curvature satisfies

$$1 \leq \sec \leq 4.$$

Remark 4.1. The above technique only works in dimension 4, but the conclusion can be extended to $\mathbb{CP}^n$ by observing that every two-plane $\pi \subset T_x\mathbb{CP}^n$ is tangent to a totally geodesic copy of $\mathbb{CP}^2$.

5 Covariant derivative along a curve and mixed derivatives

The purpose of this section is to introduce mixed derivatives of a function $f\colon \Omega \to M$, where Ω is an open set in $\mathbb{R}^n$.

To make things precise, one needs the notion of pullback connection. Given any smooth map $f\colon N \to M$, one has a pullback bundle f^*TM with fibres

$$(f^*TM)_p = T_{f(p)}M;$$

its sections are maps $s\colon N \to TM$ with $s(p) \in T_{f(p)}M$. Given a section X of TM, i.e. a vector field, one can pull it back to a section $f^*X = X \circ f$ of f^*TM. A local trivialization (i.e. a frame) of M, $e_1, \dots, e_n$, determines a local trivialization $f^*e_1, \dots, f^*e_n$ of f^*TM.

The pullback bundle f^*TM carries a pullback connection $\widetilde{\nabla}$, characterized by

$$\widetilde{\nabla}_v(f^*X) = f^*\nabla_{f_{*p}(v)}X, \quad v \in T_pN, X \in \mathcal{X}(M).$$

In other words, if $\{e_1, \dots, e_n\}$ is a frame on M and $\{\tilde{e}_1, \dots, \tilde{e}_n\}$ is the corresponding trivialization of f^*TM, then

$$\widetilde{\nabla}_v(\tilde{e}_j) = \omega^i_j(f_*v)\tilde{e}_i,$$

i.e. the connection form of $\widetilde{\nabla}$ is $\widetilde{\omega}^i_j = f^*\omega^i_j$.

If $x_1, \dots, x_n$ are coordinates about $f(p)$, the frame $\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}$ induces a local trivialization of f^*TM that one usually also denotes by $\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}$. Writing $f^i = x^i \circ f$ and given a section $s = s^i\frac{\partial}{\partial x_i}$ of f^*TM, we have

$$\begin{aligned}\widetilde{\nabla}_X(s^i\tfrac{\partial}{\partial x_i}) &= (Xs^i)\tfrac{\partial}{\partial x_i} + s^i\nabla_{f_*X}\tfrac{\partial}{\partial x_i} = (Xs^i)\tfrac{\partial}{\partial x_i} + s^i(Xf^j)\nabla_{\frac{\partial}{\partial x_j}}\tfrac{\partial}{\partial x_i}\\ &= (Xs^i)\tfrac{\partial}{\partial x_i} + (f^*\Gamma^k_{ji})s^i(Xf^j)\tfrac{\partial}{\partial x_k}. \quad (9)\end{aligned}$$

Remark 5.1. The curvature of the pullback connection is just the pullback of the curvature. This is because denoting the connection form of ∇ by (ω^i_j), the connection form of $\widetilde{\nabla}$ is $(f^*\omega^i_j)$, and therefore

$$f^*R = f^*(d\omega^i_j + \omega^i_k \wedge \omega^k_j) = d(f^*\omega^i_j) + (f^*\omega^i_k) \wedge (f^*\omega^k_j).$$

The pullback connection can be used to define mixed derivatives of a map $\gamma\colon \Omega \to M$, $\Omega \subset \mathbb{R}^m$. This will be useful in particular for $m = 2$; then $\gamma(s,t)$ can be viewed as a one-parameter family of curves in M.

First derivatives can be expressed solely in terms of the differential:

$$\frac{\partial\gamma}{\partial t_i}\colon \Omega \to TM, \quad \frac{\partial\gamma}{\partial t_i} = d\gamma(\tfrac{\partial}{\partial t_i}) = \frac{\partial\gamma^j}{\partial t_i}\tfrac{\partial}{\partial x_j}.$$

We can also think of $\frac{\partial\gamma}{\partial t}$ as a section of γ^*TM. This enables one to *define* its mixed derivative:

$$\frac{\partial^2\gamma}{\partial t_i\partial t_j} = \widetilde{\nabla}_{\frac{\partial}{\partial t_i}}\frac{\partial\gamma}{\partial t_j}.$$

Writing t, s instead of t_i and t_j and using (9), we have

$$\frac{\partial^2\gamma}{\partial s\partial t} = \widetilde{\nabla}_{\frac{\partial}{\partial s}}(\frac{\partial\gamma^i}{\partial t}\frac{\partial}{\partial x_i}) = \frac{\partial^2\gamma^i}{\partial s\partial t}\frac{\partial}{\partial x_i} + \frac{\partial\gamma^i}{\partial t}\frac{\partial\gamma^j}{\partial s}\Gamma^k_{ji}\frac{\partial}{\partial x_k}.$$

Remark 5.2. Choosing $t = t_i = t_j$ instead, we find

$$\frac{\partial^2 \gamma}{\partial t^2} = \frac{\partial^2 \gamma^i}{\partial t^2}\frac{\partial}{\partial x_i} + \frac{\partial \gamma^i}{\partial t}\frac{\partial \gamma^j}{\partial t}\Gamma^k_{ji}\frac{\partial}{\partial x_k}.$$

Proposition 5.3. *Let (M, g) be a Riemannian manifold; let $\Omega \subset \mathbb{R}^n$ be an open set, and $\gamma\colon \Omega \to M$ a smooth map. Then*

$$\frac{\partial^2 \gamma}{\partial t_i \partial t_j} = \frac{\partial^2 \gamma}{\partial t_j \partial t_i}$$

$$\frac{\partial}{\partial t_k} g(\frac{\partial \gamma}{\partial t_i}, \frac{\partial \gamma}{\partial t_j}) = g(\frac{\partial^2 \gamma}{\partial t_k \partial t_i}, \frac{\partial}{\partial t_j}) + g(\frac{\partial}{\partial t_i}, \frac{\partial^2 \gamma}{\partial t_k \partial t_j})$$

Proof. The first follows from the fact that γ^i is a function to $\mathbb{R}$, so its mixed derivatives commute, and Γ^k_{ji} is symmetric because ∇ is torsion-free.

The second follows from the fact that ∇ is Riemannian, and therefore

$$dg(\frac{\partial \gamma}{\partial t_i}, \frac{\partial \gamma}{\partial t_j}) = g(\widetilde{\nabla}\frac{\partial \gamma}{\partial t_i}, \frac{\partial \gamma}{\partial t_j}) + g(\frac{\partial \gamma}{\partial t_i}, \widetilde{\nabla}\frac{\partial \gamma}{\partial t_j})$$

□

We can define higher order partial derivatives in the same way. Indeed, we *define*

$$\frac{\partial \gamma^3}{\partial t_i \partial t_j \partial t_k} = \widetilde{\nabla}_{\frac{\partial}{\partial t_i}} \widetilde{\nabla}_{\frac{\partial}{\partial t_j}} \frac{\partial \gamma}{\partial t_k}.$$

Proposition 5.4. *Let (M, g) be a Riemannian manifold; let R be the curvature. Let $\Omega \subset \mathbb{R}^n$ be an open set, and $\gamma\colon \Omega \to M$ a smooth map. Then*

$$\frac{\partial^3 \gamma}{\partial t_i \partial t_j \partial t_k} - \frac{\partial^3 \gamma}{\partial t_j \partial t_i \partial t_k} = f^* R(\tfrac{\partial}{\partial t_j}, \tfrac{\partial}{\partial t_i})\frac{\partial \gamma}{\partial t_k}$$

Proof. Follows from

$$\widetilde{\nabla}_{\frac{\partial}{\partial t_i}} \widetilde{\nabla}_{\frac{\partial}{\partial t_j}} - \widetilde{\nabla}_{\frac{\partial}{\partial t_j}} \widetilde{\nabla}_{\frac{\partial}{\partial t_i}} - \widetilde{\nabla}_{[\frac{\partial}{\partial t_i}, \frac{\partial}{\partial t_j}]} = f^* R(\tfrac{\partial}{\partial t_i}, \tfrac{\partial}{\partial t_j}),$$

as $f^* R$ is the curvature. □

Given a curve $\gamma\colon I \to M$, a *vector field along* γ is a section of the vector bundle $\gamma^* TM$, or in other words a smooth map

$$V\colon I \to TM, \quad V(t) \in T_{\gamma(t)} M.$$

The *covariant derivative along* γ of V is defined as

$$D_t V = \widetilde{\nabla}_{\frac{\partial}{\partial t}} V.$$

If V is the velocity, i.e. $V = \frac{\partial \gamma}{\partial t}$, then

$$D_t \gamma' = \frac{\partial^2 \gamma}{\partial t^2};$$

we will write γ', γ'' instead of $\frac{\partial\gamma}{\partial t}$, $\frac{\partial^2\gamma}{\partial t^2}$.

A curve γ is called a *geodesic* if

$$\gamma''=0.$$

Given a piecewise smooth curve $\gamma\colon[a,b]\to M$, its length is given by

$$l(\gamma)=\int_a^b|\gamma'(t)|\,dt.$$

We denote by $\Omega_{p,q}$ the space of piecewise smooth curves $\gamma\colon[0,1]\to M$ with $\gamma(0)=p$, $\gamma(1)=q$. A Riemannian manifold is a metric space with $d(p,q)=\inf_{\gamma\in\Omega_{p,q}}l(\gamma)$.

The *exponential map* at $p\in M$ is a smooth map $\exp_p\colon U\to M$ from a radial open set $U\subset T_pM$ such that $\gamma(t)=\exp_p(tv)$ is the geodesic with $\gamma(0)=p$, $\gamma'(0)=v$. Restricting U one has that $\exp_p$ is a diffeomorphism; fixing an orthonormal basis of $e_1,\dots,e_n$ of T_pM, one obtains local coordinates on M, called normal coordinates. If U is a ball centered at 0, one says that $\exp_p(U)$ is a *geodesic ball.* On a geodesic ball, one can define a radial function r, smooth away from p, satisfying

$$r^2=x_1^2+\dots+x_n^2.$$

The level set $r^{-1}(k)$, which is the boundary of the closed geodesic ball of radius k, is called a geodesic sphere.

One also considers the vector field $\frac{\partial}{\partial r}$ defined by

$$r\frac{\partial}{\partial r}=x_1\frac{\partial}{\partial x_1}+\dots+x_n\frac{\partial}{\partial x_n};$$

this is a unit vector field, since it is the speed of a unit speed radial geodesic. We have:

Theorem 5.5 (Gauss' lemma)**.** *Each geodesic sphere $r^{-1}(k)$ is orthogonal to $\frac{\partial}{\partial r}$.*

Since the gradient of r is also orthogonal to a geodesic sphere, observing that

$$g(\operatorname{grad}r,r\frac{\partial}{\partial r})=dr(r\frac{\partial}{\partial r})=r=g(\frac{\partial}{\partial r},r\frac{\partial}{\partial r}),$$

we see that $\frac{\partial}{\partial r}=\operatorname{grad}r$.

The Gauss Lemma also implies that, inside a geodesic ball, r can be identified with distance from p.

Remark 5.6. More generally, if r is any smooth function on an open set $U\subset M$ such that $|\operatorname{grad}r|=1$ one sees that level surfaces are smooth and orthogonal to $\frac{\partial}{\partial r}$. Such a function is called a *distance function* (see Section 11). An example is the distance from a compact submanifold in a tubular neighbourhood.

Example 5.7. Consider the exponential map on the sphere of radius $1/\sqrt{k}$. Then for unit v in T_pM,

$$\exp_p(tv)=\frac{1}{\sqrt{k}}\sin(\sqrt{k}\,t)v+\cos(\sqrt{k}\,t)p.$$

If $|v| = r$, one obtains

$$\exp_p(v) = \frac{1}{\sqrt{k}\,r} \sin(\sqrt{k}\,r)v + \cos(\sqrt{k}\,r)p.$$

If $w \in T_v(T_pM)$ is orthogonal to v, an explicit computation shows that

$$\exp_{p*}(v; w) = \frac{1}{\sqrt{k}\,r} \sin(\sqrt{k}\,r)w;$$

thus in normal coordinates

$$g = \frac{1}{kr^2} \sin^2(\sqrt{k}\,r)g^\perp + dr^2,$$

where $g^\perp$ is the Euclidean metric restricted to $\frac{\partial}{\partial r}^\perp$. If we identify the punctured geodesic ball with $S^{n-1} \times (0, \rho)$, $g^\perp = r^2 g_{S^{n-1}}$. Thus, we can write the round metric on the sphere as

$$\frac{1}{k} \sin^2(\sqrt{k}\,r)g_{S^{n-1}} + dr^2. \tag{10}$$

Theorem 5.8 (Hopf-Rinow)**.** *Let (M, g) be a connected Riemannian manifold. The following are equivalent:*

- *(M, g) is geodesically complete, i.e. geodesics are defined on all $\mathbb{R}$;*
- *For some $p \in M$, $\exp_p$ is defined on all T_pM;*
- *The Riemannian distance makes M into a complete metric space.*

Remark 5.9. It follows from the proof of the Hopf-Rinow theorem that, on a complete Riemannian manifold, every two points p, q are connected by a geodesic of length $d(p, q)$. However, this property does not characterize complete Riemannian manifolds; consider a convex set in $\mathbb{R}^n$.

Another useful notion is the *Hessian* of a function $f : M \to \mathbb{R}$. This is defined as $\operatorname{Hess} f = \nabla df$; in local coordinates,

$$\operatorname{Hess} f = \nabla(\frac{\partial f}{\partial x_j} dx^j) = \frac{\partial^2 f}{\partial x_j \partial x_k} dx^k \otimes dx^j - \Gamma^j_{kh} \frac{\partial f}{\partial x_j} dx^k \otimes dx^h.$$

where we have used

$$\nabla_{\frac{\partial}{\partial x_k}} dx^i = -\Gamma^i_{kh} dx^h.$$

This makes it evident that $\operatorname{Hess} f$ is a symmetric tensor, as can also be deduced from the fact that the skew-symmetric part of ∇df is $d^2 f = 0$.

Notice that since covariant derivative commutes with musical isomorphisms, we have $(\nabla_X df)^\sharp = \nabla_X (df)^\sharp = \nabla_X \operatorname{grad} f$, and therefore

$$\operatorname{Hess} f(X, Y) = g(\nabla_X \operatorname{grad} f, Y)$$

The following will be useful:

Lemma 5.10. *Given a geodesic $\gamma = \exp_p tv$, we have*

$$\operatorname{Hess} f(v, v) = \frac{d^2}{dt^2}|_{t=0}(f(\gamma(t))).$$

Proof. The pullback connection commutes with contractions: if α is a section of γ^*T^*M and X a section of γ^*TM, we have

$$\big(\widetilde{\nabla}_{\frac{\partial}{\partial t}}\alpha\big)(X) = \frac{d}{dt}(\alpha(X)) - \alpha\big(\widetilde{\nabla}_{\frac{\partial}{\partial t}}X\big).$$

Now let $X = \gamma'$ and $\alpha = df$, viewed as a section of γ^*T^*M. Then $\alpha(X)$ is the function $(f\circ\gamma)'$.

If $\gamma'(0) = v$, we have

$$\begin{aligned}\operatorname{Hess} f(v,v) = (\nabla_v df)(v) = (\widetilde{\nabla}_{\frac{\partial}{\partial t}}\alpha)(X) &= \frac{d}{dt}|_{t=0}\alpha(X) - \alpha(\widetilde{\nabla}_{\frac{\partial}{\partial t}}X)\\ &= \frac{d^2}{dt^2}|_{t=0} f(\gamma(t)) - \alpha(\gamma'') = \frac{d^2}{dt^2}|_{t=0} f(\gamma(t)).\end{aligned}$$

□

Example 5.11. Consider a two-dimensional M. In polar normal coordinates the metric takes the local form $dr^2 + \rho(r,\theta)^2 d\theta^2$.

The Levi-Civita connection satisfies

$$2g(\nabla_X Y, Z) = Xg(Y,Z) + Yg(X,Z) - Zg(X,Y),$$

whenever X, Y, Z are coordinate vector fields. Hence, the only nonzero Christoffel symbols arise when at least two of the indices correspond to θ, i.e.

$$g(\nabla_{\frac{\partial}{\partial r}}\tfrac{\partial}{\partial\theta},\tfrac{\partial}{\partial\theta}) = \frac{1}{2}\frac{\partial\rho^2}{\partial r},\quad g(\nabla_{\frac{\partial}{\partial\theta}}\tfrac{\partial}{\partial\theta},\tfrac{\partial}{\partial r}) = -\frac{1}{2}\frac{\partial\rho^2}{\partial r},\quad g(\nabla_{\frac{\partial}{\partial\theta}}\tfrac{\partial}{\partial\theta},\tfrac{\partial}{\partial\theta}) = \frac{1}{2}\frac{\partial\rho^2}{\partial\theta}$$

Summing up

$$\nabla_{\frac{\partial}{\partial r}}\frac{\partial}{\partial\theta} = \frac{1}{\rho}\frac{\partial\rho}{\partial r}\frac{\partial}{\partial\theta} = \nabla_{\frac{\partial}{\partial\theta}}\frac{\partial}{\partial r},\quad \nabla_{\frac{\partial}{\partial\theta}}\frac{\partial}{\partial\theta} = -\rho\frac{\partial\rho}{\partial r}\frac{\partial}{\partial r} + \frac{1}{\rho}\frac{\partial\rho}{\partial\theta}\frac{\partial}{\partial\theta}.$$

The curvature is then

$$R(\tfrac{\partial}{\partial\theta},\tfrac{\partial}{\partial r})\tfrac{\partial}{\partial r} = \nabla_{\frac{\partial}{\partial\theta}}\nabla_{\frac{\partial}{\partial r}}\tfrac{\partial}{\partial r} - \nabla_{\frac{\partial}{\partial r}}\nabla_{\frac{\partial}{\partial\theta}}\tfrac{\partial}{\partial r} = -\nabla_{\frac{\partial}{\partial r}}(\frac{1}{\rho}\frac{\partial\rho}{\partial r}\frac{\partial}{\partial\theta}) = -\frac{1}{\rho}\frac{\partial^2\rho}{\partial r^2}\frac{\partial}{\partial\theta},$$

so that the Gaussian (or sectional) curvature is $-\frac{1}{\rho}\frac{\partial^2\rho}{\partial r^2}$.

Additionally, observe that in this case

$$\nabla \operatorname{grad} r = \nabla\tfrac{\partial}{\partial r} = d\theta\otimes\frac{1}{\rho}\frac{\partial\rho}{\partial r}\frac{\partial}{\partial\theta},\quad \operatorname{Hess} r = \rho\frac{\partial\rho}{\partial r}d\theta\otimes d\theta.$$

Applying Lemma 5.10, we see that, locally, a geodesic which is orthogonal to the geodesic sphere $\{r = r_0\}$ lies in the exterior of the sphere if $\frac{\partial\rho}{\partial r}|_{r=r_0} < 0$ and in the interior if $\frac{\partial\rho}{\partial r}|_{r=r_0} < 0$.

Example 5.12. The round sphere (10) corresponds to $\rho = 1/\sqrt{k}\sin(\sqrt{k}\,r)$ in the notation of the previous example. Then it is clear that a geodesic orthogonal to a geodesic ball lies in its interior if the ball is large enough to contain a hemisphere, and lies in the exterior if the ball is contained in a hemisphere.

6 First and second variations of energy

Given a piecewise smooth curve $\gamma\colon [a,b] \to M$, one defines its *length* and *energy* as

$$l(\gamma) = \int_a^b |\gamma'(t)|\, dt, \quad E(\gamma) = \frac{1}{2}\int_a^b g(\gamma'(t), \gamma'(t))dt.$$

It turns out that geodesics minimize both length and energy; however, the energy functional has the advantage that its minimizers have constant speed.

Theorem 6.1. *Local minima of $E\colon \Omega_{p,q} \to [0, +\infty)$ are smooth geodesics.*

The proof relies on the *first variation formula*. One considers a *piecewise smooth variation*, i.e. a continuous map

$$\overline{\gamma}\colon (-\epsilon, \epsilon) \times [a,b] \to M;$$

such that for some $a_0 = 0, \dots, a_m = b$ we have that the restrictions to each $(-\epsilon,\epsilon)\times[a_i, a_{i+1}]$ are smooth. Then each curve $\gamma_s = \overline{\gamma}(s,\cdot)$ is piecewise smooth; we say that $\overline{\gamma}$ is a *variation* of $\gamma = \gamma_0$. The variation is *proper* if $\gamma_s(a)$, $\gamma_s(b)$ are independent of s, i.e. if each γ_s is in $\Omega_{p,q}$.

Lemma 6.2. *Given a piecewise smooth variation $\overline{\gamma}\colon (-\epsilon,\epsilon)\times[a,b] \to M$, we have*

$$\frac{\partial E(\gamma_s)}{\partial s} = -\int_a^b g\left(\frac{\partial^2\overline{\gamma}}{\partial t^2}, \frac{\partial\overline{\gamma}}{\partial s}\right)dt + g\left(\frac{\partial\overline{\gamma}}{\partial t}, \frac{\partial\overline{\gamma}}{\partial s}\right)|_{t=b} - g\left(\frac{\partial\overline{\gamma}}{\partial t}, \frac{\partial\overline{\gamma}}{\partial s}\right)|_{t=a} \\ + \sum_{i=1}^{m-1} g\left(\frac{\partial\overline{\gamma}}{\partial t}|_{t=a_i^-} - \frac{\partial\overline{\gamma}}{\partial t}|_{t=a_i^+}, \frac{\partial\gamma}{\partial s}|_{t=a_i}\right)$$

Proof. By additivity, it suffices to prove the formula for smooth $\overline{\gamma}$. We compute

$$\begin{aligned}\frac{\partial E(\gamma_s)}{\partial s} &= \frac{d}{ds}\frac{1}{2}\int_a^b g\left(\frac{\partial\overline{\gamma}}{\partial t}, \frac{\partial\overline{\gamma}}{\partial t}\right)dt = \frac{1}{2}\int_a^b \frac{\partial}{\partial s} g\left(\frac{\partial\overline{\gamma}}{\partial t}, \frac{\partial\overline{\gamma}}{\partial t}\right)dt \\ &= \int_a^b g\left(\frac{\partial^2\overline{\gamma}}{\partial s\partial t}, \frac{\partial\overline{\gamma}}{\partial t}\right)dt = \int_a^b g\left(\frac{\partial^2\overline{\gamma}}{\partial t\partial s}, \frac{\partial\overline{\gamma}}{\partial t}\right)dt \\ &= \int_a^b \frac{\partial}{\partial t} g\left(\frac{\partial\overline{\gamma}}{\partial s}, \frac{\partial\overline{\gamma}}{\partial t}\right)dt - \int_a^b g\left(\frac{\partial\overline{\gamma}}{\partial s}, \frac{\partial^2\overline{\gamma}}{\partial t^2}\right)dt \\ &= g\left(\frac{\partial\overline{\gamma}}{\partial s}, \frac{\partial\overline{\gamma}}{\partial t}\right)|_{t=b} - g\left(\frac{\partial\overline{\gamma}}{\partial s}, \frac{\partial\overline{\gamma}}{\partial t}\right)|_{t=a} - \int_a^b g\left(\frac{\partial\overline{\gamma}}{\partial s}, \frac{\partial^2\overline{\gamma}}{\partial t^2}\right)dt\end{aligned}$$

□

Proof of Theorem 6.1. At a local minima, the first variation of energy is zero.

Choose a function $\lambda\colon [a,b] \to \mathbb{R}$ which is zero at each $a = a_0, \dots, a_m = b$ and positive elsewhere, and consider the variation

$$\gamma_s(t) = \exp_{\gamma(t)} sV(t), \quad V(t) = \lambda(t)\gamma''(t).$$

In Lemma 6.2, we have that each term containing $\frac{\partial\overline{\gamma}}{\partial s}|_{t=a_i}$ vanishes; therefore, at $s = 0$ we find

$$0 = -\int_a^b g(\gamma''(t), V(t))dt = -\int_a^b g(\gamma''(t), \lambda(t)\gamma''(t))dt = -\lambda(t)\int_a^b g(\gamma''(t), \gamma''(t))dt \le 0;$$

it follows that $\gamma''(t) = 0$ where it is defined.

In order to show that γ is smooth, consider a variation of the form

$$\gamma_s(t) = \exp_{\gamma(t)} sV(t), \quad V(a_i) = \begin{cases} \frac{\partial\gamma}{\partial t}|_{t=a_i^-} - \frac{\partial\gamma}{\partial t}|_{t=a_i^+}, & 0 < i < m \\ 0, & a_i = a, b \end{cases}$$

Since we already know that $\gamma'' = 0$, Lemma 6.2 gives that

$$0 = \sum_{i=1}^{m-1} g\Big(\frac{\partial\overline{\gamma}}{\partial t}|_{t=a_i^-} - \frac{\partial\overline{\gamma}}{\partial t}|_{t=a_i^+}, \frac{\partial\gamma}{\partial t}|_{t=a_i^-} - \frac{\partial\gamma}{\partial t}|_{t=a_i^+}\Big),$$

so $\frac{\partial\gamma}{\partial t}|_{t=a_i^-} = \frac{\partial\gamma}{\partial t}|_{t=a_i^+}$ at each $i = 1, \dots, m-1$. □

Proposition 6.3. *Any curve $\gamma \in \Omega(p, q)$ minimizes energy if and only if it has constant speed and minimizes length.*

Proof. We have

$$l(\gamma) = \int_0^1 |\gamma'|\,dt = \int_0^1 |\gamma'|\,1dt \leq \big(\int_0^1 |\gamma'|^2\,dt\big)^{1/2}\big(\int_0^1 dt\big)^{1/2} = \sqrt{2E(\gamma)},$$

where we have used the Cauchy-Schwartz inequality in $L^2([0,1])$. If σ is a constant-speed curve that minimizes l, for every γ we have

$$E(\sigma) = \frac{1}{2}l(\sigma)^2 \leq \frac{1}{2}l(\gamma)^2 \leq E(\gamma),$$

so σ minimizes E.

Conversely, if σ minimizes E, it has constant speed by Theorem 6.1. To show it minimizes length, let γ be a curve in $\Omega_{p,q}$. Up to reparametrization, we can assume that γ also has constant speed. Therefore

$$l(\sigma) = \sqrt{2E(\sigma)} \leq \sqrt{2E(\gamma)} = l(\gamma).$$ □

A curve γ in $\Omega(p, q)$ is a *segment* if $|\gamma'|$ is constant where defined and minimizes l. Theorem 6.1 implies that piecewise smooth segments are smooth geodesics: as they minimize length, they also minimize energy. I will use "segment" and "minimizing geodesic" interchangeably.

Theorem 6.4 (Synge)**.** *Let $\overline{\gamma}\colon (-\epsilon, \epsilon) \times [a, b] \to M$ be a smooth variation of a geodesic γ. Then*

$$\frac{d^2E(\gamma_s)}{ds^2}|_{s=0} = \int_a^b \left|\frac{\partial^2\overline{\gamma}}{\partial t\partial s}\right|^2 dt - \int_a^b g\big(R(\frac{\partial\overline{\gamma}}{\partial s}, \frac{\partial\overline{\gamma}}{\partial t})\frac{\partial\overline{\gamma}}{\partial t}, \frac{\partial\overline{\gamma}}{\partial s}\big)dt + g\big(\frac{\partial^2\overline{\gamma}}{\partial s^2}, \frac{\partial\overline{\gamma}}{\partial t}\big)|_a^b.$$

Proof. It is convenient to add a parameter and consider a "two-parameter variation" of γ, i.e. a function

$$\alpha\colon (-\delta, \delta) \times (-\delta, \delta) \times [a, b] \to M, \quad (u, s, t) \mapsto \alpha(u, s, t),$$

where, writing $\alpha_{u,s} = \alpha(u, s, \cdot)$, we have $\alpha_{0,0} = \gamma$. Specifically, we will set $\alpha(u, s, t) = \gamma(u + s, t)$.

The first variation formula gives

$$\frac{\partial E(\alpha_{u,s})}{\partial s} = -\int_a^b g\left(\frac{\partial^2\alpha}{\partial t^2}, \frac{\partial\alpha}{\partial s}\right)dt + g(\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial s})|_{t=b} - g(\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial s})|_{t=a}$$

Taking the derivative with respect to u, we find

$$\frac{\partial^2 E(\alpha_{u,s})}{\partial u\partial s} = -\tfrac{\partial}{\partial u}\int_a^b g\left(\frac{\partial^2\alpha}{\partial t^2}, \frac{\partial\alpha}{\partial s}\right)dt + \tfrac{\partial}{\partial u}g(\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial s})|_{t=b} - \tfrac{\partial}{\partial u}g(\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial s})|_{t=a}. \tag{11}$$

Assuming that $\alpha(0,0,\cdot)$ is a geodesic and computing at $u=0=s$, the first term of (11) gives

$$\begin{aligned}
&-\tfrac{\partial}{\partial u}\int_a^b g\left(\frac{\partial^2\alpha}{\partial t^2}, \frac{\partial\alpha}{\partial s}\right)dt \\
&\qquad = -\int_a^b g\left(\frac{\partial^3\alpha}{\partial u\partial t^2}, \frac{\partial\alpha}{\partial s}\right)dt - \int_a^b g\left(\frac{\partial^2\alpha}{\partial t^2}, \frac{\partial^2\alpha}{\partial u\partial s}\right)dt \\
&\qquad = -\int_a^b g\left(R(\tfrac{\partial}{\partial u}\tfrac{\partial}{\partial t})\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial s}\right)dt - \int_a^b g\left(\frac{\partial^3\alpha}{\partial t\partial u\partial t}, \frac{\partial\alpha}{\partial s}\right)dt \\
&= -\int_a^b g\left(R(\tfrac{\partial}{\partial u}\tfrac{\partial}{\partial t})\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial s}\right)dt + \int_a^b g\left(\frac{\partial^2\alpha}{\partial u\partial t}, \frac{\partial^2\alpha}{\partial t\partial s}\right)dt - g\left(\frac{\partial^2\alpha}{\partial u\partial t}, \frac{\partial\alpha}{\partial s}\right)|_a^b
\end{aligned}$$

Still working at $u=0=s$, the last two terms of (11) give

$$\tfrac{\partial}{\partial u}g(\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial s})|_a^b = g(\frac{\partial^2\alpha}{\partial u\partial t}, \frac{\partial\alpha}{\partial s})|_a^b + g(\frac{\partial\alpha}{\partial t}, \frac{\partial^2\alpha}{\partial u\partial s})|_a^b =$$

Taking the sum, we find

$$\frac{\partial^2 E(\alpha_{u,s})}{\partial u\partial s} = -\int_a^b g\left(R(\tfrac{\partial}{\partial u}\tfrac{\partial}{\partial t})\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial s}\right)dt + \int_a^b g\left(\frac{\partial^2\alpha}{\partial u\partial t}, \frac{\partial^2\alpha}{\partial t\partial s}\right)dt + g(\frac{\partial\alpha}{\partial t}, \frac{\partial^2\alpha}{\partial u\partial s})|_a^b$$

Specializing to $\alpha(u,s,t) = \gamma(u+s,t)$, we obtain the statement. □

Example 6.5. Consider a surface of revolution S in $\mathbb{R}^3$, say

$$\{x^2 + y^2 = f(z)\}.$$

Then if $f'(z) = 0$, the curve

$$\gamma[0,2\pi] \to S, \quad t \mapsto (\cos t f(z), \sin t f(z), z)).$$

is a geodesic. It is only a local minimizer of energy in $\Omega_{p,q}$, where $p = q = \gamma(0)$.

Let V be the vector field along γ obtained by projecting $\frac{\partial}{\partial z}$ on S, i.e.

$$V(t) = \tfrac{\partial}{\partial z} + \cos t f'(z)\tfrac{\partial}{\partial x} + \sin t f'(z)\tfrac{\partial}{\partial y}$$

The curves in the variation

$$\overline{\gamma}(s,t) = \exp_{\gamma(t)} s \sin\frac{1}{2}tV(t)$$

are shorter or longer than γ depending on the sign of $f''(z)$, i.e. the curvature of the surface.

As a first consequence of this formula, we find a bound on the length of segments on manifolds with positive Ricci tensor. The first versions of this theorem were proved by Bonnet and Synge using sectional curvature.

Theorem 6.6 (Myers 1941)**.** *Let (M, g) be a Riemannian manifold of dimension n. Let $\gamma \colon [0, b] \to M$ be a unit-speed geodesic such that*

$$\operatorname{ric}(\gamma', \gamma') \geq (n-1)k > 0, \quad 0 \leq t \leq b.$$

If $b > \pi/\sqrt{k}$, then γ is not minimizing.

Proof. Let $e_1, \ldots, e_n$ be a basis of vector fields along γ, orthonormal and parallel; we can assume $\gamma' = e_n$. For $j < n$, let

$$X_j(t) = \sin(\pi t/b) e_j(t).$$

To each X_j, we can associate a variation

$$\overline{\gamma}^j(s, t) = \exp_{\gamma(t)} sX_j(t).$$

This is a proper variation, so Theorem 6.4 gives

$$\begin{aligned}\frac{d}{d^2s}|_{s=0} E(\overline{\gamma}^j_s) &= \int_0^b \left|X_j'(t)\right|^2 dt - \int_0^b \operatorname{Rm}(X_j, \gamma', \gamma', X_j) dt \\ &= \int_0^b \big(\frac{\pi^2}{b^2}\cos^2(\pi t/b) - \sin^2(\pi t/b)\operatorname{Rm}(e_j, \gamma', \gamma', e_j)\big) \\ &= \int_0^b \big(\frac{\pi^2}{b^2}\cos^2(\pi t/b) - \sin^2(\pi t/b)\sec(e_j, \gamma')\big).\end{aligned}$$

If we had a lower bound on sectional curvature, we could show that this quantity is negative, giving that the real function $s \mapsto E(\overline{\gamma}^j_s)$ has a local maximum, so that γ is not energy minimizing. Since we only have a bound on Ricci curvature, we proceed as follows.

Assume that γ is not minimizing. Then by the above argument each term $\frac{d}{d^2s}|_{s=0} E(\overline{\gamma}^j_s)$ is nonnegative; taking the sum over j, we obtain

$$\begin{aligned}0 \leq \sum_{j=2}^{n} \frac{d}{d^2s}|_{s=0} E(\overline{\gamma}^j_s) &= \int_0^b (n-1)(\frac{\pi^2}{b^2}\cos^2(\pi t/b)) - \sin^2(\pi t/b)\operatorname{Rm}(e_j, \gamma', \gamma', e_j)) dt \\ &= \int_0^b (n-1)(\frac{\pi^2}{b^2}\cos^2(\pi t/b)) - \sin^2(\pi t/b)\operatorname{ric}(\gamma', \gamma')) dt \\ &\leq \int_0^b (n-1)(\frac{\pi^2}{b^2}\cos^2(\pi t/b)) - \sin^2(\pi t/b)(n-1)k) dt \\ &= \frac{b}{2}(n-1)(\frac{\pi^2}{b^2} - k) < 0,\end{aligned}$$

which is absurd. □

Example 6.7. For S^n, we have that $\sec = 1$, so we recover the known fact that there are no minimizing geodesics of length $> \pi$.

Corollary 6.8. *Let (M, g) be a complete Riemannian manifold of dimension $n \geq 2$ such that*

$$\mathrm{ric}(\xi, \xi) \geq (n-1)k\,|\xi|^2\,, \quad \xi \in T_xM;$$

then M is compact with diameter $\leq \pi/\sqrt{k}$.

In addition, the universal covering of M is compact and $\pi_1(M)$ is finite.

Proof. By the Hopf-Rinow theorem, every pairs of point is joined by a minimizing geodesic $\gamma\colon [a, b] \to M$. Suppose γ has unit speed; then it satisfies the hypotheses of Theorem 6.6, so its length cannot exceed $\pi/\sqrt{k}$. This implies that the diameter is bounded by $\pi/\sqrt{k}$. Now fix any p. By completeness, $\exp_p$ is defined on all of T_pM; by the above diameter estimate, we have that $\exp_p\left(\overline{B_{\pi/\sqrt{k}}(0)}\right)$ equals M, which is therefore compact.

The universal covering $\tilde{M}$ has a pullback metric $\tilde{g}$: if $q\colon \tilde{M} \to M$ is the projection, $\tilde{g}_p(v, w) = g(dq_p(v), dq_p(w))$. Since q is a local isometry, the Ricci operator of $\tilde{g}$ is the pullback of the Ricci operator of g, so it satisfies the same inequality, and $\tilde{M}$ must be compact.

Now observe that any fiber $q^{-1}(p)$ is both discrete (because q is a covering map) and compact (because it is closed). This implies that $q^{-1}(p)$ is finite. Since $\pi_1(M)$ acts simply transitively on the fibre, it must also be finite. □

7 Jacobi fields [7]

Consider a variation of a geodesic through geodesics $\overline{\gamma}(s, t)$, i.e. such that each γ_s is a geodesic. By Proposition (5.4),

$$0 = \frac{\partial^3 \overline{\gamma}}{\partial s \partial t^2} = \overline{\gamma}^* R(\frac{\partial \overline{\gamma}}{\partial s}, \frac{\partial \overline{\gamma}}{\partial t})\frac{\partial \overline{\gamma}}{\partial t} + \frac{\partial^3 \overline{\gamma}}{\partial t \partial s \partial t} = \overline{\gamma}^* R(\frac{\partial \overline{\gamma}}{\partial s}, \frac{\partial \overline{\gamma}}{\partial t})\frac{\partial \overline{\gamma}}{\partial t} + \frac{\partial^3 \overline{\gamma}}{\partial t^2 \partial s}.$$

Thus, if $J(t)$ is the vector field along γ_0 given by

$$J = \frac{\partial \overline{\gamma}}{\partial s}|_{s=0},$$

we see that

$$0 = J'' + R(J, \gamma')\gamma'. \tag{12}$$

A vector field along a geodesic satisfying (12) is called a *Jacobi field*. The space of Jacobi fields along a fixed geodesic will be denoted by $\mathcal{J}$.

Equation 12 is a second order ODE, so relative to a fixed geodesic γ, with $\gamma(0) = p$, we have a one-to-one-correspondence

$$\mathcal{J} \leftrightarrow T_pM \oplus T_pM, \quad J \mapsto (J(0), J'(0)). \tag{13}$$

In particular, the space of Jacobi fields along γ is a real vector space of dimension $2n$, where n is the dimension of M.

Given a geodesic γ, a variation through geodesics with $\gamma_s(0) = \gamma(0)$ can be constructed as

$$\overline{\gamma}(s, t) = \exp_p(t(\gamma'(0) + sw)). \tag{14}$$

The associated Jacobi field satisfies $J(0) = 0$ and

$$\frac{\partial J}{\partial t}|_{t=0} = \frac{\partial^2 \overline{\gamma}}{\partial t \partial s}|_{t=0,s=0} = \frac{\partial^2 \overline{\gamma}}{\partial s \partial t}|_{s=0,t=0} = \frac{\partial}{\partial s}|_{s=0}(\gamma'(0) + sw) = w.$$

Notice that J is uniquely determined by the conditions $J(0) = 0$, $J'(0) = w$. This leads to the following:

Lemma 7.1. *If γ is a geodesic with $\gamma(0) = p$, $\gamma'(0) = v$ and J is the Jacobi field along γ with $J(0) = 0$, $J'(0) = w$, then*

$$d\exp_p(v; w) = J(1).$$

Proof. Defining $\overline{\gamma}$ as in (14), we compute

$$J(t) = \frac{\partial \overline{\gamma}}{\partial s}|_{s=0} = d\exp_p(t\gamma'(0); tw).$$

The statement is obtained by setting $t = 1$. □

This result is useful per se, but also has the following consequence:

Proposition 7.2. *Suppose $\exp_p$ is nonsingular at $v \in T_pM$ and let $\gamma(t) = \exp_p tv$, $q = \gamma(1)$. For every $u \in T_qM$, there is a Jacobi field along γ with $J(0) = 0$ and $J(1) = u$.*

Remark 7.3. The vector fields γ' and $t\gamma'$ are always Jacobi fields; they are associated to the geodesic variations

$$\gamma_s(t) = \gamma(t + s), \quad \gamma_s(t) = \gamma((1 + s)t).$$

We define

$$\mathcal{J}^{\perp} = \{X \in \mathcal{J}, \quad g(X, \gamma') \equiv 0\}.$$

Proposition 7.4. *For a fixed geodesic γ, we have a decomposition*

$$\mathcal{J} = \mathcal{J}^{\perp} \oplus \operatorname{Span}\{\gamma', t\gamma'\}.$$

Proof. In light of (13) we need to show that every Jacobi field with $J(0)$ and $J'(0)$ orthogonal to γ' is in $\mathcal{J}^{\perp}$.

We first prove that for any X and Y in $\mathcal{J}$,

$$g(X', Y) - g(X, Y') = \text{constant}. \tag{15}$$

This follows from

$$\begin{aligned}\frac{d}{dt}(g(X', Y) - g(X, Y')) &= g(X'', Y) - g(X, Y'') \\ &= -\operatorname{Rm}(X, \gamma', \gamma', Y) + \operatorname{Rm}(Y, \gamma', \gamma', X) = 0.\end{aligned}$$

Now if X is a vector field with $X(0)$ and $X'(0)$ orthogonal to $\gamma'(0)$, plugging $Y = \gamma'$ in (15) gives

$$g(X'(t), \gamma'(t)) = g(X'(0), \gamma'(0)) = 0,$$

so $g(X, \gamma')' = g(X', \gamma') = 0$ and

$$g(X(t), \gamma'(t)) = g(X(0), \gamma'(0)) = 0.$$

□

In particular, any Jacobi field can be decomposed as

$$Y = Y^{\perp} + (h + kt)\gamma', \quad h, k \in \mathbb{R}, Y^{\perp} \in \mathcal{J}^{\perp}. \tag{16}$$

Given $t \in (a, b)$, we shall say that $\gamma(t)$ is a *conjugate point* of $\gamma(a)$ if there is a nonzero Jacobi field which vanishes at $t = a, t = b$. It follows from (16) that Y is in $\mathcal{J}^{\perp}$.

Remark 7.5. In the case of constant curvature $K \leq 0$, $\gamma(a)$ has no conjugate points. If $K > 0$, conjugate points are exactly $\gamma(h\pi/\sqrt{k})$, $h \in \mathbb{N}$.

Remark 7.6. If no geodesic $\gamma(t) = \exp_p tv$ has conjugate points, then $\exp_p$ is a local diffeomorphism where it is defined.

Example 7.7. Consider a metric of constant sectional curvature k, i.e.

$$R(X, Y)Z = k\big(-g\,(X, Z)\,Y + g\,(Y, Z)\,X\big).$$

Then the Jacobi equation reads

$$0 = J'' + R(J, \gamma')\gamma' = J'' - kg\,(J, \gamma')\,\gamma' + k\,|\gamma'|^2\,J$$

Let $\{e_i\}$ be a parallel frame along γ, with $e_n = \gamma$. Then a Jacobi field along γ ort takes the form $Y^i e_i(t)$, where

$$(Y^i)'' + kY^i = 0, i < n, \quad Y'' = 0.$$

Therefore Jacobi fields along γ take the form

$$J = C_k(t)A(t) + S_k(t)B(t) + \alpha\gamma' + t\beta\gamma',$$

where A, B are parallel vector fields along γ, both orthogonal to γ', α, β are constants, and S_k, C_k are solutions of

$$\psi'' + k\psi = 0, \quad S_k(0) = 0, S_k'(0) = 1, \quad C_k(0) = 1, C_k'(0) = 0.$$

Explicitly,

$$S_k(t) = \begin{cases} \frac{1}{\sqrt{k}} \sin\sqrt{k}\,t & k > 0 \\ t & k = 0 \\ \frac{1}{\sqrt{-k}} \sinh\sqrt{-k}\,t & k < 0 \end{cases}, \quad C_k(t) = \begin{cases} F\cos\sqrt{k}\,t & k > 0 \\ 1 & k = 0 \\ \cosh\sqrt{-k}\,t & k < 0 \end{cases}$$

Remark 7.8. The fact that Jacobi fields take the form of Example 7.7 characterizes metrics of constant curvature. Knowing what the Jacobi fields are, we can recover the form of the metric in normal coordinates as follows. Let $\gamma(t) = \exp_p tv$ be a geodesic. Let J be the Jacobi field along $\exp_p tv$ with $J(0) = 0, J'(0) = \sum a_i \frac{\partial}{\partial x_i}$. Assume $J'(0)$ is orthogonal to v; then $J(t) = \frac{1}{|v|}\sum S_k(t\,|v|)a_i\frac{\partial}{\partial x_i}$. Thus,

$$\left|a_i\frac{\partial}{\partial x_i}\right|^2_{\exp_p v} = |J(1)|^2 = \frac{1}{|v|^2}S_k(|v|)^2\sum a_i^2.$$

In other words, in normal coordinates we have

$$\frac{1}{r^2}S_k(r)^2 g^{\perp} + dr^2,$$

where $g^\perp$ is the Euclidean metric restricted to $\frac{\partial}{\partial r}^\perp$. If we identify the punctured geodesic ball with $S^{n-1} \times (0, \rho)$, $g^\perp = r^2 g_{S^{n-1}}$. Thus, we can write a constant curvature metric as

$$S_k(r)^2 g_{S^{n-1}} + dr^2.$$

Example 7.9. Consider a two-dimensional M. In polar normal coordinates, the metric takes the local form $dr^2 + \rho(r,\theta)^2 d\theta^2$. The Levi-Civita connection satisfies

$$2g(\nabla_X Y, Z) = Xg(Y,Z) + Yg(X,Z) - Zg(X,Y),$$

whenever X, Y, Z are coordinate vector fields. Hence, the only nonzero Christoffel symbols arise when at least two of the indices correspond to θ, i.e.

$$g(\nabla_{\frac{\partial}{\partial r}} \tfrac{\partial}{\partial \theta}, \tfrac{\partial}{\partial \theta}) = \frac{1}{2}\frac{\partial \rho^2}{\partial r}, \quad g(\nabla_{\frac{\partial}{\partial \theta}} \tfrac{\partial}{\partial \theta}, \tfrac{\partial}{\partial r}) = -\frac{1}{2}\frac{\partial \rho^2}{\partial r}, \quad g(\nabla_{\frac{\partial}{\partial \theta}} \tfrac{\partial}{\partial \theta}, \tfrac{\partial}{\partial \theta}) = \frac{1}{2}\frac{\partial \rho^2}{\partial \theta}$$

Summing up

$$\nabla_{\frac{\partial}{\partial r}} \frac{\partial}{\partial \theta} = \frac{1}{\rho}\frac{\partial \rho}{\partial r}\frac{\partial}{\partial \theta} = \nabla_{\frac{\partial}{\partial \theta}} \frac{\partial}{\partial r}, \quad \nabla_{\frac{\partial}{\partial \theta}} \frac{\partial}{\partial \theta} = -\rho \frac{\partial \rho}{\partial r}\frac{\partial}{\partial r} + \frac{1}{\rho}\frac{\partial \rho}{\partial \theta}\frac{\partial}{\partial \theta}.$$

The curvature is then

$$R(\tfrac{\partial}{\partial \theta}, \tfrac{\partial}{\partial r})\tfrac{\partial}{\partial r} = \nabla_{\frac{\partial}{\partial \theta}} \nabla_{\frac{\partial}{\partial r}} \tfrac{\partial}{\partial r} - \nabla_{\frac{\partial}{\partial r}} \nabla_{\frac{\partial}{\partial \theta}} \tfrac{\partial}{\partial r} = -\nabla_{\frac{\partial}{\partial r}} (\frac{1}{\rho}\frac{\partial \rho}{\partial r}\frac{\partial}{\partial \theta}) = -\frac{1}{\rho}\frac{\partial^2 \rho}{\partial r^2}\frac{\partial}{\partial \theta},$$

so that the Gaussian (or sectional) curvature is $-\frac{1}{\rho}\frac{\partial^2 \rho}{\partial r^2}$.

Let $J = a(t)\frac{\partial}{\partial r} + b(t)\frac{\partial}{\partial \theta}$ be a vector field along the geodesic $\gamma(t) = (r(t), \theta(t)) = (t, \theta_0)$. Then

$$J' = a'(t)\frac{\partial}{\partial r} + \big(b'(t) + b(t)\frac{\partial \log \rho}{\partial r}\big)\frac{\partial}{\partial \theta},$$

giving

$$\begin{aligned} J'' &= a''(t)\frac{\partial}{\partial r} + \big(b''(t) + b'(t)\frac{\partial \log \rho}{\partial r} + b(t)\frac{\partial^2 \log \rho}{\partial r^2}\big)\frac{\partial}{\partial \theta} + \big(b'(t) + b(t)\frac{\partial \log \rho}{\partial r}\big)\frac{\partial \log \rho}{\partial r}\frac{\partial}{\partial \theta} \\ &= a''(t)\frac{\partial}{\partial r} + \big(b''(t) + 2b'(t)\frac{\partial \log \rho}{\partial r} + b(t)\frac{\partial^2 \log \rho}{\partial r^2} + b(t)(\frac{\partial \log \rho}{\partial r})^2\big)\frac{\partial}{\partial \theta} \\ &= a''(t)\frac{\partial}{\partial r} + \big(b''(t) + 2b'(t)\frac{1}{\rho}\frac{\partial \rho}{\partial r} + b(t)\frac{1}{\rho}\frac{\partial^2 \rho}{\partial r^2}\big)\frac{\partial}{\partial \theta} \end{aligned}$$

Writing $R(J, \gamma')\gamma' = -b(t)\frac{1}{\rho}\frac{\partial^2 \rho}{\partial r^2}\frac{\partial}{\partial \theta}$, we see that J is a Jacobi field if

$$0 = J'' + R(J, \gamma')\gamma' = a''(t)\frac{\partial}{\partial r} + \big(b''(t) + 2b'(t)\frac{1}{\rho}\frac{\partial \rho}{\partial r}\big)\frac{\partial}{\partial \theta}.$$

Summing up, Jacobi fields satisfy

$$a''(t) = 0, \quad b''(t) + 2b'(t)\frac{1}{\rho}\frac{\partial \rho}{\partial r} = 0.$$

The equation can be integrated by observing that

$$(b'\rho^2)' = 2b'\rho'\rho + b''\rho^2,$$

so $b' = K\rho^{-2}$. So a basis of $\mathcal{J}$ is given by

$$\tfrac{\partial}{\partial r}, t\tfrac{\partial}{\partial r}, \tfrac{\partial}{\partial \theta}, b\tfrac{\partial}{\partial \theta},$$

where $b(t)$ is any primitive of $t \mapsto \rho^{-2}(t, \theta_0)$. Notice that $\frac{\partial}{\partial r}$ and $b\frac{\partial}{\partial \theta}$ are not smooth vector fields on M, but they are smooth vector field along the radial geodesic.

Consistently with Proposition 7.2, a suitable combination of $t\frac{\partial}{\partial r}$ and $\frac{\partial}{\partial \theta}$ gives a Jacobi field with $J(0) = 0$ and arbitrary $J(1)$; this works because the exponential map is a diffeomorphism on the patch we are considering.

Additionally, observe that in this case

$$\nabla \operatorname{grad} r = \nabla \tfrac{\partial}{\partial r} = d\theta \otimes \frac{1}{\rho}\frac{\partial \rho}{\partial r}\frac{\partial}{\partial \theta}, \quad \operatorname{Hess} r = \rho \frac{\partial \rho}{\partial r} d\theta \otimes d\theta.$$

If we take a Jacobi field of the form

$$J = \alpha \tfrac{\partial}{\partial r} + \beta t \tfrac{\partial}{\partial r} + \gamma \tfrac{\partial}{\partial \theta} + \delta b \tfrac{\partial}{\partial \theta},$$

we see that

$$\operatorname{Hess} r(J(t), J(t)) = \rho \frac{\partial \rho}{\partial r}(\gamma + \delta b(t))^2,$$

whereas

$$g(J'(t), J(t)) = g\left(\beta \frac{\partial}{\partial r} + \gamma \frac{\partial \log \rho}{\partial r}\frac{\partial}{\partial \theta} + \delta\big(\rho^2 + b(t)\frac{\partial \log \rho}{\partial r}\big)\frac{\partial}{\partial \theta}, J(t)\right),$$

and so we see that the formula of Lemma 7.11 basically only holds for $J = \frac{\partial}{\partial \theta}$.

Example 7.10. If we set $\rho(r, \theta) = \sin r$ in Example 7.9, we find $b = -\frac{1}{\tan t}$, and the Jacobi field $\frac{\partial}{\partial \theta}$ vanishes at $t = \pi$. Thus, we cannot realize $\frac{\partial}{\partial \theta} \in T_{\gamma(\pi)}M$ as $J(1)$ for a Jacobi field with $J(0) = 0$.

Notice that this is the metric on the round sphere.

Another use of Jacobi fields arises as follows.

Lemma 7.11. *Suppose that* $\exp_p \colon B(0, \rho) \to M$ *is a diffeomorphism onto its image and let* r *be the function*

$$r(x) = d(x, p).$$

Take a unit speed geodesic $\gamma = \exp_p tv$ *and a Jacobi field* J *along* γ *with*

$$J(0) = 0, \quad g(J'(0), \gamma'(0)) = 0^1.$$

Then

$$\operatorname{Hess} r(J(t), J(t)) = g(J'(t), J(t)), \quad t < \rho.$$

In addition, we have $\operatorname{Hess} r(\gamma'(t), X) = 0$ *for all* X.

[1] The hypothesis $g(J'(0), \gamma'0)$ appears to be missing in the third edition of [7], but it is clearly necessary. Consider $J = t\gamma'$ as a counterexample. The second edition contains other problems in this passage.

Proof. Since $J(0) = 0$, if $J'(0) = 0$ the Jacobi field is null and there is nothing to prove. Otherwise, we can rescale J so that $J'(0)$ has unit norm.

Let $\overline{\gamma}$ be a unit speed geodesic variation with Jacobi field J. We cannot use (14) to construct $\overline{\gamma}$, because we need $\gamma_s'(0)$ to have norm one. Therefore, we put

$$\overline{\gamma}(s,t) = \exp_p(t((\cos s)v + (\sin s)J'(0)).$$

Since every γ_s is a geodesic, by Gauss' lemma we have that $(\operatorname{grad} r)_{\gamma_s(t)} = \gamma_s'(t)$; thus, as a section of $\overline{\gamma}^*TM$, $\operatorname{grad} r$ coincides with $\frac{\partial\overline{\gamma}}{\partial t}$. We can write

$$\begin{aligned}\operatorname{Hess} r(J(t),J(t)) &= g\left(\nabla_{J(t)} \operatorname{grad} r, J(t)\right) = g\left((\overline{\gamma}^*\nabla)_{\frac{\partial}{\partial s}} \frac{\partial\overline{\gamma}}{\partial t}, J(t)\right)\\ &= g\left(\frac{\partial^2\overline{\gamma}}{\partial s\partial t}|_{s=0}, J(t)\right) = g\left(\frac{\partial^2\overline{\gamma}}{\partial t\partial s}|_{s=0}, J(t)\right) = g\left(J'(t), J(t)\right).\end{aligned}$$

For the second claim, we similarly compute

$$\operatorname{Hess} r(\gamma', X) = g(\nabla_{\frac{\partial}{\partial r}} \operatorname{grad} r, X) = g((\overline{\gamma}^*\nabla)_{\frac{\partial}{\partial t}} \frac{\partial\overline{\gamma}}{\partial t}, X) = 0. \qquad \square$$

Remark 7.12. By Proposition 7.2, in a geodesic ball $\exp_p(B_\rho(0))$ every tangent vector u can be written as $J(1)$ where J is a Jacobi field along a geodesic $\exp_p tv$ with $J(0) = 0$. In addition, by Proposition7.4 we can assume that $J'(0)$ is orthogonal to v if u is orthogonal to the radial direction. Thus, Lemma 7.11 determines the Hessian of r completely.

Example 7.13. On the unit sphere, let $p = (0,0,1)$ be the north pole, and $q = (\sin\rho, 0, \cos\rho)$ a point in the boundary of a geodesic ball $B_\rho(p)$. Then

$$T_qS^2 = \operatorname{Span}\{v,w\}, \quad v = (\cos\rho, 0, -\sin\rho), w = (0,1,0);$$

using Lemma 7.11, we know that

$$\operatorname{Hess} r(v,v) = 0 = \operatorname{Hess} r(v,w),$$

and we can compute $\operatorname{Hess}(w,w)$ by considering a unit geodesic γ from p to q and a Jacobi field J with $J(0) = 0$ and $J(\rho)$ a multiple of $(0,1,0)$, namely

$$J(t) = \sin t(0,1,0).$$

This gives

$$\operatorname{Hess} r(J(\rho), J(\rho)) = g(J'(\rho), J(\rho)) = \cos\rho\sin\rho$$

i.e.

$$\operatorname{Hess}(w,w) = \frac{1}{\sin^2\rho}\cos\rho\sin\rho = \frac{\cos\rho}{\sin\rho}.$$

In particular, it is positive when $\rho < \pi/2$, and negative when $\pi/2 < \rho\pi$. By Lemma 5.10, $\operatorname{Hess}(w,w)$ is the second derivative of r along a geodesic orthogonal to the geodesic ball $B_\rho(p)$. By the Gauss Lemma, $\gamma'(0)$ is orthogonal to $\operatorname{grad} r$, so the Hessian computation implies that 0 is a local maximum of r along $\gamma(t)$ when $\rho > \pi/2$, and a local minimum when $\rho < \pi/2$. In fact one can write explicitly γ as

$$\gamma(t) = (\cos t)(\sin\rho, 0, \cos\rho) + (\sin t)(0,1,0) = (\cos t\sin\rho, \sin t, \cos t\cos\rho),$$

and it is clear that it lies in the interior, the exterior or the boundary of the geodesic ball $B_\rho(p)$ according to whether $\rho > \pi/2$, $\rho = \pi/2$ or $\rho < \pi/2$.

Example 7.14. In the hyperbolic plane, one obtains a similar computation, except $\sin t$ and $\cos t$ are replaced by their hyperbolic counterparts, and $g(J, J') > 0$. This is consistent with the fact that the orthogonal geodesic lies on the exterior of the geodesic ball, as one can see by choosing $p = 0$ in the disc model.

8 The Ambrose-Cartan-Hicks theorem [4]

Given Riemannian manifolds (M, g), $(\overline{M}, \overline{g})$, two points $p \in M$, $\overline{p} \in \overline{M}$, and a linear isometry $I\colon T_\gamma(0)M \to T_{\overline{\gamma}(0)}\overline{M}$, any geodesic γ starting at p induces a geodesic $\overline{\gamma}$ in M starting at $\overline{p}$ by the diagram

$$\begin{tikzcd}
 & T_pM \arrow[r, "I"] \arrow[d, "\exp_p"] & T_{\overline{p}}\overline{M} \arrow[d, "\exp_{\overline{p}}"] \\
{[0,l]} \arrow[ur, dotted] \arrow[r, "\gamma"] \arrow[rr, bend right] & M & \overline{M}
\end{tikzcd} \qquad \overline{\gamma}(t) = \exp_{\overline{p}}(tI(\gamma'(0))).$$

In addition, it induces a one-parameter family of linear isometries

$$I_t^\gamma : T_\gamma(t)M \to T_{\overline{\gamma}(t)}\overline{M}$$

by composing I with parallel transport along γ and $\overline{\gamma}$. If $\{e_1, \dots, e_n\}$ is an orthonormal frame at p, extended to parallel vector fields $e_1(t), \dots, e_n(t)$ along γ, and $\overline{e}_1(0) = I(e_1), \dots, \overline{e}_n = I(e_n)$, with $\overline{e}_1(t), \dots, \overline{e}_n(t)$ similarly defined by parallel translation, then I_t maps $e_j(t)$ into $\overline{e}_j(t)$. Notice that $I_t^\gamma \circ D_t = D_t \circ I_t^\gamma$, because for any vector field Y along γ,

$$Y'(t) = (Y^i)'(t)e_i(t), \quad \overline{Y}'(t) = (Y^i)'(t)\overline{e}_i(t).$$

We will write I_t instead of I_t^γ for brevity when γ is understood.

Notice that this construction carries geodesics to geodesics on possibly unrelated Riemannian manifolds. However, we will see that when curvature is preserved, this determines an isometry.

Lemma 8.1. *Let and let $I\colon T_pM \to T_{\overline{p}}\overline{M}$ be a linear isotropy at a point. Suppose that $B_r(p)$ and $B_r(\overline{p})$ are geodesic balls, and consider the diffeomorphism*

$$\phi\colon B_r(p) \to B_r(\overline{p}), \quad \phi(\exp_p(v)) = \exp_{\overline{p}}(I(v)).$$

Suppose that for any geodesic $\gamma(t) = \exp_p tv$, we have

$$I_t(R(v, w)u) = \overline{R}(I_t(v), I_t(w))I_t(u).$$

Then ϕ is an isometry and $d\phi_{\gamma(t)} = I_t$.

Proof. Let γ be the geodesic from p to a point q contained in $B_r(p)$ and let J be a Jacobi field along γ such that $J(0) = 0$. Define $\overline{J}$ along $\overline{\gamma}$ by $\overline{J}(t) = I_t^\gamma(J(t))$. Then $\overline{J}$ is a Jacobi field, and $|J(t)| = \left|\overline{J}(t)\right|$. We claim that

$$\overline{J}(1) = d\phi_q(J(1)). \tag{17}$$

This will prove the statement, because for any $u \in T_qM$ there exists a Jacobi field J with $J(0) = 0$ and $J(1) = u$, and by construction $d\phi_q(u) = \overline{J}(1)$ has the same norm as $J(1) = u$, i.e. $d\phi_q = I_1$ is a linear isometry.

To prove (17), observe that

$$d\exp_p(\gamma'(0); J'(0)) = J(1), \quad d\exp_{\overline{p}}(\overline{\gamma}'(0); \overline{J}'(0)) = \overline{J}(1).$$

The claim follows applying the chain rule to the diagram

$$\begin{tikzcd} T_pM \supset B_r(0) \arrow[r, "I"] \arrow[d, "\exp_p"] & B_r(0) \subset T_{\overline{p}}\overline{M} \arrow[d, "\exp_{\overline{p}}"] \\ B_r(p) \arrow[r, "\phi"] & B_r(\overline{p}) \end{tikzcd}$$

□

We will need to consider broken geodesics, i.e. piecewise smooth curves in some $\Omega_{p,q}$ obtained by concatenating geodesics.

Given a broken geodesic γ in M from p to q, and a linear isometry $I\colon T_pM \to T_{\overline{p}}\overline{M}$, we can define a broken geodesic $\overline{\gamma}$ and a one-parameter family of linear isometries $I_t^\gamma\colon T_{\gamma(t)}M \to T_{\overline{\gamma}(t)}\overline{M}$ with the property that on each interval $[a,b]$ where γ is smooth, $\overline{\gamma}|_{[a,b]}$ and $I|_{[a,b]}$ are induced from $\gamma|_{[a,b]}$ in the sense above.

Recall that the injectivity radius at p, denoted by inj_p, is the largest r such that $\exp_p\colon B_r(0) \to M$ is a diffeomorphism onto its image, i.e. $B_r(p)$ is a geodesic ball. The injectivity radius of M is

$$\operatorname{inj} M = \inf_{p\in M} \operatorname{inj}_p .$$

It turns out that the injectivity radius is continuous, but we will only need a weaker property, namely that the injectivity radius is positively bounded below on any compact set $K \subset M$; equivalently, every point of M has a neighbourhood U such that there exists an ϵ for which $B_\epsilon(p)$ is a geodesic ball for all $p \in U$. This follows from considering the map

$$TM \to M \times M, \quad (p; v) \mapsto (\exp_p(v), p),$$

which is an embedding when restricted to a neighbourhood of the zero section.

Example 8.2. Consider a surface of revolution in $\mathbb{R}^3$ of the form

$$x^2 + y^2 = f(z).$$

It is complete because it is closed in $\mathbb{R}^3$: any Cauchy sequence for the Riemannian distance is also a Cauchy sequence for the Euclidean distance, hence it converges. Notice that each critical value r of f determines a periodic geodesic of length $2\pi r$. If the critical values of f converge to zero, the injectivity radius is zero.

Theorem 8.3 (Ambrose-Cartan-Hicks)**.** *Let M, $\overline{M}$ be complete Riemannian manifolds, with M simply connected, and $I\colon T_pM \to T_{\overline{p}}\overline{M}$ a linear isometry. Suppose that for all broken geodesics γ in M, the induced one-parameter family of linear isometries I_t satisfies*

$$I_t(R(v,w)u) = \overline{R}(I_t(v), I_t(w))I_t(u).$$

Then for all broken geodesics α, γ in $\Omega_{p,q}$ the induced broken geodesics satisfy $\overline{\alpha}(1) = \overline{\gamma}(1)$, and the map $\Phi\colon M \to \overline{M}$ defined by $\gamma(1) \to \overline{\gamma}(1)$ is a local isometry.

Proof. Let $\alpha, \gamma \in \Omega_{p,q}$ be broken geodesics. We need to show that $\overline{\alpha}(0) = \overline{\gamma}(1)$.

The proof is in three steps. First, assume that $\alpha, \gamma, \overline{\alpha}, \overline{\gamma}$ are contained in geodesic balls $B_r(p)$, $B_r(\overline{p})$. Notice that α, γ are *broken* geodesics, so they need not coincide. However, by the lemma, the map ϕ is an isometry; hence, $\overline{\alpha} = \phi \circ \alpha$, $\overline{\gamma} = \phi \circ \gamma$, so their terminal points are the same, and the claim is proved in this case. Notice also that $I_1^\gamma = d\phi = I_1^\alpha$.

Now, choose points $0 = t_0 < \ldots < t_n = 1$ such that γ and α are segments in $[t_i, t_{i+1}]$ and assume that for each i, some geodesic ball centered at $\gamma(t_i)$ contains the points $\alpha(t_{i+2})$, $\alpha(t_{i+1})$, $\gamma(t_{i+1})$ and the segments that connect each pair of these points; we also assume that the same holds replacing α, γ with $\overline{\alpha}$, $\overline{\gamma}$.

For each $k = 1, \ldots, n$, let $\tau_k \colon [t_{k-1}, t_k] \to M$ be the segment from $\gamma(t_{k-1})$ to $\alpha(t_k)$, and let

$$\beta_k(t) = \begin{cases} \gamma(t) & 0 \le t \le t_{k-1} \\ \tau_k(t) & t_{k-1} \le t \le t_k \end{cases}.$$

By induction on k, we show that $\overline{\beta}_k(t_k) = \overline{\alpha}(t_k)$ and that

$$I_{t_k}^{\beta_k} = I_{t_k}^{\alpha}.$$

Since $\beta_n = \gamma$, this will give $\overline{\alpha}(1) = \overline{\gamma}(1)$.

For $k = 1$, $\beta_1 = \tau_1$ is a segment in a geodesic ball around p, going from p to $\alpha(t_1)$. However, $\alpha|_{[0,t_1]}$ is also a segment, so they coincide and there is nothing to prove.

If the claim holds for k, observe that the curves

$$\delta(t) = \begin{cases} \tau_k(t) & t_{k-1} \le t \le t_k \\ \alpha(t) & t_k \le t \le t_{k+1} \end{cases}, \quad \sigma(t) = \begin{cases} \gamma(t) & t_{k-1} \le t \le t_k \\ \tau_{k+1}(t) & t_k \le t \le t_{k+1} \end{cases}$$

are broken geodesics contained in a geodesic ball, as they are obtained by joining the four points $\gamma(t_{k-1})$, $\alpha(t_k)$, $\alpha(t_{k+1})$, $\gamma(t_k)$ by segments. Considering the linear isometry $I_{t_{k-1}}^\gamma \colon T_{\gamma(t_{k-1})}M \to T_{\overline{\gamma}(t_{k-1})}\overline{M}$, we have induced broken geodesics $\overline{\delta}$, $\overline{\sigma}$, which satisfy the same conditions. Thus, by the first step $\overline{\delta}(t_{k+1}) = \overline{\sigma}(t_{k+1})$. Adjoining the geodesic $\gamma|_{[0,t_{k-1}]}$, we see that $\overline{\beta}_{k+1}(t_{k+1}) = \overline{\sigma}(t_{k+1}) = \overline{\delta}(t_{k+1}) = \overline{\alpha}(t_{k+1})$, proving the induction step.

Finally, let γ and α be arbitrary, with $\gamma(1) = \alpha(1)$. Since M is simply connected, there is a homotopy $h \colon I \times I \to M$ from $\gamma = h_0$ to $\alpha = h_1$ with $h(s, 0) = p$, $h(s, 1) = q$. Let r be the minimum injectivity radius at points in the image of h. We can subdivide the interval I in $0 = t_0 < \cdots < t_n = 1$ and $0 = s_0 < \cdots < s_m = 1$ in such a way that γ and α are segments in the intervals $[t_i, t_{i+1}]$, and the diameter of each rectangle $h([s_j, s_{j+1}] \times [t_i, t_{i+1}])$ is less than $r/4$. This implies that the geodesic ball of radius r around $h(s_j, t_i)$ contains $h(s_j, t_{i+1})$, $h(s_{j+1}, t_{i+2})$ and $h(s_{j+1}, t_{i+2})$ as well as the segments that join any two of them (because they have length $< r/2$). This property is preserved if the partitions $\{t_i\}$ and $\{s_j\}$ are refined.

Now define broken geodesics γ_j by adjoining segments from $h(s_j, t_i)$ to $h(s_j, t_{i+1})$ for all i. By construction, $\gamma_0 = \gamma$ and $\alpha = \gamma_1$. In addition, for each j we have a broken geodesic $\overline{\gamma}_j$ from $h(s_j, 0) = p$.

We can arrange for a geodesic ball of radius $\overline{r}$ around $\overline{\gamma}_j(s_j, t_i)$ to contain $\overline{\gamma}_j(t_{i+1})$, $\overline{\gamma}_{j+1}(t_{i+1})$, $\overline{\gamma}_{j+1}(t_{i+2})$ as well as the segments that join any two of them.

This is because having fixed the partition $\{t_i\}$, the length of the broken geodesics γ_j is bounded; thus, the geodesics $\overline{\gamma}_j$ lie in some compact subset of $\overline{M}$, where the injectivity radius can be assumed to be $\geq \overline{r}$. Since the correspondence $\gamma \to \overline{\gamma}$ is continuous, we can refine the $\{s_j\}$ so that $d(\overline{\gamma}_j(t), \overline{\gamma}_{j+1}(t)) < \overline{r}/4$. Adding further breakpoints $\{t_i\}$, we can assume that $d(\overline{\gamma}_j(t_i), \overline{\gamma}_j(t_i+1)) < \overline{r}/4$.

Thus, each pair γ_j, γ_{j+1} satisfies the second step, and $\overline{\gamma}_j(1) = \overline{\gamma}_{j+1}(1)$. This gives $\overline{\gamma}(1) = \overline{\alpha}(1)$.

Now let q be arbitrary, and let γ be a geodesic from p to q; let $\overline{\gamma}$ be the induced geodesic, with $\overline{q} = \overline{\gamma}(1)$. By the lemma, the map

$$\phi = \exp_{\overline{q}} \circ I_1^\gamma \circ \exp_q^{-1}$$

is an isometry from a geodesic ball $B_r(q)$ to $B_r(\overline{q})$. However, for any $x \in B_r(q)$, we can obtain $\Phi(x)$ by considering a broken geodesic from p to q, then q to x, and this shows that $\Phi|_{B_r(q)} = \phi$. So it is an isometry. □

Lemma 8.4. *Let $f \colon \tilde{M} \to M$ be a local isometry, where $\tilde{M}$ is complete and M connected. Then f is a covering map.*

Proof. As f is a local isometry, geodesics of M lift to geodesics of $\tilde{M}$, and conversely, geodesics of $\tilde{M}$ project to geodesics of M. In particular, geodesics through $f(x) \in M$ are defined for all $\mathbb{R}$, so M is complete.

To see that f is surjective, pick $\tilde{x} \in \tilde{M}$ and $y \in M$, and choose a geodesic γ joining $f(x)$ to y. Then γ lifts to a geodesic $\tilde{\gamma}$ from $\tilde{x}$ to some $\tilde{y}$, where necessarily $f(\tilde{y}) = y$.

To see that f is a covering map, let $V = B_\epsilon(x)$ be a geodesic ball in M, meaning that it is the diffeomorfic image of $B_\epsilon(0) = T_xM$ under $\exp_x$. Let $\tilde{x}_\alpha$ denote the preimages of x. For every α, let U_α denote the metric ball of center x_α and radius ϵ. The U_α are pairwise disjoint: any minimizing geodesic connecting $\tilde{x}_\alpha$ a $\tilde{x}_\beta$ projects down to a geodesic γ loop based at x. Since V does not contain geodesic loops, γ must leave and reenter V, so it has length $\geq 2\epsilon$. Thus, the distance between the centers of the metric balls U_α is at least 2ϵ, which forces the U_α to be disjoint.

It is clear that $f(U_\alpha) \subset V$, as f maps geodesics to geodesics. In addition, $f \colon U_\alpha \to V$ is invertible: given $y \in V$, the geodesic from x to y lifts to a geodesic of length $< \epsilon$ which is contained in U_α; its terminal point is the unique preimage of y in U_α.

The fact that $f|_{U_\alpha} \colon U_\alpha \to V$ is a diffeomorphism now follows from the fact that it is a local isometry, hence a local diffeomorphism. □

Theorem 8.5. *If M is a connected, complete Riemannian manifold of constant curvature k, then the universal covering of M with the pullback metric is one of*

$$\begin{cases} S^n_{\frac{1}{\sqrt{k}}} & k > 0 \\ \mathbb{R}^n & k = 0 \\ H^n_{\frac{1}{\sqrt{-k}}} & k < 0 \end{cases}$$

Proof. Let N_k be one of the model spaces appearing in the statement; each is a complete, simply connected manifold of constant curvature k. Fixing a point $p \in N_k$ and a point $\overline{p} \in M$, and any isometry $I \colon T_pN_k \to T_{\overline{p}}M$, we can construct a local isometry Φ as in Theorem 8.3. Then Φ is a covering map by Lemma 8.4. □

9 Manifolds of negative sectional curvature

Theorem 9.1 (Cartan-Hadamard)**.** *If (M, g) is complete with* $\sec \leq 0$ *at every point, then* $\exp_p \colon T_pM \to M$ *is a covering map. In particular, the universal covering of M is diffeomorphic to $\mathbb{R}^n$.*

Proof. We claim that geodesics do not have conjugate points. By Lemma 7.1, this implies that $d\exp_p$ is invertible at each point. Thus, the map $\exp_p \colon T_pM \to M$ is a local diffeomorphism. We can pull back the metric of M to T_pM so that $\exp_p$ becomes an isometry. Since straight lines through 0 in T_pM are geodesics for the pullback metric, they are defined for all time. Hence, T_pM is complete by the Hopf-Rinow theorem. Then Lemma 8.4 implies that $\exp_p$ is a covering map.

To see that there are no conjugate points, suppose that J is a Jacobi field along a geodesic, $J(0) = 0$, $J'(0) \neq 0$. Then

$$\frac{1}{2}\frac{d}{dt}|J|^2 = g(J', J).$$

Since $\sec \leq 0$, we have

$$\frac{d}{dt}g(J', J) = g(J'', J) + |J'|^2 = -g(R(J, \gamma')\gamma', J) + |J'|^2 \geq |J'|^2.$$

Therefore

$$g(J', J) \geq g(J'(0), J(0)) + \int_0^t |J'|^2\, dt = \int_0^t |J'|^2\, dt > 0,$$

where we have used $J'(0) \neq 0$. Thus,

$$\frac{1}{2}|J(t)|^2 > 0, t > 0,$$

so there can be no conjugate points. □

Denote as usual by r the distance function from a fixed point p. Since r is not smooth at p, we will consider the modified distance function

$$f_p = \frac{1}{2}r^2 = \frac{1}{2}d(p, \cdot)^2;$$

we will occasionally write f instead of f_p when the reference point p is understood. We have

$$\nabla df = \nabla r dr = dr \otimes dr + r\nabla dr = dr \otimes dr + r \operatorname{Hess} r,$$

and we know how to compute $\operatorname{Hess} r$ thanks to Lemma 5.10.

For general Riemannian manifolds, f becomes nonsmooth at points where radial geodesics cease to be minimizing ([7, Corollary 5.7.11]). However, if $\exp_p \colon T_pM \to M$ is a diffeomorphism as in the Cartan-Hadamard theorem, then r is the distance from zero in normal coordinates, so f is smooth everywhere.

Lemma 9.2. *If (M, g) is simply connected, complete and with* $\sec \leq 0$*, then f is proper and strictly convex along geodesics.*

Proof. Properness is a consequence of the Cartan-Hadamard theorem: the preimage of $[0, r^2]$ is the closed geodesic ball $B_r(p)$.

Given a geodesic γ, we must show that $f \circ \gamma$ is convex. By Lemma 5.10, it suffices to show that $\operatorname{Hess} f$ is positive definite, or equivalently that $\operatorname{Hess} r$ is positive definite on $\frac{\partial}{\partial r}^{\perp}$.

Let $q = \exp_p(tv)$ be any point on the unit geodesic $\gamma(t) = \exp_p tv$, and $u \in T_qM$ be orthogonal to $\frac{\partial}{\partial r}$. Since $\exp_p$ is a diffeomorphism, there exists a Jacobi field J such that $J(0) = 0$ and $J(t) = u$. By Lemma 7.11, we have

$$\operatorname{Hess} r(u,u) = g(J'(1), J(1)) > 0,$$

where the last inequality is obtained as in the proof of the Cartan-Hadamard theorem. □

Theorem 9.3 (Cartan)**.** *If (M, g) is simply connected, complete with* $\sec \leq 0$*, then any isometry with finite order has a fixed point.*

Proof. Let $F \colon M \to M$ be an isometry of finite order. Take any $p \in M$ and let $\{p, F(p), \ldots, F^k(p)\}$ be its orbit.

Let f be the modified distance function from p. We know that f is proper and strictly convex along geodesics. The same applies to $f \circ F^i$ for any i, since $f \circ F^i$ is the modified distance from $F^{-i}(p)$. It follows that the function

$$g(q) = \max\{f(q), f(F(q)), \ldots, f(F^k(q))\}.$$

is also proper and strictly convex along geodesics (though not smooth). Since g is nonnegative and proper, the completeness of M implies that g has a minimum at some q.

By construction, $g(q)$ is the maximum of f restricted to the orbit of F, so it is invariant under F. Therefore, g has a minimum at $F(q)$.

If q and $F(q)$ were distinct, the geodesic from q to $F(q)$ would contain two minima, which is absurd. It follows that q is a fixed point for F. □

Corollary 9.4. *If (M, g) is complete with* $\sec \leq 0$*, then the fundamental group is torsion free.*

Proof. We need to show that $\pi_1(M)$ contains no nontrivial element of finite order.

Consider the universal covering $\tilde{M}$ with the pullback metric. Then $\pi_1(M)$ acts on $\tilde{M}$, freely and by isometries. If $F \in \pi_(M)$ is an element of finite order, then Theorem 9.3 implies that F has a fixed point. Therefore, F is the identity. □

Up to here, we have used the fact that $\operatorname{Hess} f$ is positive if $\sec \leq 0$. In Euclidean space, we have

$$f_0 = \frac{1}{2}(x_1^2 + \ldots + x_n^2), \quad \operatorname{Hess} f_0 = \nabla(x_1 dx_1 + x_n dx_n) = dx_1 \otimes dx_1 + \ldots + dx_n \otimes dx_n = g.$$

With the assumption $\sec \leq 0$, this becomes an inequality. We will state this locally, so we need to restrict f_p so that it is smooth.

Lemma 9.5. *Let U be a geodesic ball around p. If* $\sec \leq 0$*, then* $\operatorname{Hess} f_p \geq g$ *on U, i.e.* $\operatorname{Hess} f_p(v,v) \geq g(v,v)$*. If* $\sec < 0$*, then* $\operatorname{Hess} f_p(v,v) > g(v,v)$ *unless v is in* $\operatorname{Span}\left\{\frac{\partial}{\partial r}\right\}$.

Proof. We know that $\operatorname{Hess} f_p = r \operatorname{Hess} r + dr \otimes dr$, where $\operatorname{Hess} r$ is zero on $\frac{\partial}{\partial r}$. Therefore, we only need to show

$$\operatorname{Hess} f_p(v,v) = r \operatorname{Hess} r(v,v) \geq g(v,v), \quad v \in \frac{\partial}{\partial r}^{\perp}.$$

Let J be a Jacobi field along a unit-speed radial geodesic so that $J(0) = 0$, $J(t_0) = v$. Then by Lemma 7.11, we have

$$\operatorname{Hess} r(J(t), J(t)) = g(J'(t), J(t));$$

thus, we must show that

$$tg(J'(t), J(t)) \geq |J(t)|^2 .$$

In other words, we must show that $\lambda(t) \leq t$, where

$$\lambda(t) = \frac{|J(t)|^2}{g(J'(t), J(t))}.$$

As $t \to 0$, we have $|J(t)|^2 = o(t)$, as it vanishes at zero with its derivative, whereas

$$g(J'(t), J(t)) \sim t\big(g(J'(0), J'(0)) - R(J(0), \gamma'(0))\gamma'(0)\big) \sim |J'(0)|^2 t;$$

therefore λ extends continuously setting $\lambda(0) = 0$. In addition

$$\begin{aligned}\lambda' = \frac{2g(J',J)^2 - |J|^2\left(|J'|^2 - R(J',\gamma')\gamma', J\right)}{g(J'(t),J(t))^2} &\leq \frac{2g(J',J)^2 - |J|^2 |J'|^2}{g(J'(t),J(t))^2} \\ &\leq \frac{2g(J',J)^2 - g(J',J)^2}{g(J'(t),J(t))^2} = 1,\end{aligned}$$

where we have used $\sec \leq 0$ and the Cauchy-Schwartz inequality. Hence, $\lambda(t) \leq t$ as claimed.

If $\sec < 0$, then $\lambda' < 1$, $\lambda(t) < t$ as soon as $t > 0$ and we obtain a strict inequality. Notice that this only applies if v, hence $J'(0)$ has a nonzero component orthogonal to $\frac{\partial}{\partial r}$. $\square$

Recall that three geodesics γ_a, γ_b, γ_c form a *geodesic triangle* in M if there are three points each pair of which is connected by one of the geodesics, say

$$\gamma_a \colon [0,a] \to M, \quad \gamma_b \colon [0,b] \to M, \quad \gamma_c \colon [0,c] \to M;$$

we can assume that the geodesic are unit speed and

$$\gamma_a(a) = \gamma_b(0), \quad \gamma_b(b) = \gamma_c(0), \quad \gamma_c(c) = \gamma_a(0).$$

We can then define the length of each side of the triangle as $a = l(\gamma_a)$, $b = l(\gamma_b)$, $c = l(\gamma_b)$, and the (internal) angles by

$$\cos\alpha = g(-\gamma_b'(b), \gamma_c'(0)), \quad 0 \leq \alpha \leq \pi,$$

with β, γ defined similarly.

Proposition 9.6. *Let M be complete with* $\sec \leq 0$*. Pick a geodesic triangle with sides lengths a, b, c, and let α be the angle opposite a. Then*

$$a^2 \geq b^2 + c^2 - 2bc \cos \alpha.$$

If $\sec < 0$*, then strict inequality holds unless the triangle is degenerate, i.e. a single geodesic contains all sides.*

Proof. Up to taking the universal cover $\tilde{M}$, we can assume that $\exp_p \colon T_pM \to M$ is a diffeomorphism for $p = \gamma_b(0)$.

We have

$$\frac{d}{dt} f_p(\gamma_c(t)) = rg(\tfrac{\partial}{\partial r}, \gamma_c'(t)),$$
$$\frac{d^2}{dt^2} f_p(\gamma_c(t)) = \operatorname{Hess} f_p(\gamma_c'(t), \gamma_c'(t)) \geq |\gamma_c'(t)|^2 = 1,$$

so

$$f_p(\gamma_c(t)) \geq f_p(\gamma_c(0)) + t\frac{d}{dt}|_{t=0} f_p(\gamma_c(t)) + \frac{1}{2}t^2 = f_p(\gamma_c(0)) + btg(\tfrac{\partial}{\partial r}, \gamma_c'(0)) + \frac{1}{2}t^2;$$

for $t = c$ we get

$$\frac{1}{2}a^2 = \frac{1}{2}d(p, \gamma_c(c))^2 \geq \frac{1}{2}b^2 - bc\cos\alpha + \frac{1}{2}c^2.$$

If $\sec < 0$, we can assume that strict inequality holds, except if $\gamma_c'(0)$ is linearly dependent on $\frac{\partial}{\partial r} = \gamma_b'(b)$, which corresponds to degenerate triangles. □

Remark 9.7. In the Euclidean plane, equality holds in Proposition 9.6. Indeed, consider a triangle with vertices $(0, 0)$, $(b \cos \alpha, b \sin \alpha)$, $(c, 0)$. Then

$$d((b \cos \alpha, b \sin \alpha), (c, 0)) = b^2 + c^2 - 2bc \cos \alpha.$$

Corollary 9.8. *Let M be complete with* $\sec < 0$*. Then the sum of the angles of a nondegenerate geodesic triangle is $< \pi$.*

Proof. Let a, b, c be the edge lengths and α, β, γ the angles. Consider a Euclidean triangle with the same edge lengths, and call its angles α', β', γ'. Then

$$b^2 + c^2 - 2bc \cos \alpha < a^2 = b^2 + c^2 - 2bc \cos \alpha',$$

so $\cos \alpha > \cos \alpha'$, i.e. $\alpha < \alpha'$. Repeating for β, γ we find

$$\alpha + \beta + \gamma < \alpha' + \beta' + \gamma' = \pi.$$

□

The analogous statement for quadrilaterals (which follows immediately by splitting a quadrilateral along the diagonal) is instrumental in the proof of the following:

Theorem 9.9 (Preissmann)**.** *If M is compact and* $\sec < 0$*, then any Abelian subgroup of $\pi_1(M)$ is cyclic. In particular, no compact product $M \times N$ admits a metric with* $\sec < 0$*.*

We first point out that the second part of the stament follows from the first: compact implies complete, so the universal covering of a compact manifold with negative curvature is $\mathbb{R}^n$, which implies that both M and N have nontrivial fundamental group. Taking the product of any cyclic group in $\pi_1(M)$ and any cyclic group in $\pi_1(N)$ we obtain an abelian subgroup of $\pi_1(M \times N)$ which is not cyclic.

For the proof, we will need some facts and definitions which apply beyond the context of negative curvature. An *axis* for an isometry F is a geodesic $\gamma\colon \mathbb{R} \to M$ such that $F \circ \gamma$ is a reparametrization of γ. Since $F \circ \gamma$ is again a geodesic with the same speed, we get $F(\gamma(t)) = \gamma(\pm t + a)$. If $F(\gamma(t)) = \gamma(-t+a)$, then F fixes the point $\gamma(a/2)$. If $F(\gamma(t)) = \gamma(t+a)$ we say that the axis has *period* a. The period depends on the parametrization. Notice that in the context of simply connected manifolds with negative sectional curvature, geodesics are injective maps, so the period is uniquely determined.

Example 9.10. A rotation of S^2 around the line Span $\{v\}$ has a single axis, which is the geodesic orthogonal to Span $\{v\}$. The period is the angle of rotation, or any multiple.

Example 9.11. Consider the hyperbolic half-plane H with the metric $\frac{1}{y^2}(dx^2 + dy^2)$. Then the map

$$F(x, y) = (2x, 2y)$$

is an isometry, since it maps the orthonormal frame $\{y\frac{\partial}{\partial x}, y\frac{\partial}{\partial y}\}$ to itself. The half-line $t \mapsto (0, e^t)$ is a geodesic, since

$$D_t e^t \frac{\partial}{\partial y} = e^t \nabla_{e^t \frac{\partial}{\partial y}} \frac{\partial}{\partial y} + e^t \frac{\partial}{\partial y} = e^{2t}(-e^{-t}\frac{\partial}{\partial y}) + e^t \frac{\partial}{\partial y} = 0.$$

It is also an axis.

Notice that on the quotient $H/\langle F\rangle$, the axis is mapped into a periodic geodesic.

For a fixed isometry F, we define the function

$$x \mapsto \delta_F(x) = d(x, F(x)).$$

Lemma 9.12. *Let M be complete and $F\colon M \to M$ be an isometry. If δ_F has a positive minimum, then F has an axis.*

Proof. Let δ_F have a mimimum at p and let $\gamma\colon [0,1] \to M$ be a segment from p to $F(p)$.

For $t \in [0,1]$, we have

$$\begin{aligned}\delta_F(p) \le \delta_F(\gamma(t)) = d(\gamma(t), F(\gamma(t)) &\le d(\gamma(t), \gamma(1)) + d(\gamma(1), F(\gamma(t)) \\ &= d(\gamma(t), \gamma(1)) + d(\gamma(0), \gamma(t)) = d(\gamma(0), \gamma(1)).\end{aligned}$$

Thus equality holds, and this means that we can obtain a segment from $\gamma(t)$ to $F(\gamma(t))$ by joining the segment $\gamma|_{[t,1]}$ to $F \circ \gamma|_{[0,t]}$ (the extension is smooth because it minimizes length and due to Theorem 6.1). Thus, $F \circ \gamma$ is an extension of γ, i.e. $F(\gamma(t)) = \gamma(t+1)$. By repeating the argument, this holds for any t. $\square$

Lemma 9.13. *Let M be compact and $\tilde{M}$ the universal cover. If F is a nontrivial deck transformation, then δ_F has a minimum, bounded below by $2\,\mathrm{inj}(M)$. The image in M of the corresponding axis has minimal length in its free homotopy class of loops in M.*

Proof. Since M is compact, it is complete, and so is $\tilde{M}$.

Let $\pi\colon \tilde{M} \to M$ be the projection and fix $\tilde{x} \in \tilde{M}$, $\pi(\tilde{x}) = x$. We have a one-to-one correspondence between $\pi_1(M, x)$ and $\pi^{-1}(x)$: a loop α is mapped to the terminal point of its lifting $\tilde{\alpha}$ starting at $\tilde{x}$. In particular, any curve from $\tilde{x}$ to $F(\tilde{x})$ covers a (homotopically) nontrivial loop. In particular, any such curve has length at least $2\,\mathrm{inj}_x(M)$, since it covers a loop based at x, and any loop that stays inside a geodesic ball is trivial. This shows that $\delta_F(\tilde{x}) \geq 2\,\mathrm{inj}_x(M)$.

We must show that δ_F has a minimum. Consider a sequence $\{\tilde{q}_i\}$ in $\tilde{M}$ such that $\delta_F(\tilde{q}_i) \to \inf \delta_F$. For each $\tilde{q}_i$, the segment from $\tilde{q}_i$ to $F(\tilde{q}_i)$ projects to a geodesic loop based at $q_i = \pi(\tilde{q}_i)$ of length $\delta_F(\tilde{q}_i)$. Let this loop be

$$[0,1] \ni t \to \exp_{q_i} tv_i,$$

so that $\delta_F(\tilde{q}_i) = |v_i|$. By compactness of M, we can extract a subsequence of $\{q_i\}$ such that $q_i \to q$, $v_i \to v \in T_qM$. Then $q_i = \exp_{q_i} v_i$ converges to $q = \exp_q v$, i.e. $t \mapsto \exp_q tv$ is a loop. Let U be a trivializing neighbourhood of q for the covering, and choose i large enough so that $q_i \in U$. Then if $\tilde{q}_i$ is in some connected component $\tilde{U}$ of $\pi^{-1}(U)$, the terminal point of the geodesic that covers $\exp_{q_i} tv$ is in the component $F(\tilde{U})$. In the limit, we see that the geodesic that covers $\exp_q tv$ also ends inside $F(\tilde{U})$, so its terminal point is $F(q)$. Since its length is $|v| = \inf \delta_F$, this shows that δ_F has a minimum.

Now let γ' be a loop in the same free homotopy class as the geodesic loop $\gamma(t) = \exp_q tv$. This means that we have a (continuous) variation γ_s of γ, with $\gamma_1 = \gamma'$, and where each γ_s is a loop. By the above argument, lifting each γ_s to $\tilde{M}$ we obtain a curve from some p_s to $F(p_s)$. Therefore the length of each γ_s is greater than or equal to $\min \delta_F$, which is the length of γ. $\square$

Remark 9.14. It follows from the lemma that on a compact manifold M, every element of $\pi_1(M)$ is represented by a geodesic loop.

Proof of Preissmann's theorem. We can identify $\pi_1(M)$ with the group of deck transformations of the universal cover $\tilde{M} \to M$. Note that there is at least one nontrivial deck transformation F, because M is compact and $\tilde{M}$ is diffeomorphic to $\mathbb{R}^n$.

We claim that F has a unique axis (up to reparametrization). Notice that since F has no fixed points, axes are of the form $F(\gamma(t)) = \gamma(t + a)$.

By Lemma 9.13, F has an axis γ. Suppose that it also has an axis γ'. Then if γ and γ' intersect at p, they also intersect at $F(p) \neq p$; since they are geodesics and $\exp_p\colon T_p\tilde{M} \to M$ is a diffeomorphism, they coincide.

If γ and γ' do not intersect, pick $p_1 = \gamma(0)$ and $p_2 = \gamma'(0)$. Then join p_1 to p_2 by a segment σ. We have that $F(p_1)$ and $F(p_2)$ are joined by the segment $F \circ \gamma$. Now consider that the angle between $\gamma'(0)$ and $\sigma'(0)$ is the same as the angle between $(F \circ \gamma)'(0)$ and $(F \circ \sigma)'(0)$. Thus, the sum of the angles of the quadrilateral with edges γ, γ', σ and $F(\sigma)$ is 2π. Since sec < 0, this is absurd by Corollary 9.8.

Thus, the axis for F is unique; up to reparametrization of γ, we can suppose it has period 1, i.e. $F(\gamma(t)) = \gamma(t+1)$.

Let H be an abelian group of deck transformations containing F; we must show that H is cyclic. For any G in H, we have that $G \circ \gamma$ is also an axis for F, as

$$(G \circ \gamma)(t+1) = (G \circ F \circ \gamma)(t) = (F \circ G \circ \gamma)(t).$$

This implies that $G \circ \gamma$ is a reparametrization of γ, by uniqueness. Thus, γ is an axis for every deck transformation G in H.

We have a map $H \to \mathbb{R}$ which associates to each F its period. This is a homomorphism: if F_1 has period a_1 and F_2 has period a_2, then $F_1 F_2$ has period $a_1 + a_2$, as

$$F_1 F_2 \gamma(t) = F_1(\gamma(t + a_2)) = \gamma(t + a_1 + a_2).$$

An element of H with period 0 fixes the whole geodesic, and so it is the identity, because H is a group of deck transformations. The image of H is discrete, since by Lemma 9.13 every isometry G with axis γ satisfies $\delta_G(x) \geq 2\,\mathrm{inj}(M)$. Therefore, H is cyclic. □

Remark 9.15. If (M, g), $(\hat{M}, \hat{g})$ are Riemannian manifolds, then the curvature of the product metric $g + \hat{g}$ on $M \times \hat{M}$ is determined as follows. Write

$$g = g_{ij} dx^i dx^j, \quad \hat{g} = \hat{g}_{ij} d\hat{x}^i d\hat{x}^j;$$

then a computation with the Koszul formula shows that the Levi-Civita connection on $\overline{M} = M \times \hat{M}$ is determined by the Christoffel symbols Γ_{ij}^k and $\hat{\Gamma}_{ij}^k$ of g and $\hat{g}$ by

$$\overline{\nabla}_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_j} = \Gamma_{ij}^k \frac{\partial}{\partial x_k}, \quad \overline{\nabla}_{\frac{\partial}{\partial \hat{x}_i}} \frac{\partial}{\partial \hat{x}_j} = \hat{\Gamma}_{ij}^k \frac{\partial}{\partial \hat{x}_k}, \quad \overline{\nabla}_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial \hat{x}_j} = 0 = \overline{\nabla}_{\frac{\partial}{\partial \hat{x}_i}} \frac{\partial}{\partial x_j}.$$

Thus, the only nonzero components in the curvature tensors are

$$\overline{R}(\tfrac{\partial}{\partial x_i}, \tfrac{\partial}{\partial x_j})\tfrac{\partial}{\partial x_k} = R(\tfrac{\partial}{\partial x_i}, \tfrac{\partial}{\partial x_j})\tfrac{\partial}{\partial x_k}, \quad \overline{R}(\tfrac{\partial}{\partial \hat{x}_i}, \tfrac{\partial}{\partial \hat{x}_j})\tfrac{\partial}{\partial \hat{x}_k} = \hat{R}(\tfrac{\partial}{\partial \hat{x}_i}, \tfrac{\partial}{\partial \hat{x}_j})\tfrac{\partial}{\partial \hat{x}_k},$$

with obvious notation. Thus, $\overline{R} = R + \hat{R}$ in the obvious sense. Taking traces, we se that the Ricci tensors similarly satisfy

$$\overline{\mathrm{ric}} = \mathrm{ric} + \hat{\mathrm{ric}}.$$

In particular, the product of two manifolds with negative Ricci also has negative Ricci. This should be contrasted with Preissmann's theorem.

10 Klingenberg's lemma and the cut locus

In this section we introduce the cut locus for an arbitrary complete Riemannian manifold. No assumptions will be made on the curvature. We will partly follow [4], to work around a couple of problems in the presentation of [7].

On a complete manifold, having fixed a point p, every point is reached by some segment starting at p. Define the segment domain as the set

$$\mathrm{seg}(p) = \{v \in T_p M \mid [0,1] \ni t \to \exp tv \text{ is a segment}\},$$

or equivalently

$$\operatorname{seg}(p) = \{v \mid r(\exp_p(v)) = |v|\}.$$

The map $\exp_p\colon \operatorname{seg}(p) \to M$ is surjective, but generally not injective.

It is easy to see that $\operatorname{seg}(p)$ is star shaped; its intersection with a line through the origin is an interval, whose extrema are called cut points. In other words, $\exp(v)$ is the cut point of the geodesic $t \mapsto \exp_p(tv)$ if the geodesic $\exp_p tv$ is minimizing up to $t = 1$ but not after t. The set of all cut points is called the *cut locus.*

Remark 10.1. If M has finite diameter, e.g. if M is compact, then every geodesic has a cut point. Otherwise, a geodesic may be minimizing up to $+\infty$.

Lemma 10.2. *A geodesic γ has a cut point if and only if there exists a value of t for which either*

1. *$\gamma(t)$ is conjugate to $\gamma(0)$ along γ; or*
2. *there is another segment $\sigma \neq \gamma$ from $\gamma(0)$ to $\gamma(t)$.*

The cut point is $\gamma(t_0)$, where t_0 is the minimal t for which (1) or (2) hold.

Proof. Write $\gamma(t) = \exp_p tv$. Suppose that $\gamma(t_0) = \exp_p t_0 v$ is a cut point; we must prove that (1) or (2) hold. Suppose that $\gamma(t_0)$ is not a conjugate point; then $t_0 v$ is a regular point for $\exp_p\colon T_pM \to M$. Consider the sequence $v_n = (t_0 + 1/n)v$, which lies outside the segment domain; for each v_n, choose w_n in the segment domain such that $\exp_p(w_n) = \exp_p(v_n)$. Up to extracting a subsequence, we can assume $w_n \to w$. Then by continuity $\exp_p(t_0 v) = \exp_p(w)$. Now since $t_0 v$ is a regular point, $\exp_p$ is injective around v, so if we had $t_0 v = w$ then $w_n = v_n$ for large n, which is absurd. Thus, we have obtained an element $w \neq t_0 v$ such that $\exp_p(t_0 v) = \exp_p(w)$. Since the segment domain is closed, w is in the segment domain. Thus, $|t_0 v| = |w|$ and (2) holds.

To prove minimality of t_0, suppose that $t_1 < t_0$ satisfies (1) or (2). If (2) holds, we can construct a segment from $\gamma(0)$ to $\exp v$ by adjoining σ to $\gamma|_{[t_1,t_0]}$; since segments are smooth geodesics, this has to coincide with γ, which is absurd. If (1) holds, we can show that γ is not a segment by constructing a variation $\gamma_s\colon [0, t_0] \to M$ as follows. Let J be a nonzero Jacobi field on $[0, t_1]$ with $J(0) = 0$, $J(t_1) = 0$; we extend J to a piecewise smooth vector field by declaring it to be zero after t_1. Let Y be a parallel vector field along γ such that $Y(t_1) = -J'(t_1)$. Let $\phi\colon [0, t_0] \to \mathbb{R}$ be a smooth function that vanishes at the boundary and is identically 1 in a neighbourhood of t_1, and let $X_\epsilon = J + \epsilon\phi Y$. Consider the proper variation $\gamma_s = \exp_{\gamma(t)} sX_\epsilon(t)$; then

$\frac{\partial \gamma_s}{\partial s}|_{s=0} = X_\epsilon$ and $\frac{\partial^2 \gamma_s}{\partial s^2} = 0$. Then Synge's formula gives

$$\begin{aligned}
&\frac{\partial^2}{\partial s^2}|_{s=0} E(\gamma_s) \\
&\quad = \int_0^{t_0} \left|\frac{\partial^2 \gamma_s}{\partial s \partial t}\right|^2 dt - \int_0^{t_0} \mathrm{Rm}(\frac{\partial \gamma_s}{\partial s}, \frac{\partial \gamma}{\partial t}, \frac{\partial \gamma}{\partial t}, \frac{\partial \gamma_s}{\partial s}) dt + g(\frac{\partial^2 \gamma_s}{\partial s^2}, \frac{\partial \gamma}{\partial t})|_0^{t_0} \\
&\qquad + g(\frac{\partial^2 \gamma_s}{\partial s^2}(t_1), \frac{\partial \gamma}{\partial t}|_{t=t_1^+} - \frac{\partial \gamma}{\partial t}|_{t=t_1^-}) \\
&\quad = \int_0^{t_0} |X_\epsilon'|^2 dt - \int_0^{t_0} \mathrm{Rm}(X_\epsilon, \gamma', \gamma', X_\epsilon) dt \\
&= \int_0^{t_1} (|J'|^2 + 2\epsilon g(J', \phi' Y)) dt - \int_0^{t_1} \mathrm{Rm}(J, \gamma', \gamma', J) dt - 2\epsilon \int_0^{t_1} \mathrm{Rm}(J, \gamma', \gamma', \phi Y) dt + o(\epsilon) \\
&= \int_0^{t_1} (-g(J'', J) - \mathrm{Rm}(J, \gamma', \gamma', J)) dt + 2\epsilon \int_0^{t_1} g(J', \phi' Y)) dt + 2\epsilon \int_0^{t_1} g(J'', \phi Y) dt + o(\epsilon) \\
&\qquad = 2\epsilon g(J', \phi Y)_0^{t_1} + o(\epsilon) = -2\epsilon |J'(t_1)|^2 + o(\epsilon),
\end{aligned}$$

which is negative for small $\epsilon > 0$. This shows that γ does not minimize energy, hence it is not a segment.

Thus, if a cut point $\gamma(t_0)$ exists, then t_0 is the minimal t for which (1) or (2) hold. Conversely, if there is some $t = t_0$ satisfying (1) or (2) then, arguing as above, γ is not minimizing after that value of t, so a cut point must exist. □

Example 10.3. For the round sphere, having fixed p, the cut locus is $\{-p\}$, which satisfies both condition 1 and 2 in Lemma 10.2.

Example 10.4. For the flat torus, the cut locus is a rectangle. In this case there are no conjugate points, so each cut point satisfies only condition 2.

Since the conditions of Lemma 10.2 are symmetric between the points 0 and t_0, we obtain:

Proposition 10.5. *If $\gamma(t_0)$ is a cut point of γ, then $\gamma(0)$ is a cut point of $t \mapsto \gamma(t_0 - t)$.*

Proposition 10.6. *The function*

$$F\colon \{v \in T_pM \mid |v| = 1\} \to \mathbb{R}_+ \cup \{+\infty\}, \quad F(v) = \begin{cases} t_0 & \exp_p tv \text{ has a cut point at } t_0 \\ +\infty & \text{otherwise} \end{cases}$$

is continuous.

Proof. Let $\{v_n\}$ be a sequence in T_pM, $|v_n| = 1$, with limit v, and let $F(v_n) = t_n$, where possibly $t_n = +\infty$. We must show that t_n converges to $F(v)$, and we will do so by showing

$$\limsup t_n \leq F(v) \leq \liminf t_n.$$

For every $\epsilon > 0$, we have a subsequence t_{n_k} such that $t_{n_k} \geq t = \limsup t_n - \epsilon$. This implies that the points tv_{n_k} are in the segment domain, and so is the limit tv_n; therefore, $F(v) \geq t$, and since ϵ is arbitrary, $F(v) \geq \limsup t_n$.

It remains to show that $F(v) \leq \liminf t_n$.

If $\liminf t_n = +\infty$, there is nothing to prove. Otherwise up to a subsequence we may assume that the $\{t_n\}$ are finite and they converge to $\bar{t} = \liminf t_n$. If, up to taking a further subsequence, the $t_n v_n$ are conjugate points, then $d\exp_p$ is singular at $t_n v_n \to \bar{t}v$, so $F(v) \leq \bar{t}$ and we are done.

Otherwise, we can assume that each $\exp_p t_n v_n$ is joined to 0 by another segment, i.e. there is $t_n w_n \neq t_n v_n$ in the segment domain with $\exp_p(t_n w_n) = \exp_p(t_n v_n)$. Up to a subsequence, $t_n w_n$ converges to $\bar{t}w$, with $|w| = 1$. If $\bar{t}v$ is a critical point for $\exp_p$, then $\exp_p(\bar{t}v)$ is in the cut locus, and $F(v) \leq \bar{t}$. Otherwise, $\exp_p$ is injective around $\bar{t}v$, so we cannot have $v = w$, for otherwise $w_n = v_n$ eventually. This shows that there are two segments joining p to $\bar{t}v$, so again $F(v) \leq \bar{t}$. □

Corollary 10.7. *The injectivity radius at p is the minimal distance from p to its cut locus. In particular, if the injectivity radius at p is finite, then there is a cut point of minimal distance to p.*

Proof. Let ρ be the minimal distance from p to its cut locus (i.e. the minimum of the function F defined in Proposition 10.6), where possibly $\rho = +\infty$.

Then geodesics of length $< \rho$ through p have no conjugate points, so $d\exp_p$ is nonsingular on the ball $B_\rho(0) \subset T_pM$. It is also injective: since the ball is in the segment domain, if $\exp_p(v)$ and $\exp_p(w)$ coincide, then $|v| = |w|$, so $v = w$, for otherwise $\exp_p(v)$ would be joined to 0 by two different segments, and it would have to be in the cut locus.

For the second part, observe that there is at least one cut point, and take the minimum of the function F. □

Remark 10.8. The cut point of minimal distance is not unique: consider the torus as a quotient of $[-1, 1] \times [-1, 1]$ with the usual identifications and the flat metric. Then the geodesic $\gamma(t) = (t, 0)$ has a cut point at $\gamma(1) = (1, 0)$, and the geodesic $\alpha(t) = (0, t)$ has a cut point at $\alpha(1) = (0, 1)$.

The following result will mostly be used in Section **??**.

Lemma 10.9 (Klingenberg)**.** *Choose $p \in M$ such that inj_p is finite, and let $\exp_p v$ be a cut point of minimal distance to p. Then at least one of the following holds:*

1. *there exists exactly one $w \neq v$ in T_pM such that $t \mapsto \exp tw$ is also a segment from p to $\exp_p v = \exp_p w$; in this case, the two segments join smoothly at $\exp v$;*
2. *$\exp_p(v)$ is a critical value for $\exp_p\colon \mathrm{seg}(p) \to M$.*

Proof. Suppose that v is a regular point for $\exp_p\colon T_pM \to M$. Then there are two segments joining p to $\exp_p v$, i.e. $q = \exp_p v = \exp_p w$ with w in $\mathrm{seg}(p)$.

If q is a critical value for $\exp_p$, (2) holds. Suppose otherwise. We have two distinct segments from p to q, call them $\gamma_v(t) = \exp_p tv$, $\gamma_w = \exp_p tw$. We must prove that that they concatenate smoothly at q, i.e. $\gamma_v'(1) = -\gamma_w'(1)$. Suppose for a contradiction that $\gamma_v'(1) \neq -\gamma_w'(1)$. Then there exists a vector $u \in T_qM$ which forms an angle $> \pi/2$ with both $\gamma_v'(1)$ and $\gamma_w'(1)$. Let $c(s)$ be a curve with $c(0) = q$, $c'(0) = u$. Since v, w are regular points for $\exp_p$, there exist curves $v(s)$, $w(s)$ with $\exp_p v(s) = c(s)$, $\exp_p w(s) = c(s)$, $v(0) = v$, $w(0) = w$. We can apply Gauss' lemma to the radial function $r = \sqrt{x_1^2 + \ldots + x_n^2}$ (which defines distance from p within the geodesic ball); then $\operatorname{grad} r = \frac{\partial}{\partial r}$ at v and w. Since $g(v'(0), \frac{\partial}{\partial r}) < 0$, $g(w'(0), \frac{\partial}{\partial r}) < 0$, both $v(s)$ and $w(s)$ enter the geodesic ball; this contradicts injectivity of $\exp_p$ in the geodesic ball.

It remains to show that w is unique. This follows because w is determined by where the geodesic $\exp_p tv$ reenters the geodesic ball. $\square$

11 The distance function ([7])

It is clear that the distance function $r(q) = d(p, q)$ is not C^1 at p. Similarly, one can see that it fails to be C^1 at cut points. Indeed, segments from p are integral curves of $\operatorname{grad} r$ that contain p; if $\operatorname{grad} r$ is continuous at q, any segment from p to q can be extended, so q is not a cut point. At other points, r turns out to be smooth:

Lemma 11.1. *Choose $p \in M$ and let $r(x) = d(x, p)$. Then r is smooth outside the cut locus and $\{p\}$.*

Proof. Let $q \neq p$ be a point outside the cut locus; let $v \mapsto \exp_p tv$ be the segment from p to $q = \exp v$. The continuity of the map F of Proposition 10.6 implies that $\exp_p tw$ does not have cut points at $t \leq 1$ for w in a neighbourhood of v. Therefore, $r(\exp w) = |w|$. In addition, $\exp_p$ is a diffeomorphism around v because q is not a cut point. Therefore, r is smooth around q. $\square$

We will generalize the situation and replace r with an arbitrary *distance function*, i.e. a continuous function $r \colon M \to \mathbb{R}$ which is smooth on some open set of M where it satisfies $|\operatorname{grad} r| = 1$. We will also write $\frac{\partial}{\partial r}$ instead of $\operatorname{grad} r$, consistently with the Gauss Lemma applied to the "ordinary" distance function $r(x) = d(x, p)$.

Let S be the $(1, 1)$ tensor defined by

$$g(S(X), Y) = \operatorname{Hess} r(X, Y),$$

i.e. $S(X) = \nabla_X \frac{\partial}{\partial r}$. Then we will write

$$g(S^2(X), Y) = \operatorname{Hess}^2 r(X, Y).$$

Notice that S is self-adjoint, hence

$$g(S^2(X), Y) = g(S(X), S(Y)) = g(\nabla_X \tfrac{\partial}{\partial r}, \nabla_Y \tfrac{\partial}{\partial r}).$$

Proposition 11.2. *Let r be a distance function on (M, g). Then*

$$\nabla_{\frac{\partial}{\partial r}} \operatorname{Hess} r(X, Y) + \operatorname{Hess}^2 r(X, Y) = -R(X, \tfrac{\partial}{\partial r}, \tfrac{\partial}{\partial r}, Y).$$

Proof. Since $\left|\frac{\partial}{\partial r}\right| = 1$, we have

$$\operatorname{Hess}(\tfrac{\partial}{\partial r}, X) = \operatorname{Hess} r(X, \tfrac{\partial}{\partial r}) = g(\nabla_X \tfrac{\partial}{\partial r}, \tfrac{\partial}{\partial r}) = 0.$$

In other words, $\nabla_{\frac{\partial}{\partial r}}\frac{\partial}{\partial r}=0$. Thus,

$$\begin{aligned}R(X,\tfrac{\partial}{\partial r})\tfrac{\partial}{\partial r}&=\nabla_X\nabla_{\frac{\partial}{\partial r}}\tfrac{\partial}{\partial r}-\nabla_{\frac{\partial}{\partial r}}\nabla_X\tfrac{\partial}{\partial r}-\nabla_{[X,\frac{\partial}{\partial r}]}\tfrac{\partial}{\partial r}\\&=-\nabla_{\frac{\partial}{\partial r}}(S(X))-S([X,\tfrac{\partial}{\partial r}])\\&=-(\nabla_{\frac{\partial}{\partial r}}S)(X)-S(\nabla_{\frac{\partial}{\partial r}}X)-S([X,\tfrac{\partial}{\partial r}])\\&=-(\nabla_{\frac{\partial}{\partial r}}S)(X)-S(\nabla_X\tfrac{\partial}{\partial r})=-(\nabla_{\frac{\partial}{\partial r}}S)(X)-S^2(X).\end{aligned}$$

Contracting with Y, and using that musical isomorphisms commute with ∇, we obtain the statement. □

On a Riemannian manifold, let vol be the volume form. We define the divergence of a vector field by

$$(\operatorname{div}X)\mathrm{vol}=\mathcal{L}_X\mathrm{vol}.$$

In an orthonormal frame $\{e_i\}$, we have

$$\operatorname{div}X=-\sum_i g(\mathcal{L}_Xe_i,e_i).\tag{18}$$

Then one defines the Laplacian by

$$\Delta f=\operatorname{div}\operatorname{grad}f.$$

By (18), we have

$$\begin{aligned}\Delta f=-\sum_i g([\operatorname{grad}f,e_i],e_i)&=-\sum_i g(\nabla_{\operatorname{grad}f}e_i,e_i)+g(\nabla_{e_i}\operatorname{grad}f,e_i)\\&=\sum_i g(\nabla_{e_i}\operatorname{grad}f,e_i)=\operatorname{tr}\operatorname{Hess}f.\end{aligned}$$

Proposition 11.3. *Let r be a distance function on (M,g). If M has dimension n, on the open set where r is smooth,*

$$\tfrac{\partial}{\partial r}\Delta r+\frac{(\Delta r)^2}{n-1}\leq\tfrac{\partial}{\partial r}\Delta r+|\operatorname{Hess}r|^2=-\operatorname{ric}(\tfrac{\partial}{\partial r},\tfrac{\partial}{\partial r}).$$

If equality holds, then $\operatorname{Hess}r(X,X)=\frac{1}{n-1}\Delta r$ *for any unit X orthogonal to $\frac{\partial}{\partial r}$.*

Proof. We have

$$\nabla_{\frac{\partial}{\partial r}}\operatorname{Hess}r(X,Y)+\operatorname{Hess}^2r(X,Y)=-R(X,\tfrac{\partial}{\partial r},\tfrac{\partial}{\partial r},Y).$$

We can rewrite this as

$$\nabla_{\frac{\partial}{\partial r}}\operatorname{Hess}r(X,Y)+g(\nabla_X\tfrac{\partial}{\partial r},\nabla_Y\tfrac{\partial}{\partial r})=-R(X,\tfrac{\partial}{\partial r},\tfrac{\partial}{\partial r},Y).$$

Fixing an orthonormal frame $e_1,\dots,e_n$ and contracting, we obtain

$$\sum_i\nabla_{\frac{\partial}{\partial r}}\operatorname{Hess}r(e_i,e_i)+g(\nabla_{e_i}\tfrac{\partial}{\partial r},\nabla_{e_i}\tfrac{\partial}{\partial r})=-\sum_i R(e_i,\tfrac{\partial}{\partial r},\tfrac{\partial}{\partial r},e_i)=-\operatorname{ric}(\tfrac{\partial}{\partial r},\tfrac{\partial}{\partial r}).$$

i.e.

$$\tfrac{\partial}{\partial r}\Delta r + \left|\nabla \tfrac{\partial}{\partial r}\right|^2 = -\operatorname{ric}(\tfrac{\partial}{\partial r}, \tfrac{\partial}{\partial r}).$$

Since $\nabla\frac{\partial}{\partial r} = \operatorname{Hess} r$, we are reduced to showing

$$\left|\operatorname{Hess} r\right|^2 \geq \frac{1}{n-1}(\Delta r)^2. \tag{19}$$

To this end, complete $\frac{\partial}{\partial r} = e_1$ to an orthonormal frame $e_1, \dots, e_n$, and write

$$\begin{aligned}\left|\operatorname{Hess} r\right|^2 = \sum_{i,j}\operatorname{Hess}(e_i, e_j)^2 &= \sum_{i>1, j>1}\operatorname{Hess}(e_i, e_j)^2\\ &\geq \frac{1}{n-1}\big(\sum_{i>1}\operatorname{Hess}(e_i, e_i)\big)^2 = \frac{1}{n-1}(\Delta r)^2.\end{aligned}$$

where we have used Cauchy-Schwartz on square matrices of order $n-1$ in the form

$$\operatorname{tr} A^2 = g(A, I)^2 \leq |A|^2\,|I|^2 = |g|^2\,(n-1).$$

If equality holds in (19), then equality holds in the Cauchy-Schwarz inequality, i.e. A and I are linearly dependent. $\square$

We aim at proving an estimate for the Laplacian of the distance function. The idea is to consider the solutions of $\psi'' + k\psi = 0$ which we called S_k, i.e.

$$S_k(t) = \begin{cases} \frac{1}{\sqrt{k}}\sin\sqrt{k}\,t & k > 0\\ t & k = 0\\ \frac{1}{\sqrt{-k}}\sinh\sqrt{-k}\,t & k < 0\end{cases}.$$

In particular, when $k > 0$ S_k vanishes at $t = \pi/\sqrt{k}$.

Notice that $\rho = S_k'/S_k$ satisfies

$$\rho' + \rho^2 + k = 0.$$

For the following lemma, we adapt the proof of [6].

Lemma 11.4 (Riccati comparison estimate)**.** *Let $\rho\colon (0,b) \to \mathbb{R}$ be a smooth function and $k \in \mathbb{R}$. Then:*

1. *if $\rho' + \rho^2 + k \leq 0$, then $\rho(t) \leq \frac{S_k'(t)}{S_k(t)}$ where defined (which implies $b < \pi/\sqrt{k}$ for $k > 0$);*
2. *if $\rho' + \rho^2 + k \geq 0$ and $\lim_{t\to 0}\rho(t) = +\infty$, then $\rho(t) \geq \frac{S_k'(t)}{S_k(t)}$ where defined (which entails $t < \pi/\sqrt{k}$ for $k > 0$).*

Proof. Let τ be a function such that $\tau' + \tau^2 + k = 0$. Let F be a primitive of $\rho + \tau$. Then

$$\frac{d}{dt}(\rho - \tau)e^F = (\rho^2 - \tau^2 + \rho' - \tau')e^F = (\rho^2 + \rho' + k)e^F.$$

If $\rho' + \rho^2 + k \leq 0$, then $(\rho - \tau)e^F$ is nonincreasing; it follows that if $\rho(t_0) > \tau(t_0)$, then $\rho > \tau$ for $t \leq t_0$. Similarly, if $\rho' + \rho^2 + k \geq 0$ and $\rho(t_0) < \tau(t_0)$, then $\rho < \tau$ for $t \leq t_0$.

Now suppose for a contradiction that $\rho(t_0) > \frac{S_k'(t_0)}{S_k(t_0)}$ for some t_0. By continuity, there exists $\epsilon > 0$ such that $\rho(t_0) > \frac{S_k'(t_0-\epsilon)}{S_k(t_0-\epsilon)}$. Applying the above observation with $\tau(t) = \frac{S_k'(t-\epsilon)}{S_k(t-\epsilon)}$, we see that

$$\rho(t) > \frac{S_k'(t-\epsilon)}{S_k(t-\epsilon)}, \quad \epsilon < t \leq t_0.$$

Taking the limit for $t \to \epsilon^+$, we see that ρ must go to infinity at $t = \epsilon$, which is absurd.

The second part is similar. Suppose for a contradiction that $\rho(t_0) < \frac{S_k'(t_0)}{S_k(t_0)}$ for some t_0. By continuity, there exists $\epsilon > 0$ such that $\rho(t_0 - \epsilon) < \frac{S_k'(t_0)}{S_k(t_0)}$. Then arguing as above we find $\rho(t-\epsilon) < \frac{S_k'(t)}{S_k(t)}$ for all $\epsilon < t < t_0$; since we are assuming that $\rho \to +\infty$ as $t \to 0$, this is a contradiction.

The parenthetical remarks in the case $k > 0$ follow from the fact that S_k'/S_k is only defined for $t < \pi/\sqrt{k}$, and $\lim_{t\to\pi/\sqrt{k}} S_k'/S_k = -\infty$. In the case $\rho' + \rho^2 + k \leq 0$, this implies that $b < 0$. In the case $\rho' + \rho^2 + k \geq 0$, this implies that the inequality must be restricted to $t < \pi/\sqrt{k}$. $\square$

We now go back to following [7].

Lemma 11.5 (Calabi)**.** *If (M, g) is complete and* $\mathrm{Ric} \geq (n-1)k$*, with $k \in \mathbb{R}$, then the distance function $r(x) = d(x, p)$ satisfies*

$$\Delta r(x) \leq (n-1)\frac{S_k'(r(x))}{S_k(r(x))}$$

at points x where r is smooth.

Proof. Let $\gamma(t) = \exp_p tv$ be a unit speed segment, so that $r(\gamma(t)) = t$. Let $\rho(t) = \frac{1}{n-1}\Delta r(\gamma(t))$.

By Proposition 11.3, $\rho' + \rho^2 + k \leq 0$, and we can apply the Riccati comparison estimate. Notice that when $k > 0$, r cannot be smooth after $t = \pi/\sqrt{k}$, and we recove Myers' theorem. $\square$

12 Subharmonic functions and the Calabi lemma at the cut locus

In this section we generalize Calabi's lemma so that it applies at points where x is not smooth. The first step is making sense of the inequality in the statement at points where Δr is not defined. The starting observation is the following:

Lemma 12.1. *If f, h are smooth functions on M such that $f(p) = h(p)$ and $f(x) \geq h(x)$ in a neighbourhood of p, then*

$$df_p = dh_p, \quad \mathrm{Hess}\, f_p \geq \mathrm{Hess}\, h|_p, \quad \Delta f(p) \geq \Delta h(p).$$

Proof. By linearity, we can assume $h = 0$. Consider a unit geodesic γ through p. Then $f \circ \gamma$ has a local minimum at 0, hence

$$(f\circ\gamma)'(0) = 0, \quad (f\circ\gamma)''(0) \geq 0.$$

This implies that $df_p(\gamma'(0)) = 0$ and $\mathrm{Hess}(\gamma'(0), \gamma'(0)) \geq 0$. By varying γ, we see that $df_p = 0$, $\mathrm{Hess}\, f|_p \geq 0$; taking the trace, we also see that $\Delta f(p) \geq 0$. $\square$

This motivates the following definition. Given a symmetric bilinear form B on T_pM, a smooth function satisfies $\mathrm{Hess}\, f_p \geq B$ if $\mathrm{Hess}\, f_p - B$ is positive semidefinite. we say that a *continuous* function f satisfies $\mathrm{Hess}\, f|_p \geq B$ if for every $\epsilon > 0$ there exists a smooth function f_ϵ defined on a neighbourhood of p, called a *support function from below*, such that

1. $f_\epsilon(p) = f(p)$;
2. $f_\epsilon \leq f$ in some neighbourhood of p;
3. $\mathrm{Hess}\, f_\epsilon|_p \geq B - \epsilon g_p$.

It follows from Lemma 12.1 that when f is smooth, this is equivalent to $\mathrm{Hess}\, f_p - B$ being positive semidefinite.

We say that $\mathrm{Hess}\, f|_p > B$ if $\mathrm{Hess}\, f|_p \geq B'$ for some $B' > B$, i.e. if there is an ϵ such that $\mathrm{Hess}\, f|_p \geq B + \epsilon g$.

Inverting the inequalities, we similary define support functions from above and write $\mathrm{Hess}\, f|_p \leq B$ when analogous conditions hold. We then write $\mathrm{Hess}\, f|_p = B$ to mean that $\mathrm{Hess}\, f|_p \geq 0$ and $\mathrm{Hess}\, f_p \leq 0$.

We give analogous definitions for $\Delta f(p)$ (replace B with a scalar and Hess with Δ). A continuous function on (M, g) is *subharmonic* if $\Delta \geq 0$, and *superharmonic* if $\Delta \leq 0$. We will need the following version of the maximum principle:

Theorem 12.2 (Maximum principle)**.** *If $f\colon M \to \mathbb{R}$ is continuous and subharmonic, then f is constant in a neighbourhood of every local maximum. In particular, if N is connected and f has a global maximum, then it is constant.*

Proof. Let p be a local maximum. If $\Delta f(p) > 0$, we obtain a contradiction because there is a support function from below f_ϵ with $\Delta f_\epsilon(p) > 0$, which also has a local maximum at p, against Lemma 12.1.

Now let $B_r(p)$ be a geodesic ball around p, on which p is a maximum point. Suppose for a contradiction that f is not constant on $B_r(p)$. Thus, there is some q such that $f(q) \neq f(p)$; up to decreasing r, we can assume that q lies in the boundary $\partial B_r(p)$ of the geodesic ball. Set

$$V = \{x \in \partial B_r(p) \mid f(x) = f(p)\};$$

notice that q does not lie in V.

We will obtain a contradiction by considering a function $\overline{f} = f + \delta h$ with positive Laplacian and a local maximum, which contradicts what we proved at the beginning. Here $\delta > 0$ is a constant and h is a smooth function on $\overline{\partial} B_p(r)$ such that

$$\begin{cases} h < 0 \text{ on } V, \\ h(p) = 0, \\ \Delta h > 0 \text{ on } \overline{B}_p(r). \end{cases} \tag{20}$$

In order to construct h, let ϕ be a smooth function on $\overline{B}_r(p)$ such that

$$\phi(p) = 0, \quad \phi < 0 \text{ on } V, \quad d\phi \neq 0.$$

For instance, if in normal coordinates $q = \exp_p(1, 0, \dots, 0)$, we can set

$$\phi = \epsilon x_1 - (x_2^2 + \ldots + x_n^2)$$

for small $\epsilon > 0$. Then we set $h = e^{\alpha\phi} - 1$, with α a positive constant. The first two conditions in (20) are trivially satisfied; for the third, we compute

$$\operatorname{Hess} f = \nabla dh = \nabla(\alpha e^{\alpha\phi} d\phi) = \alpha^2 e^{\alpha\phi} d\phi \otimes d\phi + \alpha e^{\alpha\phi} \operatorname{Hess} \phi;$$

taking the trace,

$$\Delta f = \alpha e^{\alpha\phi}(\alpha \left|d\phi\right|^2 + \Delta\phi),$$

which for large α is positive on the compact set $\overline{B}_r(p)$.

Now let

$$\overline{f} \colon \overline{\partial} B_p(r) \to \mathbb{R}, \quad \overline{f} = f + \delta h.$$

This function has a maximum because it is defined on a compact set; for small $\delta > 0$, $\overline{f}$ has a maximum in the interior of the geodesic ball. This is because $f \leq f(p) - \epsilon$ on the compact set

$$K = \{x \in \partial B_r(p) \mid h(x) \geq 0\},$$

and so we have $\overline{f} \leq f(p)$ on K if δ is small enough; since $\overline{f} \leq f \leq f(p)$ on $\partial B_r(p) \setminus K$, this shows that $\overline{f} \leq f(p) = \overline{f}(p)$ on $\partial B_r(p)$.

On the other hand, we have $\Delta \overline{f} > 0$: indeed, let f_ϵ be a support function from below for f at q; then $\Delta f_\epsilon(q) \geq -\epsilon$. The function $f_\epsilon + h\delta$ is a support function for $\overline{f} = f + h\delta$, and

$$(\Delta f_\epsilon + h\delta)(q) \geq -\epsilon + \delta \Delta h(q),$$

which for small ϵ becomes positive.

Now $\overline{f}$ has a local maximum on $B_r(p)$ and positive Laplacian, which is a contradiction by the first part of the proof. □

Theorem 12.3. *If $f \colon M \to \mathbb{R}$ is continuous and $\Delta f = 0$, then f is smooth.*

Proof. Fix $p \in M$ and choose $\Omega = \overline{B_r(p)}$ contained in a normal neighbourhood of p. By a standard result in PDE, the Dirichlet boundary value problem

$$\Delta u = 0, \quad \Delta|_{\partial\Omega} = f|_{\partial\Omega}$$

has a solution, which is continuous on Ω and smooth in the interior of Ω. By the maximum principle, $f - u$ and $u - f$ attain their maximum value on the boundary of Ω, so they are both zero. □

Lemma 12.4 (Calabi)**.** *If (M, g) is complete and* $\operatorname{Ric} \geq (n-1)k$*, with $k \in \mathbb{R}$, then the distance function $r(x) = d(x, p)$ satisfies*

$$\Delta r(x) \leq (n-1)\frac{S_k'(r(x))}{S_k(r(x))}$$

at every x such that $S_k(r(x)) > 0$.

Proof. For $q \in M$, let σ be a unit speed segment from p to q, $\sigma(L) = q$. For $0 < \epsilon < L$, write $r_\epsilon(x) = \epsilon + d(\sigma(\epsilon), x)$; then $r_\epsilon(q) = L$, and

$$r_\epsilon(x) = d(p, \sigma(\epsilon)) + d(\sigma(\epsilon, x)) \geq r(x).$$

We claim that the r_ϵ are smooth. Suppose for a contradiction that for some ϵ the function r_ϵ is not smooth at q. Then q is a cut point for the geodesic from $\sigma(\epsilon)$ to q. By Proposition 10.5, this implies that $\sigma(\epsilon)$ is a cut point for the geodesic from q to $\sigma(\epsilon)$, which is absurd because σ is a segment from p to q.

Thus, the r_ϵ are smooth, and they are support functions from above for r at q. By Lemma 11.5, we have

$$\Delta r_\epsilon(q) \leq (n-1)\frac{S_k'(d(q,\sigma(\epsilon)))}{S_k(d(q,\sigma(\epsilon)))} = (n-1)\frac{S_k'(r(q)-\epsilon)}{S_k(r(q)-\epsilon)}.$$

If $S_k(r(q)) > 0$, the right hand side goes to $\frac{S_k'(r(q))}{S_k(r(q))}$ as $\epsilon \to 0$, and we obtain

$$\Delta r(q) \leq (n-1)\frac{S_k'(r(q))}{S_k(r(q))}.$$

□

Remark 12.5. For $x \neq p$, the function $S_k(r(x))$ can only vanish if $k > 0$ and $d(x,p) = \pi/\sqrt{k}$, which by the Myers theorem can only happen if the diameter is exactly $\pi/\sqrt{k}$ and $d(x,p) = \pi/\sqrt{k}$. We will see in Theorem **??** that this only happens when M is isometric to the round sphere of curvature k.

13 The Cheeger-Gromoll splitting theorem [7]

In this section we prove the Cheeger-Gromoll theorem, stating that a complete Riemannian manifold with $\mathrm{Ric} \geq 0$ that contains a line is a Riemannian product $M \times \mathbb{R}$. We start with a general discussion of rays and lines, which will be useful in applications of the theorem.

On a Riemannian manifold, a *ray* is a unit speed geodesic $\gamma\colon [0,\infty) \to M$ such that

$$d(\gamma(s), \gamma(t)) = |t - s|\,. \tag{21}$$

Thus, γ is unit speed and its restriction to any closed interval is a segment. We will say the ray *emanates* from $\gamma(0)$. A *line* is a unit speed geodesic $\gamma\colon (-\infty,\infty)$ that also satisfies (21).

Example 13.1. Let $\{g_t\}$ be a one parameter family of metrics on M. This induces a metric $g = g_t + dt^2$ on $M \times \mathbb{R}$, sometimes referred to as a *generalized cylinder*. If $x_1, \ldots, x_n$ are coordinates on M, then $x_1, \ldots, x_n, t$ are coordinates on $M \times R$ and the Koszul formula gives

$$g(\nabla_{\frac{\partial}{\partial t}} \tfrac{\partial}{\partial t}, \tfrac{\partial}{\partial x_i}) = 0, \quad g(\nabla_{\frac{\partial}{\partial t}} \tfrac{\partial}{\partial x_i}, \tfrac{\partial}{\partial x_j}) = \frac{1}{2}\tfrac{\partial}{\partial t} g_{ij}.$$

Thus, the curves $t \mapsto (t,p)$ are lines.

Example 13.2. A paraboloid $x^2 + y^2 = z$ does not contain any line. This is because all geodesics on the paraboloid are tangent to a parallel $\{z = z_0\}$ (with z_0 being the minimum height along the geodesic) and cross every meridian an infinite number of times. Indeed, write the geodesic as

$$\gamma(t) = (r\cos\theta, r\sin\theta, r^2),$$

where r, θ are functions of t. Then

$$\gamma'(t) = r'(\cos\theta, \sin\theta, 2r) + \theta'(-r\sin\theta, r\cos\theta, 0).$$

In particular

$$(r')^2(1 + 4r^2) + r^2(\theta')^2 = 1;$$

moreover Clairaut's relation gives

$$K \equiv r\langle\gamma', (-\sin\theta, \cos\theta, 0)\rangle = r^2\theta'.$$

Since r is not constant along the geodesic (parallels are not geodesics), we can reparametrize the geodesic in terms of r; then

$$\frac{\partial\theta}{\partial r} = \frac{Kr^{-2}}{\sqrt{(1 - K^2r^{-2})/(1 + 4r^2)}} = \frac{K}{r}\sqrt{\frac{1 + 4r^2}{r^2 - K^2}} \geq \frac{K}{r}.$$

This shows that θ is unbounded, and theferere a geodesic intersect every meridian an infinite number of times.

Since parallels are not geodesics, a geodesic cannot be asymptotic to a parallel; hence, at the minimum value of r along a geodesic, the geodesic is tangent to a parallel.

These two facts imply that any geodesic which is not a meridian meets itself an infinite number of times; hence, it is not a line.

We say that a manifold M is *connected at infinity* if for every compact subset $K \subset M$ there is a compact subset $C \supset K$ such that any two points in $M \setminus C$ can be joined by a curve in $M \setminus K$. Customarily, in the definition $M \setminus C$ is allowed to be disconnected. This seems to be an unnecessary complication, due to the following:

Lemma 13.3. *If M is connected at infinity, any compact set K is contained in a compact set C that does not disconnect M.*

Proof. Let K be an arbitrary compact set. Because M is connected at infinity, there exists a compact set $C \supset K$ such that any two points of $M \setminus C$ are joined by a curve in $M \setminus K$. Let $\{U_\alpha\}_{\alpha\in I}$ be the connected components of $M \setminus K$. Since any two points of $M \setminus C$ must lie in the same component of $M \setminus K$, we see that there is only one of the U_α that is not contained in C, call it U_β. Then U_β is open, because a manifold is locally connected; thus, $C' = C \setminus U_\beta$ is compact. Writing

$$C' = \bigcup_{\alpha\neq\beta} U_\alpha \cup K,$$

we see that $M \setminus C' = V_\beta$ is connected. □

We say that M is *disconnected at infinity* if it is not connected at infinity.

Remark 13.4. A compact manifold M is trivially connected at infinity, even if it not connected. Notice that the arbitrary compact subset $K \subset M$ in the definition can still disconnect M (consider the equator in the sphere).

Remark 13.5. A manifold with two noncompact connected components is trivially disconnected at infinity.

It follows from the two remarks above that the notion of connectedness at infinity is only meaningful for noncompact, connected manifolds.

Example 13.6. For $n > 1$, $\mathbb{R}^n$ is connected at infinity, since every compact K is contained in some closed ball $\overline{B_r(0)}$.

Example 13.7. If N is compact and connected, then $M = N \times \mathbb{R}$ is disconnected at infinity, since $N \times \{0\}$ divides N in two connected components.

Example 13.8. If N is compact and connected, then $N = M \times \mathbb{R}^2$ is connected at infinity: every compact K is contained in some $M \times \overline{B_r(0)}$, whose complementary is connected.

Lemma 13.9. *Let M be complete, connected and noncompact. Then:*

1. *for every p, there is a ray emanating from p;*
2. *if M is disconnected at infinity, then M contains a line.*

Proof. For the first part, let q_i be a sequence going to infinity, i.e. for every compact K, q_i is eventually outside K. Let γ_i be a sequence of unit speed segments from p to q_i, say $\gamma_i(t) = \exp_p tv_i$, with $|v_i| = 1$. By the compactness of the sphere in T_pM, we can assume up to extracting a subsequence that v_i converges to v; write $\gamma(t) = \exp tv$. By continuity of the exponential, we have

$$\lim_{i\to\infty} \gamma_i(t) = \gamma(t).$$

In particular, $d(\gamma(t), \gamma(s)) = \lim d(\gamma_i(t), \gamma_i(s)) = |t - s|$. Thus, γ is a ray.

For the second part, assume that M is disconnected at infinity. Then there is a compact set K such that for any compact $C \supset K$, there exist two points outside C that are not connected by a curve in $M \setminus K$. Consider an exhaustion by compact sets $M = \bigcup_{i\in\mathbb{N}} C_i$, and choose points a_i, b_i in $M \setminus C_i$ not connected by a curve in $M \setminus K$. Thus, the unit speed segment γ_i from a_i to b_i intersects K at some p_i; up to time translation, we can write $\gamma_i = \exp_{p_i} t_i v_i$, with $|v_i| = 1$. Up to a subsequence, we can assume that $(p_i; v_i)$ converges to $(p; v)$ with $p \in K$ and $|v| = 1$; then $\exp_{p_i}(tv_i)$ converges to $\exp_p(tv)$ for all t. Since the γ_i are segments on an interval of length $d(a_i, b_i) \to +\infty$, $t \mapsto \exp tv$ is a line. □

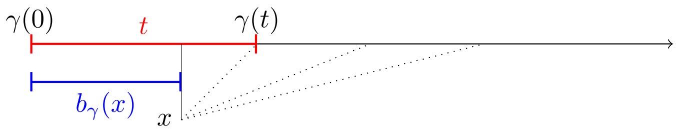
Let (M, g) be complete and noncompact. Let $\gamma\colon [0, \infty)$ be a ray, and define

$$b_t(x) = d(x, \gamma(t)) - t.$$

Remark 13.10. At points $x = \gamma(s)$, we have

$$b_t(\gamma(s)) = |t - s| - t = \begin{cases} -s & s \leq t \\ s - 2t & s \geq t \end{cases}.$$

Figure 1: Busemann function in $\mathbb{R}^n$



Lemma 13.11. *Let (M, g) be complete, let γ be a ray and define the functions b_t as above. Then:*

1. *for fixed x, the function $t \mapsto b_t(x)$ is decreasing and bounded in absolute value by $d(x, \gamma(0))$;*
2. *$|b_t(x) - b_t(y)| \leq d(x, y)$;*
3. *if $\mathrm{Ric} \geq 0$, $\Delta b_t(x) \leq \frac{n-1}{b_t(x)+t}$.*

Proof. The first item follows from the triangle inequality: if $s < t$, we have

$$b_t(x) - b_s(x) = d(x, \gamma(t)) - t - d(x, \gamma(s) + s \leq d(\gamma(t), \gamma(s) - (t - s) = 0,$$

and boundedness follows from

$$|b_t(x)| = |d(x, \gamma(t)) - t| = |d(x, \gamma(t)) - d(\gamma(0), \gamma(t)| \leq |d(x, \gamma(0)| \,.$$

The second item follows from

$$|b_t(x) - b_t(y)| = |d(x, \gamma(t)) - d(y, \gamma(t)| \leq d(x, y).$$

The third item is Calabi's lemma for $k = 0$. $\square$

We define the *Busemann function*

$$b_\gamma \colon M \to \mathbb{R}, \quad b_\gamma(x) = \lim_{t \to +\infty} b_t(x);$$

this is well defined by Lemma 13.11, and by the same satisfies

$$|b_\gamma(x) - b_\gamma(y)| \leq d(x, y), \quad |b_\gamma(x)| \leq d(x, \gamma(0)).$$

In particular, b_γ is continuous. It should be thought of as a "shifted" distance from infinity in the direction of the ray, as in

$$d(x, \gamma(+\infty)) = d(\gamma(0), +\infty) + b_\gamma(x) \quad \text{(ideally)}. \tag{22}$$

Remark 13.12. By Remark 13.10, $b_\gamma(\gamma(s)) = -s$.

Example 13.13. In Euclidean space,

$$b_\gamma(x) = (x - \gamma(0)) \cdot \gamma'(0),$$

(see Figure 1).

Given a ray γ and any point $p \in M$, we can choose a sequence $\gamma(t_n)$ such that the unit speed segments from p to $\gamma(t_n)$ converge to a ray $\tilde{\gamma}\colon [0, +\infty) \to M$ with $\tilde{\gamma}(0) = p$; such a ray (in general not unique) will be called an *asymptote* for γ from p.

Along an asymptote, the "distance from infinity" b_γ decreases in the same way as along γ; the triangle inequality then implies that, moving away from p, b_γ cannot increase more than $b_{\tilde{\gamma}}$:

Proposition 13.14. *If γ is a ray and $\tilde{\gamma}$ an asymptote for γ from p, then the Busemann functions are related by*

1. $b_\gamma(\tilde{\gamma}(t)) = b_\gamma(p) - t$;
2. $b_\gamma(x) - b_\gamma(p) \leq b_{\tilde{\gamma}}(x)$.

Proof. Let σ_n be the unit speed segments from p to $\gamma(t_n)$, converging to $\tilde{\gamma}$. Fix $s > 0$; for large n, $\sigma_n(s)$ lies in the segment from p to $\gamma(t_n)$, so

$$d(p, \gamma(t_n)) = d(p, \sigma_n(s)) + d(\sigma_n(s), \gamma(t_n)).$$

Then

$$\begin{aligned} b_\gamma(p) = \lim_{n\to+\infty} d(p, \gamma(t_n)) - t_n &= \lim_{n\to+\infty} d(p, \sigma_n(s)) + d(\sigma_n(s), \gamma(t_n)) - t_n \\ &= d(p, \tilde{\gamma}(s)) + \lim_{n\to+\infty} (d(\tilde{\gamma}(s), \gamma(t_n)) - t_n) \\ &\qquad\qquad\qquad = s + b_\gamma(\tilde{\gamma}(s)), \end{aligned}$$

proving the first item.

The second follows, as

$$d(x, \gamma(s)) - s \leq d(x, \tilde{\gamma}(t)) + d(\tilde{\gamma}(t), \gamma(s)) - s;$$

taking the limit as $s \to +\infty$,

$$b_\gamma(x) \leq d(x, \tilde{\gamma}(t)) + b_\gamma(\tilde{\gamma}(t)) = d(x, \tilde{\gamma}(t)) + b_\gamma(p) - t;$$

taking now the limit as $t \to +\infty$ we find $b_\gamma(x) \leq b_\gamma(p) + b_{\tilde{\gamma}}(x)$. $\square$

Lemma 13.15. *If* $\operatorname{Ric} \geq 0$*, then* $\Delta b_\gamma \leq 0$.

Proof. Let $\tilde{\gamma}$ be an asymptote from p. Then $b_{\tilde{\gamma}}(p) = 0$, and by Proposition 13.14 $b_\gamma(p) + b_{\tilde{\gamma}} \geq b_\gamma$. By Lemma 13.11 the functions $b_t(x)) = d(\tilde{\gamma}(t), x) - t$ converge to $b_{\tilde{\gamma}}(x)$ from above; they are smooth around p because $\tilde{\gamma}(t)$ is not a cut point, so $b_{\gamma(p)} + b_t$ is a support function for $b\gamma$ at p; again by Lemma 13.11, we have

$$\Delta b_t(p) \leq \frac{n-1}{b_t(p) + t} = \frac{n-1}{t},$$

so $\Delta b_\gamma \leq 0$. $\square$

Theorem 13.16 (Cheeger-Gromoll)**.** *If (M, g) is noncompact, complete, connected, it contains a line and* $\operatorname{Ric} \geq 0$*, then it is isometric to $H \times \mathbb{R}$ with a product metric.*

Proof. The proof makes use of a linear distance function r, i.e. a smooth function $r\colon M \to \mathbb{R}$ such that

$$|\operatorname{grad} r| = 1, \quad \operatorname{Hess} r = 0.$$

Having defined such a function, let $N = r^{-1}(0)$ with the induced metric, and let

$$\Phi\colon N \times \mathbb{R} \to M$$

be the map defined by the flow of $\operatorname{grad} r$. This is a diffeomorphism around a point $(p, 0)$, giving local coordinates $x_1, \dots, x_n, r$ on M, where $x_1, \dots, x_n$ are coordinates on N; in these coordinates, the coordinate field $\frac{\partial}{\partial r}$ coincides with $\operatorname{grad} r$. By construction, the metric takes the form of a generalized cylinder $g_r + dr^2$. Since $\nabla \frac{\partial}{\partial r} = \operatorname{Hess} r = 0$, $\nabla_{\frac{\partial}{\partial r}} \frac{\partial}{\partial x_i} = \nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial r} = 0$, so the computation of Example 13.1 shows that g_r is constant in r. Thus, Φ is an isometry about $(p, 0)$. Repeating the argument at different points (p', t) one sees that Φ is a global isometry.

In order to construct r, let γ be a line. Let $b^\pm$ be the Busemann functions associated to the positive and negative rays of γ, i.e.

$$b^+(x) = \lim_{t\to+\infty} d(x, \gamma(t)) - t, \quad b^-(x) = \lim_{t\to+\infty} d(x, (\gamma(-t)) - t.$$

By the triangle inequality,

$$b^+(x) + b^-(x) = \lim_{t\to+\infty} d(x, \gamma(t)) + d(x, \gamma(-t) - 2t \geq 0;$$

moreover at points of γ we have $b^+(\gamma(t)) + b^-(\gamma(t)) = 0$. Thus, $b^+ + b^-$ has a global minimum, and $\Delta(b^+ + b^-) \leq 0$; by the minimum principle it is constant. This implies

$$b^- = -b^+, \quad \Delta b^+ = 0 = \Delta b^-.$$

We need to show that $\operatorname{grad} b^\pm$ has unit norm; then the $b^\pm$ are linear distance function and we can conclude as above.

Fix $p \in M$. Let $\tilde\gamma^\pm$ be asymptotes from p for the positive and negative rays in γ. Set

$$b_t^\pm(x) = d(x, \tilde\gamma^\pm(t)) - t;$$

so that $b_t^\pm(x)$ converges to $b_{\tilde\gamma^\pm}(x)$ from above; then by Proposition 13.14 we have

$$b_t^+(x) \geq b_{\tilde\gamma^+}(x) \geq b^+(x) - b^+(p),$$
$$b_t^-(x) \geq b_{\tilde\gamma^-}(x) \geq b^-(x) - b^-(p),$$

since $b^+ = -b_-$, this gives

$$b_t^+(x) + b_t^-(x) \geq 0,$$

with equality at p. Since b_t^+ and b_t^- are smooth at p, we have $db_t^+(p) = -db_t^-(p)$, and

$$b_t^+(x) \geq b^+(x) - b^+(p) \geq -b_t^-(x)$$

implies that b^+ is differentiable at p with $db^+(p) = db_t^+(p)$; in particular, $|\operatorname{grad} b^+| = 1$ everywhere.

Moreover, $r = b^+$ is harmonic, so it is smooth. By Proposition 11.3, we have

$$0 = \partial_r \Delta r + \frac{(\Delta r)^2}{n-1} \leq \partial_r \Delta r + |\mathrm{Hess}\, r|^2 = -\operatorname{ric}(\tfrac{\partial}{\partial r}, \tfrac{\partial}{\partial r}) \leq 0,$$

so the Hessian is zero. □

The Cheeger-Gromoll splitting theorem has several interesting consequences.

Corollary 13.17. *$S^2 \times S^1$ and $S^3 \times S^1$ do not admit any Ricci-flat metric.*

Proof. Suppose for a contradiction that $S^k \times S^1$ has a Ricci-flat metric with $k = 2, 3$, and let g be the pullback metric on the universal covering $S^k \times \mathbb{R}$. Since $S^k \times \mathbb{R}$ is disconnected at infinity, it contains a line by Lemma 13.9. By the Cheeger-Gromoll theorem, $(S^k \times \mathbb{R}, g) = (H \times \mathbb{R}, h + dt^2)$, where (H, h) is a Ricci-flat manifold of dimension $k = 2, 3$. For dimensional reasons, h is a flat metric. It is also simply connected, because so is $H \times \mathbb{R} = S^k \times \mathbb{R}$, and complete. So by the classification of complete simply connected manifolds with constant curvature, $H = \mathbb{R}^2$, which gives a contradiction. □

A group of isometries of $\mathbb{R}^k$ is called a *crystallographic group* if it is discrete and acts cocompactly. Bieberbach's theorem states that the intersection of a crystallographic group Γ with the subgroup $\mathbb{R}^k \subset \mathrm{Isom}(\mathbb{R}^k)$ consisting of translations is normal and of finite index in Γ, and isomorphic to $\mathbb{Z}^k$ (see [8]).

We know from Myers' theorem that the universal covering of a complete manifold with $\mathrm{Ric} \geq \frac{n-1}{k}$ is compact. The following theorem is a generalization to $\mathrm{Ric} \geq 0$:

Theorem 13.18 (Structure theorem for $\mathrm{Ric} \geq 0$, Cheeger-Gromoll)**.** *Let (M, g) be a compact Riemannian manifold with* $\mathrm{Ric} \geq 0$. *Then:*

1. *the universal cover with the pullback metric is a Riemannian product $N \times \mathbb{R}^k$, where N is compact;*
2. *the isometry group of $N \times \mathbb{R}$ is* $\mathrm{Isom}(N) \times \mathrm{Isom}(\mathbb{R}^k)$*;*
3. *there exists a finite normal subgroup $G \subset \pi_1(M)$ such that $\pi_1(M)/G$ is a crystallographic group.*

Proof. By the splitting theorem, the universal cover is a Riemannian product $\tilde{M} = N \times \mathbb{R}^k$, where N does not contain any line. We will see that N is compact. First, we will prove (2).

Observe first that the projections of a line in $N \times \mathbb{R}^k$ on each factor is either constant or a line. We have a natural action of $\mathbb{R}^k$ on $N \times \mathbb{R}^k$ by translations, allowing us to write

$$(x, y) + v = (x, y + v), \quad x \in N, y, v \in \mathbb{R}^k.$$

Since N does not contain any line, the lines of $\tilde{M} = N \times \mathbb{R}^k$ take the form

$$\gamma(t) = (x, y + tv) = (x, y) + tv, \quad v \in \mathbb{R}^k. \tag{23}$$

Let $F \colon \tilde{M} \to \tilde{M}$ be an isometry. Then F maps a line as in (23) to a line of the same form. In particular, $(F \circ \gamma')$ is the velocity of a line, so it has no component along N, and we can write

$$F(x, y + tv) = F(x, y) + t dF_{x,y}(0, v);$$

thus, the first component of $F(x, y)$ only depends on x, i.e. we can write $F(x, y) = (F_1(x), F_2(x, y))$. Thus, the differential takes the block form

$$dF_{x,y} = \begin{pmatrix} \frac{\partial}{\partial x_i} F_1 & 0 \\ \frac{\partial}{\partial x_i} F_2 & \frac{\partial}{\partial y_j} F_2 \end{pmatrix}.$$

Since $dF_{x,y}$ is a linear isometry, $\frac{\partial}{\partial x_i} F_2 = 0$, i.e. $F_2(x, y)$ only depends on y. This implies that F_1 is an isometry of M and F_2 an isometry of $\mathbb{R}^k$.

The fundamental group $\pi_1(M)$ acts on $\tilde{M}$ by isometries and the quotient is compact. This implies that every $\pi_1(M)$-orbit intersects some compact set $W = W_1 \times W_2 \subset M \times \mathbb{R}^k$. Suppose for a contradiction that N is not compact. Then by Lemma 13.9, it contains a ray $\gamma\colon [0, +\infty) \to N$. Let t_n be a sequence such that $\gamma(t_n)$ goes to infinity (i.e. for any compact set K, $\gamma(t_n)$ is eventually outside K); for each n, we can choose an isometry in $\pi_1(M)$ mapping $(\gamma(t_n), 0)$ to an element of W, which by projection determines an isometry F_n of N that maps $\gamma(t_n)$ to an element of W_1. Up to a subsequence, we can assume that $F_n(t_n)$ converges to p and $dF_n(\gamma'(t_n))$ converges to some $(p; v) \in TN$. Thus, the rays

$$\gamma_n(t) = F_n(\gamma(t + t_n)), \quad t \geq -t_n.$$

converge to a geodesic $\gamma_\infty(t) = \exp_p tv$. Since the γ_n are rays, $d(\gamma_\infty(t, s)) = \lim d(\gamma_n(t), \gamma_n(s)) = |s - t|$, so γ_∞ is a line, which is a contradiction.

To prove (3), let G be the normal subgroup of π_1 that acts trivially on $\mathbb{R}^k$ (i.e., the kernel of the projection to $\mathrm{Isom}(\mathbb{R}^k)$). Then G acts freely with discrete orbits on $\tilde{M}$, because so does π_1, and hence on N. Since N is compact, G is finite. By construction, $H = \pi_1(M)/G$ is a group of isometries of $\mathbb{R}^k$, and it acts with discrete orbits; to see that it is a crystallographic group, all that is left to observe is that the surjective map $\tilde{M} = N \times \mathbb{R}^k \to \mathbb{R}^k$ induces a surjective map $M = \tilde{M}/\pi_1(M) \to \mathbb{R}^k/H$. □

Corollary 13.19. *If M is a compact Riemannian manifold with* Ric ≥ 0 *and the universal cover is contractible, then the universal covering is Euclidean space and M is flat.*

Proof. By the structure theorem of Cheeger and Gromoll, $\tilde{M} = N \times \mathbb{R}^k$, which can only be contractible if N is a point. □

The following can be seen as a generalization of Myers' theorem.

Corollary 13.20. *If M is compact with* Ric ≥ 0 *and* $\mathrm{Ric}_p > 0$ *at some p, then $\pi_1(M)$ is finite.*

Proof. By the structure theorem of Cheeger and Gromoll, $\tilde{M} = N \times \mathbb{R}^k$ with the product metric, so the Ricci tensor at each point is zero in the directions of $\mathbb{R}^k$. This is only possible if $k = 0$, i.e. the universal covering is compact. □

Corollary 13.21. *If M is compact with* Ric ≥ 0*, then $b_1(M) \leq \dim M = n$, with equality holding if and only if M is a flat torus.*

Proof. Consider the epimorphism

$$h\colon \pi_1(M) \to H_1(M, \mathbb{Z}) \cong \mathbb{Z}_1^{b_1} \times T,$$

where T is a finite abelian group. The finite subgroup G of the Cheeger-Gromoll theorem is mapped to T; thus, h induces an epimorphism $\pi_1(M)/G \to \mathbb{Z}^{b_1}$. We know that $\pi_1(G)$ has a finite index subgroup isomorphic to $\mathbb{Z}^k$; thus, $b_1 \leq k \leq n$.

If equality holds, then $k = n$, i.e. $\tilde{M} = \mathbb{R}^n$. So the subgroup G is trivial. Moreover the restriction of h to $\mathbb{Z}^n$ must be injective because the image has finite index. Thus the kernel of h cannot intersect the finite index subgroup $\mathbb{Z}^n$, and hence must be finite. However, $\pi_1(M)$ acts freely on $\mathbb{R}^n$, so it contains no elements of finite order (see Theorem 9.3), i.e. h is injective. Thus, $\pi_1(M) \cong \mathbb{Z}^n \times T$, and by the same token T is trivial. So $M = \mathbb{R}^n/\mathbb{Z}^n$ is a flat torus. $\square$

Recall that a Riemannian manifold is *homogeneous* if for any two points p, q there is an isometry mapping p to q.

Theorem 13.22. *Let M be a homogeneous Riemannian manifold with* $\mathrm{Ric} \geq 0$. *Then M is the Riemannian product of a compact homogeneous manifold N with Euclidean space.*

Proof. A homogeneous Riemannian manifold is complete because inj_p is the same at every point, so any geodesic can be extended indefinitely.

Write $M = N \times \mathbb{R}^k$, where N does not contain any line. Then N is homogeneous. If N is noncompact, let γ be a ray, and let F_n be isometries such that $F_n(\gamma(n)) = \gamma(0)$. Then the rays $\gamma_n \colon [-n, +\infty)$, $\gamma_n(t) = F_n(\gamma(t+n))$ are the same geodesic as γ, so γ is a line, which is absurd. $\square$

Corollary 13.23. *A homogeneous Ricci-flat Riemannian manifold is flat.*

Sketch of proof. Follows from Theorem 13.22 and the fact that a Killing field on a Ricci-flat compact manifold is parallel. $\square$

14 Maximal diameter rigidity [7]

We know from Myers' theorem that the diameter of a complete metric with $\mathrm{Ric} \geq (n-1)k$ is bounded by $\pi/\sqrt{k}$. The bound is an equality for the sphere; this actually characterizes the round metric on the sphere, by the following:

Theorem 14.1 (Cheng)**.** *Let M be a compact manifold with diameter $\pi/\sqrt{k}$ and* $\mathrm{Ric} \geq (n-1)k$. *Then M is isometric to the sphere.*

Proof. Let p, q be points with $d(p,q) = \pi/\sqrt{k}$. Define

$$r(x) = d(p,x), \quad \tilde{r}(x) = d(q,x).$$

Then $r + \tilde{r} \geq \pi/\sqrt{k}$ by the triangle inequality, and equality holds on the segment from p to q. We claim equality holds everywhere. Indeed, by Calabi's lemma we have

$$\Delta(r+\tilde{r}) \leq (n-1)\frac{S_k'(r(x))}{S_k(r(x))} + (n-1)\frac{S_k'(\tilde{r}(x))}{S_k(\tilde{r}(x))}.$$

Since S_k'/S_k is decreasing for $k > 0$, we obtain

$$\Delta(r+\tilde{r}) \leq (n-1)\frac{S_k'(r(x))}{S_k(r(x))} + (n-1)\frac{S_k'(\pi/\sqrt{k} - r(x))}{S_k(\pi/\sqrt{k} - r(x))} = 0. \tag{24}$$

The last equality follows either from the explicit expression of S_k or by the observation that the Calabi lemma is an equality on the sphere of curvature k.

Thus, $r + \tilde{r}$ is superharmonic and has a global minimum; by the minimum principle, it is constant, i.e.

$$r + \tilde{r} = \pi/\sqrt{k}.$$

It follows that for any x, any segment from p to x can be joined to a segment from x to q to obtain a segment, which must necessarily be smooth. Therefore, x is not in the cut locus of any segment from p, and r is smooth at x. This implies that $r, \tilde{r}$ are smooth outside of p, q.

Since $r + \tilde{r}$ is costant, equality holds in (24), and in particular

$$\Delta r = (n-1)\frac{S_k'(r(x))}{S_k(r(x))}.$$

Recalling that $\rho(t) = S_k'(t)/S_k(t)$ satisfies $\rho' + \rho^2 + k = 0$, and applying Proposition 11.3, we see that

$$-(n-1)k = \tfrac{\partial}{\partial r}\Delta r + \frac{(\Delta r)^2}{n-1} \leq \tfrac{\partial}{\partial r}\Delta r + |\operatorname{Hess} r|^2 = -\operatorname{ric}(\tfrac{\partial}{\partial r}, \tfrac{\partial}{\partial r}) \leq -(n-1)k.$$

In particular, the inequality $(\Delta r)^2 \leq (n-1)\,|\operatorname{Hess} r|^2$ is an equality. As in the proof of Proposition 11.3, $\operatorname{Hess} r$ is a symmetric operator on $\frac{\partial}{\partial r}^\perp$, and denoting I the identity on $\frac{\partial}{\partial r}^\perp$, Cauchy-Schwartz is an equality, so

$$\operatorname{Hess} r = \frac{1}{n-1}(\Delta r)g_r = \frac{S_k'(r(x))}{S_k(r(x))}g_r.$$

Now Proposition 11.2 gives

$$-kg_r(X,Y) = \nabla_{\frac{\partial}{\partial r}} \operatorname{Hess} r(X,Y) + \operatorname{Hess}^2 r(X,Y) = -R(X, \tfrac{\partial}{\partial r}, \tfrac{\partial}{\partial r}, Y).$$

We cannot directly invoke the classification of metrics with constant curvature, since this condition only holds in the radial direction; however, the fact that curvature determines the metric is really a consequence of Jacobi fields being determined. In this case, Jacobi fields are exactly as in a space of constant curvature. We know that along a radial geodesics $\exp_p tv$, with $|v| = 1$, Jacobi fields orthogonal to $\frac{\partial}{\partial r}$ take the form $J(t) = S_k(|v|)V$, where V is parallel, and $J'(0) = V$. If we allow $|v|$ to be arbitrary, the Jacobi field with $J(0) = 0$ and $J'(0) = V$ can be written as $J(t) = \frac{1}{|v|}S_k(t\,|v|)V$. This implies

$$\left|d\exp_p(v; V(0))\right| = \left|d\exp_p(v; J'(0))\right| = |J(1)| = \frac{1}{|v|}\,|S_k(|v|)|\,|V|\,.$$

Thus, in normal coordinates,

$$g = \frac{1}{r^2}S_k(r)^2 g^\perp + dr^2,$$

where $g^\perp$ is the Euclidean metric restricted to $\frac{\partial}{\partial r}^\perp$, or in polar normal coordinates,

$$g = S_k(r(x))^2 g_{S^{n-1}} + dr^2,$$

where $g_{S^{n-1}}$ is the metric on the sphere (see Remark 7.8). Thus, $M \setminus \{q\}$ is isometric to the complement of a point N in the sphere with curvature k. Symmetrically, $M \setminus \{p\}$ is isometric to the complement of $-N$, yielding a global isometry of M with the sphere. □

15 Rauch comparison [7]

We have already seen that in negative curvature, the Hessian of the distance function is bounded below by g (Lemma 9.5). This can be generalized to give more general bounds both above and below under the assumption that sec is bounded. The following result should also be compared with Calabi's lemma, which from a bound on Ric (the trace of R) deduces a bound on Δr (the trace of $\operatorname{Hess} r$).

Theorem 15.1 (Rauch comparison)**.** *Suppose $k \leq \sec \leq K$, and let $g = dr^2 + g_r$ in polar coordinates inside a geodesic ball. Then*

$$\frac{S_K'(r)}{S_K(r)} g_r \leq \operatorname{Hess} r \leq \frac{S_k'(r)}{S_k(r)} g_r.$$

Proof. We follow different techniques for the upper and lower bound. In both cases, let $\gamma \colon [0, b]$ be a unit speed radial geodesic inside the geodesic ball, $\gamma(0) = p$.

Assuming $\sec \leq K$, let

$$\rho(t) = \frac{g(J', J)}{|J|^2} = \operatorname{Hess} r(\frac{J}{|J|}, \frac{J}{|J|}),$$

for J a Jacobi field along γ with $J(0) = 0$ and $J'(0)$ orthogonal to $\gamma'(0)$. Then

$$\begin{aligned}\rho' = \frac{(g(J'', J) + |J'|^2)\,|J|^2 - 2g(J', J)^2}{|J|^4} \\ = -\sec(J, \gamma') + \frac{|J'|^2\,|J|^2 - 2g(J', J)^2}{|J|^2} \geq -K - \rho^2.\end{aligned}$$

In addition $\rho(t) \sim \frac{1}{t}$, since $g(J', J) \sim t\,|J'(0)|^2$. Then apply Lemma 11.4 to conclude that $\frac{S_K'(t)}{S_K(t)} \leq \rho(t)$.

Now assume $\sec \geq k$, and consider a radial geodesic γ, let V be a unit parallel vector field along γ orthogonal to γ', and define

$$\rho = \operatorname{Hess} r(V, V);$$

then, since V is parallel along γ, we have

$$\rho' = (\nabla_{\frac{\partial}{\partial r}} \operatorname{Hess} r)(V, V).$$

By Proposition 11.2, we have

$$\rho' + \operatorname{Hess}^2 r(V, V) = -\operatorname{Rm}(V, \tfrac{\partial}{\partial r}, \tfrac{\partial}{\partial r}, V).$$

Applying Cauchy-Schwartz in the form

$$\operatorname{Hess} r(V, V)^2 = g(\nabla_V \tfrac{\partial}{\partial r}, V)^2 \leq \left|\nabla_V \tfrac{\partial}{\partial r}\right|^2 |V|^2 = \left|\nabla_V \tfrac{\partial}{\partial r}\right|^2 = \operatorname{Hess}^2 r(V, V),$$

we obtain

$$\rho' = -\operatorname{Hess}^2 r(V, V) - \sec(V, V) \leq -\rho^2 - k.$$

and again we conclude by Lemma 11.4. $\square$

Remark 15.2. Along a unit speed radial geodesic, with $J \in \mathcal{J}^\perp$, $J(0) = 0$, the formula says

$$\frac{S_K'(r)}{S_K(r)} |J|^2 \leq \operatorname{Hess} r(J,J) = g(J,J') \leq \frac{S_k'(r)}{S_k(r)} |J|^2 .$$

If the geodesic is not unit speed, say $\gamma(v) = \exp_p tv$, then $r(t) = t\,|v|$, so $\frac{\partial}{\partial t} S_k(r(t)) = S_k'(r(t))\,|v|$, and similarly $\frac{\partial}{\partial t} J$, so again

$$\frac{\frac{\partial}{\partial t} S_K(r(t))}{S_K(r(t))} |J|^2 \leq g(J,J') \leq \frac{\frac{\partial}{\partial t} S_k(r(t))}{S_k(r)} |J|^2 .$$

As a consequence we also obtain another theorem (also known as Rauch's theorem, see [4]), which compares the length of curves to related curves in a space form. The idea is that curves become longer as the curvature decreases (consider the case of a sphere).

Corollary 15.3 (Rauch). *Let M, $\tilde{M}$ be Riemannian manifolds; fix a point $p \in M$ and a point $\tilde{p} \in \tilde{M}$, and let $I \colon T_pM \to T_{\tilde{p}}\tilde{M}$ be an isometry. Suppose that $\exp_p$ has no critical points in $B_\rho(0) \subset T_pM$ and that $\exp_{\tilde{p}}$ has no critical points in $B_\rho(0) \subset T_{\tilde{p}}\tilde{M}$. For any curve c in $B_\rho(0) \subset T_pM$, consider the curves*

$$\gamma = \exp_p \circ c, \quad \tilde{\gamma} = \exp_{\tilde{p}} \circ I \circ c.$$

Then:

- *if M has $\sec \leq K$ and $\tilde{M}$ has constant sectional curvature K, then $l(\gamma) \geq l(\tilde{\gamma})$;*
- *if M has $k \leq \sec$ and $\tilde{M}$ has constant sectional curvature k, then $l(\gamma) \leq l(\tilde{\gamma})$.*

Proof. Suppse that $\sec \leq K$. Let $\overline{\gamma}$ be a geodesic variation obtained by $\overline{\gamma}(s,t) = \exp_p(tc(s))$. Then each γ_s is a geodesic endowed with a Jacobi field

$$J_s(t) = \frac{\partial \overline{\gamma}}{\partial s} = t d\exp_p(tc(s); c'(s)).$$

In particular, we have $J_s(1) = \gamma'(s)$, $J_s'(0) = c'(s)$.

The Jacobi field $J_s = \frac{\partial \overline{\gamma}}{\partial s}$ vanishes at 0, so it takes the form $ht\gamma_s' + J_s^\perp$, where the constant h can be determined by

$$g(c'(s), \gamma_s'(0)) = g(J_s'(0), \gamma_s'(0)) = hg\,|\gamma_s'(0)|^2 .$$

By Rauch comparison applied to $J^\perp$, we have

$$\frac{\partial}{\partial t} \log S_K(r(\gamma_s(t))) \leq \frac{\partial}{\partial t} \log \left| J_s^\perp \right| . \tag{25}$$

Now consider similarly the variation

$$\tilde{\gamma}_s = \exp_{\tilde{p}}(tI(c(s))).$$

Its Jacobi field is $\tilde{J}_s$,

$$\tilde{J}_s(1) = \tilde{\gamma}'(s), \quad \tilde{J}_s = ht\tilde{\gamma}_s' + \tilde{J}_s^\perp,$$

where the constant h is the same as before because I is an isometry, and for the same reason $(\tilde{J}_s^{\perp})'(0)$ has the same norm as $(J_s^{\perp})'(0)$.

Since $\tilde{M}_K$ has constant curvature K, we know that $\left|J_s^{\perp}\right|$ is a constant multiple of $S_k(r(\tilde{\gamma}_s(t)))$, so combining with (25) we find

$$\tfrac{\partial}{\partial t} \log \left|\tilde{J}_s^{\perp}\right| \leq \tfrac{\partial}{\partial t} \log \left|J_s^{\perp}\right|.$$

Thus $t \mapsto \left|\tilde{J}_s^{\perp}\right|^2 / \left|J_s^{\perp}\right|^2$ is a smooth nonincreasing function which takes the value 1 at $t = 0$, and therefore $\left|\tilde{J}_s\right| \leq |J_s|$. At $t = 1$, this gives $|\tilde{\gamma}'(s)| \leq |\gamma'(s)|$. This implies that $\tilde{\gamma}$ is shorter than γ.

The case $k \leq \sec$ is similar. □

The methods of the Rauch comparison theorems can also be used to obtain a partial generalization of the Cartan-Hadamard theorem.

Theorem 15.4. *If M is complete and $\sec \leq K$, with $K > 0$, then for any p, $\exp_p\colon B(0, \pi/\sqrt{K}) \to M$ has no critical points.*

Proof. We must show that geodesics shorter than $\pi/\sqrt{K}$ have no conjugate points.

Let $\gamma\colon [0, b)$ be a unit speed geodesic, and let J be a nonzero Jacobi field along γ, with $\gamma(0) = 0$, $\gamma'(0)$ orthogonal to $\gamma'(0)$. As we computed in the proof of Theorem 15.1, $\rho(t) = \frac{g(J',J)}{|J|^2}$ satisfies $\rho' + \rho^2 + K \geq 0$, so by Lemma 11.4 we have that $\rho \geq \frac{S_k'(t)}{S_k(t)}$, $t < \min\{b, \pi/\sqrt{K}\}$. We can rewrite this as

$$\frac{d}{dt} \log |J|^2 \geq \frac{d}{dt} \log S_K^2,$$

which shows that $|J|^2 / S_k^2$ is nondecreasing. We have $S_K(t) \sim t$ as $t \to 0$, and $|J(t)|^2 = t^2(J'(0), J'(0)) + o(t^2)$, so

$$\lim_{t \to 0^+} \frac{|J(t)|^2}{S_K(t)^2} = \lim_{t \to 0^+} \frac{t^2(J'(0), J'(0)) + o(t^2)}{t^2} = |J'(0)|^2.$$

Therefore $|J(t)|^2 \geq |J'(0)|^2 S_K^2(t)$, which does not vanish for $0 < t < \pi/\sqrt{K}$. □

16 Positive sectional curvature

Recall from Section 9 that an axis for an isometry F is a geodesic γ such that $F \circ \gamma$ is a reparametrization of γ. On a complete manifold, if the displacement function δ_F has a positive minimum, then there is an axis. Additionally, recall from Lemma 9.13 that any nontrivial deck transformation of the universal cover of a compact manifold has an axis that maps to a geodesic loop in M which has minimal length in its free homotopy class.

Theorem 16.1 (Synge)**.** *Let M be a compact manifold with $\sec > 0$.*

- *If M is even dimensional and orientable, then M is simply connected.*
- *If M is odd dimensional, then M is orientable.*

Proof. Let $\pi\colon \tilde{M} \to M$ be the universal cover. Observe that $\tilde{M}$ is orientable, and M is orientable if and only if all deck transformations are orientation-preserving.

We will prove the two assertions simultaneously, assuming for a contradiction that π has a nontrivial deck transformation F which preserves or reverses orientation according to whether the dimension is even or odd.

By Lemma 9.13, F has an axis $\tilde{\gamma}$; this yields a geodesic loop $\gamma\colon [0,b] \to M$ which has minimal length in its free homotopy class. Assume γ has unit speed.

Now consider parallel translation along $\gamma, \tilde{\gamma}$; this gives isometries $\tilde{L}$ and L that make the following diagram commute:

$$\begin{tikzcd}
T_{\tilde{\gamma}(0)}\tilde{M} \arrow[r, "\tilde{L}"] \arrow[d] & T_{\tilde{\gamma}(b)}\tilde{M} \arrow[r, "dF^{-1}"] \arrow[d] & T_{\tilde{\gamma}(0)}\tilde{M} \arrow[dl] \\
T_{\gamma(0)}M \arrow[r, "L"] & T_{\gamma(0)}M &
\end{tikzcd}$$

Since $\tilde{L}$ preserves the orientation, $\tilde{L} \circ dF^{-1}$ preserves or reverses orientation according to whether F does. Thus, $\det L = (-1)^n$.

We claim that L has 1 as an eigenvalue of multiplicity at least two. Indeed, the real eigenvalues can only be ± 1, so $(-1)^n = \det L = (-1)^k$, where n is the dimension of M and k is the multiplicity of -1. So the sum of the eigenspaces with eigenvalue $\neq -1$ has even dimension, and this implies that the multiplicity of 1 is even. Notice that the multiplicity cannot be zero, as $L(\gamma'(0)) = \gamma'(0)$.

Let v be a unit eigenvector, $L(v) = v$, orthogonal to $\gamma'(0)$. Then parallel translation defines a closed vector field V along γ. Consider the geodesic variation

$$\gamma_s = \exp_{\gamma(t)} sV(t),$$

and write Synge's formula,

$$\begin{aligned}
\frac{\partial^2}{\partial s^2}E(\gamma_s)|_{s=0} = \int_0^b \frac{\partial^2 \gamma_s}{\partial t \partial s} tdt - \int_0^b g(R(V,\gamma')\gamma', V)dt + g(\frac{\partial^2 \gamma_s}{\partial s^2}, \gamma')(0,b) - g(\frac{\partial^2 \gamma_s}{\partial s^2}, \gamma')(0,0) \\
= -\int_0^b \sec(V,\gamma')dt < 0.
\end{aligned}$$

This is a contradiction because the curves γ_s are loops in the same free homotopy class as γ, which minimizes energy by the argument of Proposition 6.3 (though endpoints are not fixed here). $\square$

Remark 16.2. Any nonorientable manifold has a 2 : 1 orientable cover, so Synge's theorem implies that the fundamental group of a manifold of even dimension with positive sectional curvature has at most order two. For instance, this shows that $\mathbb{RP}^2 \times \mathbb{RP}^2$ does not carry such a metric.

Theorem 15.4 can be turned into a result on the injectivity radius by using Klingenberg's lemma (Lemma 10.9).

Recall that a curve γ is a *geodesic loop* if it is a geodesic which is not injective, i.e. meets itself. A *periodic geodesic* is a geodesic such that $\gamma(t+k) = \gamma(t)$ for some k. Periodic geodesics are geodesics loops, but the converse does not hold.

Proposition 16.3 (Klingenberg)**.** *If M is compact with* $\sec \leq K$*,* $K > 0$*, then*

$$\operatorname{inj}_p \geq \min\left\{\pi/\sqrt{K}, \frac{1}{2}\textit{length of shortest geodesic loop based at } p\right\}.$$

In addition, either:

- $\operatorname{inj} \geq \pi/\sqrt{K}$*; or*
- inj *is one half the length of the shortest geodesic loop, which is a periodic geodesic.*

Proof. By Klingenberg's lemma, $\operatorname{inj} p$ is the length of a segment $[0,1] \ni t \mapsto \exp_p tv$. If $|v| \geq \pi/\sqrt{K}$, there is nothing to prove. Otherwise, we know from Theorem 15.4 that there are no conjugate points along the segment, and $\exp_p$ is nonsingular at v. By the lemma again, we see that $\exp_p v = \exp_p w$; the two segments $t \mapsto \exp tv$ and $t \mapsto \exp tw$ can be joined to form a loop, which is smooth at $\exp_p v$. Since both are segments, they have the same length, so the length of the loop is $2\,|v|$.

For the second part, observe that if for all points $\operatorname{inj}_p \geq \pi/\sqrt{K}$, then the claim holds trivially. Otherwise, use compactness and choose p such that $\operatorname{inj} = \operatorname{inj}_p < \pi/\sqrt{K}$. Then there is a geodesic loop based at p of length $2\operatorname{inj}_p$, say $\gamma\colon [0, 2\operatorname{inj} p] \to M$. If $\gamma(\operatorname{inj} p) = q$, and γ is obtained by concatenating the only two segments from p to q. We have that $\operatorname{inj}_q = \operatorname{inj}_p < \pi/\sqrt{K}$, so symmetrically there are exactly two segments from q to p which form a geodesic loop, which are necessarily the segments forming γ. This shows that γ is periodic. □

Theorem 16.4 (Klingenberg)**.** *If M is compact of even dimension with* $0 < \sec \leq 1$*, then* $\operatorname{inj} \geq \pi$ *if M is orientable, and* $\operatorname{inj} \geq \pi/2$ *if M is not orientable.*

Proof. Assume M orientable, and suppose for a contradiction that $\operatorname{inj} M < \pi$. Then by Proposition 16.3 there is a periodic geodesic of length $2\operatorname{inj} M$. So let $\gamma\colon [0, 2\operatorname{inj}] \to M$ be a unit speed geodesic loop. Since M is oriented, parallel transport $L\colon T_{\gamma(0)}M \to T_{\gamma(2\operatorname{inj})}M$ preserves the orientation. Arguing as in Theorem 16.1, we see that L must fix an orthogonal direction, giving a closed parallel unit vector field $V(t)$ along γ. The associated variation $\gamma_s = \exp_{\gamma(t)} sV(t)$ is such that each γ_s is a loop and $s \mapsto l(\gamma_s)$ has a local maximum at $s = 0$, i.e. each γ_s has length $< 2\operatorname{inj} M$.

For any $s \neq 0$, since γ_s is a loop, it is contained in the geodesic ball $B_{\operatorname{inj} M}\gamma_s(0)$. Choose t_s maximizing distance between $\gamma_s(0)$ and $\gamma_s(t_s)$; then

$$\gamma_s(t_s) = \exp_{\gamma_s(0)}(v_s), \quad v_s \in \operatorname{seg}(\gamma_s(0)).$$

For $s \to 0$, the length of γ_s converges to $2\operatorname{inj} M$, so $|v_s|$ converges to $\operatorname{inj} M$. By compactness, for some sequence $s_n \to 0$, we get $|v_{s_n}| \to v_0$, $|v_0| = \operatorname{inj} M$. Since $\gamma_s(t_s)$ is the farthest point from $\gamma_s(0)$ in the curve γ_s, it follows that the segment $t \mapsto \exp_{\gamma_s(0)} tv_s$ is orthogonal to $\gamma_s'(t_s)$. Thus, $\exp_{\gamma(0)} tv_0$ meets $\gamma(t)$ orthogonally at $t = \operatorname{inj} M$. This implies that $\gamma(\operatorname{inj} M)$ is joined by three distinct segments to $\gamma(0)$; by Lemma 10.9 this is only possible if $\gamma(\operatorname{inj} M)$ is a critical value for $\exp_{\gamma(0)}$; however, we are assuming $\operatorname{inj} M < \pi$, so this is absurd by Theorem 15.4.

If M is not orientable, then it has a $2:1$ orientable cover with injectivity radius $\geq \pi$. By the triangular inequality, geodesic balls of radius $\pi/2$ map diffeomorphically to M, so they are also geodesic balls. □

17 The sphere theorem

For this section we mainly follow [3] (but see also [4]).

We aim at proving the following, in the case where M is even dimensional:

Theorem 17.1 (Berger)**.** *Let M be a compact, simply connected manifold with $\frac{1}{4} < \sec \leq 1$. Then M is homeomorphic to a sphere.*

The "pinching" constant $\frac{1}{4}$ cannot be improved, and indeed strict inequality is necessary. On the other hand, the statement can be improved by replacing "homemomorphic" with "diffeomorphic"; this is a much more difficult and recent result due to Brendle and Schoen (see [2]).

Remark 17.2. Since M is assumed to be compact, the pinching condition on the curvature can also be written as

$$\frac{1}{4} < \delta \leq \sec \leq 1.$$

We will write the pinched curvature condition in this form in the following lemmas.

The proof of the sphere theorem relies on several results. One of these is an estimate on the injectivity radius. We have seen that in even dimensions, $\operatorname{inj} M \geq \pi$ (orientability is automatic since we assume simply connected). In odd dimensions, the following holds:

Theorem 17.3 (Klingerberg-Sakai)**.** *If M is compact, simply connected and $\frac{1}{4} \leq \sec \leq 1$, then $\operatorname{inj} M \geq \pi$.*

We will not prove this here, so our proof of Berger's sphere theorem is only complete for even dimensions; see [5, 2.6.A.1] for a proof. Notice that Berger's sphere theorem only requires a weaker version of the theorem with the assumption $\frac{1}{4} < \sec \leq 1$; this version of Theorem 17.3 was proved long before by Klingerberg.

Remark 17.4. The fact that simpy connectedness is necessary can be seen by considering the quotients $\mathrm{SU}(2)/\mathbb{Z}_k$, where $\mathbb{Z}_k$ is a cyclic group in $U(1)$. The subgroup $U(1) \subset \mathrm{SU}(2)$ is a periodic geodesic of length 2π, so the quotient has a periodic geodesic of length $2\pi/k$ and its injectivity radius can be made arbitrarily small.

Remark 17.5. Consider the Berger sphere, where an orthonormal basis $e_1 = \frac{1}{a}\underline{i}$, $e_2 = \underline{j}$, $e_3 = \underline{k}$. We proved in Section 3 that $\nabla_{e_1} e_1 = 0$, so there is a periodic geodesic of length $2\pi a$, i.e. $\operatorname{inj} = \pi a$. On the other hand,

$$a^2 \leq \sec \leq 4 - 3a^2,$$

so for the rescaled metric $(4 - 3a^2)g$ we get

$$\operatorname{inj} = \pi a \sqrt{4 - 3a^2}, \quad \frac{a^2}{4 - 3a^2} \leq \sec \leq 1.$$

For $a < \frac{1}{\sqrt{3}}$, $\operatorname{inj} < \pi$, and $\frac{a^2}{4-3a^2}$ can be made arbitrarily close to $\frac{1}{9}$. This shows that it it not possible to replace $[\frac{1}{4}, 1]$ with $[\frac{1}{9} - \epsilon, 1]$ in Theorem 17.3.

The second result is the following:

Lemma 17.6 (Berger)**.** *Let M be compact, fix q and choose p in M such that $d(p,q)$ is maximal. Then for all $w \in T_pM$ there exists a segment γ from $p = \gamma(0)$ to q with $g(\gamma'(0), w) \geq 0$.*

Proof. The statement holds trivially if $w = 0$. Otherwise, let $\sigma(t) = \exp_p tw$. Pick a sequence $p_n \to p$ of points on $\sigma(t)$ and choose $v_n \in \operatorname{seg}(p_n)$ such that $\exp_{p_n}(v_n) = q$, i.e. $[0,1] \to M$, $t \mapsto \exp_{p_n} tv_n$ is a segment. Then $|v_n|$ converges to $d(p,q)$, so up to a subsequence we may assume that v_n converges to some $v \in T_pM$, $\exp_p v = q$. We claim that we can choose the sequences $\{p_n\}, \{v_n\}$ so that $g(v_n, \sigma') \geq 0$; in the limit, we obtain $g(v, \sigma'(0)) = g(v,w) \geq 0$; thus, this gives the geodesic $\gamma(t) = \exp_p tv$ as in the statement.

In order to prove the claim, suppose for a contradiction that there exists a $t_0 > 0$ such that for every $0 \leq \tau < t_0$ and every segment γ_τ from $\sigma(\tau)$ to q we have $g(\gamma_\tau'(0), \sigma'(\tau)) < 0$. For fixed τ, choose q_0 on γ_τ such that the subsegment of γ_τ connecting $\sigma(\tau)$ to q_0 is inside a geodesic ball centered at q_0. Then by Gauss' lemma the points of $\sigma(t)$ become closer to q_0 as $t < \tau$ (at least for t sufficiently close to τ). So by the triangular inequality

$$d(\sigma(t), q) \leq d(\sigma(t), q_0) + d(q_0, q) < d(\sigma(\tau), q_0) + d(q_0, q) = d(\sigma(\tau), q).$$

Since τ is arbitrary, this shows that the distance from q is strictly increasing along $\sigma(t)$ in the interval $(0, t_0)$, which is a contradiction because $d(p,q)$ is maximal. $\square$

Remark 17.7. In Berger's lemma, compactness is only used to have a point p with maximal distance from q. However, notice that a complete manifold such that a point q has distance $\leq \rho$ from every other point is covered by $\exp_q(\overline{B_\rho(0)})$, and therefore is compact.

The next lemma can be proved in more than one way; we will follow Tsukamoto's proof using Rauch comparison.

Lemma 17.8 (Berger, Tsukamoto)**.** *Let M be compact, simply connected, with sectional curvature satisfying*

$$\frac{1}{4} < \delta \leq \sec \leq 1;$$

let p, q be points such that $d(p,q) = \operatorname{diam} M$ and choose ρ such that

$$\frac{\pi}{2\sqrt{\delta}} < \rho < \pi.$$

Then M is the union of the geodesic balls $B_\rho(p)$, $B_\rho(q)$.

Proof. We use Klingenberg's theorem (which also holds in odd dimensions in these hypotheses), which implies that $\operatorname{inj} M \geq \pi$; this shows that $B_\rho(p)$, $B_\rho(q)$ are indeed geodesic balls.

Suppose for a contradiction that there exists $r \in M$ with

$$d(p,r) \geq d(q,r) \geq \rho.$$

We claim that $\partial B_\rho(p)$ and $\partial B_\rho(q)$ have nonempty intersection. Indeed, a segment from q to r intersects $\partial B_\rho(q)$ in q'. This point does not belong to $B_\rho(p)$, for otherwise we would have

$$d(r,q') \geq d(r,p) - d(q',p) > d(r,p) - \rho \geq d(r,q) - \rho = d(r,q').$$

This implies that $\partial B_\rho(q)$ is not contained in $B_\rho(p)$. On the other hand, by (5), $\mathrm{ric} \geq (n-1)\delta$, so the Bonnet-Myers theorem (Corollary 6.8) gives $\mathrm{diam} \leq \pi/\sqrt{\delta} < 2\rho$; together with the observation that M cannot be contained in a single geodesic ball, we find

$$\rho < d(p,q) < 2\rho.$$

Thus $\partial B_\rho(q)$ meets $B_\rho(p)$ and therefore its boundary, so $\partial B_\rho(p)$ and $\partial B_\rho(q)$ intersect at some r_0, i.e.

$$d(r_0,p) = d(r_0,q) = \rho.$$

Let λ be a segment from p to r_0. By Berger's lemma, there is a segment γ from p to q with $g(\lambda'(0),\gamma'(0)) \geq 0$. Then γ intersects $\partial B_\rho(p)$ in some point s.

We can apply Rauch's theorem and compare the triangle r_0ps with a triangle in the sphere with curvature δ, i.e. with radius $\frac{1}{\sqrt{\delta}}$; notice that $\frac{1}{\sqrt{\delta}} \geq 1$, so $B_\rho(\hat{p})$ is a geodesic ball in that sphere. The corresponding isosceles triangle $\hat{r}_0\hat{p}\hat{s}$ in the sphere has angle $\leq \pi/2$ at $\hat{p}$. This means that, in appropriate coordinates, $\hat{r}_0$ and $\hat{s}$ are in the same orthant; therefore, the geodesic arc that connects them is also contained in the orthant and has length bounded by $\frac{\pi}{2\sqrt{\delta}}$, i.e. $d(\hat{s},\hat{r}_0) \leq \frac{\pi}{2\sqrt{\delta}}$. By Rauch comparison, the curve in M corresponding to the geodesic realizing the distance has length $\leq \frac{\pi}{2\sqrt{\delta}}$, and therefore

$$d(s,r_0) \leq \frac{\pi}{2\sqrt{\delta}} < \rho.$$

Now let s_0 be the point of γ closest to r_0; s_0 must belong to the interior of the segment, since $d(p,r_0) = \rho = d(q,r_0)$. Therefore, the segment from r_0 to s_0 is orthogonal to γ, and

$$d(s_0,r_0) \leq \frac{\pi}{2\sqrt{\delta}} < \rho.$$

Since $d(p,q) \leq \pi/\sqrt{\delta}$, we have $d(p,s_0) \leq \frac{\pi}{2\sqrt{\delta}}$ or $d(q,s_0) \leq \frac{\pi}{2\sqrt{\delta}}$; suppose the first holds. As above, we can consider the triangle ps_0r_0, which has angle $\pi/2$ at s_0 and two sides of length $\leq \frac{\pi}{2\sqrt{\delta}}$; so, the corresponding triangle in the sphere is contained in a single orthant in suitable coordinates. Therefore, by Rauch comparison $d(p,r_0) \leq \frac{\pi}{2\sqrt{\delta}} < \rho$, which is absurd. The case $d(q,s_0) \leq \frac{\pi}{2\sqrt{\delta}}$ is similar. $\square$

Lemma 17.9. *Let M be compact, simply connected, with sectional curvature satisfying*

$$\frac{1}{4} < \delta \leq \sec \leq 1;$$

let p,q be points such that $d(p,q) = \operatorname{diam} M$ and choose ρ such that

$$\frac{\pi}{2\sqrt{\delta}} < \rho < \pi.$$

On every geodesic γ of length ρ, $\gamma(0) = p$, there is exactly one point m equidistant from p and q.

Proof. By the injectivity radius estimate $\operatorname{inj} \geq \pi$, we have that γ is a segment and its terminal point lies on the boundary of $B_\rho(p)$; thus, by Lemma 17.8, $\gamma(\rho)$ belongs to $B_\rho(q)$. Thus the continuous function $t \mapsto d(p, \gamma(t)) - d(q, \gamma(t))$ has a zero, which gives the equidistant point $\gamma(t) = m$.

Suppose for a contradiction that m' is another equidistant point. Without loss of generality, assume m' is between p and m. Then

$$d(q, m) = d(p, m) = d(p, m') + d(m', m) = d(q, m') + d(m', m).$$

Let σ be the segment from q to m. Then m' lies on σ, in between q and m. Since σ and γ contain the unique segment from m' to m, they are the same geodesic, so $p = q$, which is absurd. $\square$

Theorem 17.10 (Berger)**.** *Let M be a compact, simply connected manifold with $\frac{1}{4} < \sec \leq 1$. Then M is homeomorphic to a sphere.*

Proof. Let p, q be points of M such that $\operatorname{diam} M = d(p, q)$.

We know from Lemma 17.9 that for any unit v in T_pM, there is a unique point m on $[0, \rho] \to \exp_t v$ which has equal distance $\tau(v)$ from p and q. The function τ is continuous, because its graph is the compact set

$$\{(t, v) \mid d(p, \exp_p(tv)) = d(q, \exp_p(tv))\},$$

which projects homeomorphically on the unit sphere in T_pM. We can extend τ to $T_pM \setminus \{0\}$ by $\tau(tv) = \tau(v)$, $t > 0$.

This gives a map

$$T_pM \supset \overline{B_1(0)} \to M, \quad v \mapsto \begin{cases} \exp_p \tau(v)v & v \neq 0 \\ p & v = 0 \end{cases},$$

which is a homeomorphism onto the image, call it D_1, because it is continuous and injective. By construction, any point r of D_1 satisfies $d(r, p) \leq d(r, q)$, and equality holds if and only if r is on the image of the boundary.

We can construct in a similar way a map

$$T_qM \supset \overline{B_1(0)} \to M,$$

which maps homeomorphically onto a disc D_2. We then have that M is the union of two discs glued along the boundaries, so it is homeomorphic to a sphere (notice that there is only one orientation-reversing homeomorphism $S^n \to S^n$ up to homotopy). $\square$

References

[1] M. Berger, A panoramic view of Riemannian geometry. *Springer, 2003.*

[2] S. Brendle, R. Schoen. Manifolds with 1/4-pinched curvature are space forms. J. Amer. Math. Soc.22(2009), no.1, 287–307.

[3] Do Carmo, Riemannian Geometry. *Birkhauser, 1992.*

[4] Cheeger and Ebin. Comparison theorems in Riemannian geometry.

[5] Klingerberg. Riemannian Geometry. De Gruyter 1995.

[6] Meyer. Toponogov's Theorem and Applications.

[7] P. Petersen, Riemannian Geometry. Third edition. *Springer, 2006.*

[8] J.A. Wolf. Spaces of constant curvature. McGraw-Hill.