NOTES ON RIEMANNIAN GEOMETRY

DANNY CALEGARI

Abstract. These are notes for an introductory course on Riemannian Geometry. These notes follow a course given at the University of Chicago in Spring 2013.

Contents

1. Smooth manifolds

1.1. Linear algebra.

Definition 1.1 (Tensor product)**.** Let V and W be real vector spaces. The *tensor product* of V and W, denoted $V \otimes W$, is the real vector space *spanned by* elements of the form $v \otimes w$ with $v \in V$ and $w \in W$, and subject to the relations

(1) distributivity: $(v_1+v_2)\otimes w = (v_1\otimes w)+(v_2\otimes w)$ and $v\otimes(w_1+w_2) = (v\otimes w_1)+(v\otimes w_2)$
(2) linearity: if $\lambda \in \mathbb{R}$ then $\lambda(v \otimes w) = (\lambda v) \otimes w = v \otimes (\lambda w)$

Warning 1.2. It is *not* true in general that every element of $V \otimes W$ has an expression of the form $v \otimes w$. Rather, every element of $V \otimes W$ can be written (not uniquely) as a *finite sum* of the form $\sum_i v_i \otimes w_i$.

Example 1.3. If $\{v_i\}$ is a basis for V and $\{w_j\}$ is a basis for W then $\{v_i \otimes w_j\}$ is a basis for $V \otimes W$. Hence we have

$$\dim(V \otimes W) = \dim(V) \times \dim(W)$$

Example 1.4. There are isomorphisms $U \otimes (V \otimes W) \cong (U \otimes V) \otimes W$ given by $u \otimes (v \otimes w) \to (u \otimes v) \otimes w$ and $V \otimes W \cong W \otimes V$ given by $v \otimes w \to w \otimes v$.

Date: April 1, 2015.

Example 1.5. If V is a real vector space, the tensor product of V with itself n times is denoted $V^{\otimes n}$.

The disjoint union of $V^{\otimes n}$ over all n is an associative algebra, with $\otimes$ as the operation. This algebra is called the *tensor algebra* TV.

Definition 1.6 (Symmetrization and antisymmetrization)**.** Let I be the (two-sided) ideal in TV generated by all elements of the form $v \otimes v$ for $v \in V$. The tensor product operation descends to the quotient $\Lambda V := TV/I$ and makes it into an algebra, the *exterior algebra* of V.

In a similar way, we can define I' to be the (two-sided) ideal generated by all elements of the form $v \otimes w - w \otimes v$ and define $SV := TV/I'$ to be the *symmetric algebra* of V.

The grading on TV descends to ΛV and SV and makes them into graded algebras. The pieces of rank n are denoted $\Lambda^n V$ and $S^n V$ respectively.

1.2. **Duals and wedge product.** We denote the dual $\mathrm{Hom}(V, \mathbb{R})$ of V by V^* (if V is a complex vector space, we write V^* for $\mathrm{Hom}(V, \mathbb{C})$ by abuse of notation). Then $(V^{\otimes n})^* = (V^*)^{\otimes n}$ for each n, with pairing defined by

$$(\xi_1 \otimes \cdots \otimes \xi_n)(v_1 \otimes \cdots \otimes v_n) = \xi_1(v_1)\xi_2(v_2) \cdots \xi_n(v_n)$$

We would like the relation $\Lambda^n V^* = (\Lambda^n V)^*$ to hold; but we have defined $\Lambda^n V$ as a *quotient* of $V^{\otimes n}$ so it would be more natural (although an abuse of notation), and much more convenient for the purposes of computation, to define $\Lambda^n V^*$ as a *subalgebra* of $(V^*)^{\otimes n}$.

We introduce the notation of wedge product, defined by the formula

$$\xi_1 \wedge \cdots \wedge \xi_n := \sum_\sigma (-1)^{\mathrm{sign}(\sigma)} \xi_{\sigma(1)} \otimes \cdots \otimes \xi_{\sigma(n)}$$

where the sum is taken over all permutations σ of the indices, and the sign of the permutation is the parity of the number of elementary transpositions needed to express it. With this definition, the pairing of $\Lambda^n V^*$ with $V^{\otimes n}$ takes the form

$$(\xi_1 \wedge \cdots \wedge \xi_n)(v_1 \otimes \cdots \otimes v_n) = \det(\xi_i(v_j))$$

Observe that this formula vanishes on the ideal I, so it is well defined on the quotient $\Lambda^n V$. If we define

$$\mathrm{Alt}_n(\xi_1 \otimes \cdots \otimes \xi_n) := \frac{1}{n!} \xi_1 \wedge \cdots \wedge \xi_n$$

then it is easy to observe that Alt_n is a *projection*; i.e. $\mathrm{Alt}_n \circ \mathrm{Alt}_n = \mathrm{Alt}_n$.

It is sometimes also convenient to think of $\Lambda^n V^*$ as the space of antisymmetric multilinear operator on V^n, in which case we write

$$(\xi_1 \wedge \cdots \wedge \xi_n)(v_1, v_2, \cdots, v_n) = \det(\xi_i(v_j))$$

The main example we have in mind is that V should be $T_p M$, the tangent space to a smooth manifold M at a point p, and V^* is the cotangent space $T_p^* M$. It is common to consider sections of the bundle $\Lambda^n T^* M$ for various n (i.e. differential forms), but much more unusual to consider sections of $\Lambda^n TM$ for $n > 1$.

In order for wedge product (as defined above) to be associative, we must introduce some fudge factors. Let $\alpha \in (V^*)^{\otimes p}$ and β in $(V^*)^{\otimes q}$. Then

$$(p+q)!\,\mathrm{Alt}_{p+q}(\alpha \otimes \beta) = \mathrm{Alt}_{p+q}(p!\,\mathrm{Alt}_p(\alpha)) \otimes (q!\,\mathrm{Alt}_q(\beta))$$

Hence for $\alpha \in \Lambda^p V^*$ and $\beta \in \Lambda^q V^*$ we define

$$\alpha \wedge \beta := \frac{p!\,q!}{(p+q)!}\mathrm{Alt}_{p+q}(\alpha \otimes \beta)$$

and observe that the resulting product makes $\Lambda^* V^*$ into an associative algebra, extending the definition implicit in the notation above.

1.3. Inner products.

Definition 1.7 (Inner product)**.** If V is a real vector space, an *inner product* on V is an element $q \in V^* \otimes V^*$. It is *symmetric* if $q(u,v) = q(v,u)$ and *antisymmetric* if $q(u,v) = -q(v,u)$ for all $u,v \in V$. Symmetric and antisymmetric inner products can be thought of as elements of S^2V^* and $\Lambda^2 V^*$ respectively.

An inner product is *nondegenerate* if for all v there is u so that $q(v,u) \neq 0$.

If we choose a basis e_i for V and write $v \in V$ as a (column) vector in this basis, then every inner product corresponds to a matrix Q, and $q(u,v) = u^TQv$. A change of basis replaces Q by a matrix of the form S^TQS for some invertible matrix S. An inner product is (anti)symmetric if and only if Q is (anti)symmetric in the usual sense.

Example 1.8. If q is antisymmetric and nondegenerate, then V has even dimension $2n$, and q is conjugate to a symplectic form. That is, q is represented (in some basis) by the $2n \times 2n$ matrix

$$J := \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

where I is the $n \times n$ identity matrix.

Example 1.9. If q is symmetric and nondegenerate, then q is represented by a diagonal matrix Q with p diagonal entries equal to 1 and $n-p$ diagonal entries equal to -1. The *signature* of q is the tuple $(p, n-p)$ (sometimes if the dimension n is understood the signature is defined to be the difference $p-(n-p) = 2p-n$).

1.4. Lie algebras.

Definition 1.10 (Lie algebra)**.** A (real) *Lie algebra* is a (real) vector space V with a bilinear operation $[\cdot,\cdot] : V \otimes V \to V$ called the *Lie bracket* satisfying the following two properties:

(1) (antisymmetry): for any $v, w \in V$, we have $[v,w] = -[w,v]$; and
(2) (Jacobi identity): for any $u,v,w \in V$, we have $[u,[v,w]] = [[u,v],w] + [v,[u,w]]$.

If we write the operation $[u,\cdot] : V \to V$ by ad_u, then the Jacobi identity can be rewritten as the formula

$$\mathrm{ad}_u[v,w] = [\mathrm{ad}_u v, w] + [v, \mathrm{ad}_u w]$$

i.e. that ad_u satisfies a "Leibniz rule" (one also says it acts as a *derivation*) with respect to the Lie bracket operation.

Example 1.11. Let V be any vector space, and define $[\cdot,\cdot]$ to be the zero map. Then V is a *commutative* Lie algebra.

Example 1.12. Let V be an associative algebra, and for any $u, v \in V$ define $[u, v] := uv - vu$. This defines the structure of a Lie algebra on V.

Example 1.13. Let V be Euclidean 3-space, and define the Lie bracket to be cross product of vectors.

Example 1.14 (Heisenberg algebra). Let V be 3-dimensional with basis x, y, z and define $[x, y] = z$, $[x, z] = 0$, $[y, z] = 0$. The 1-dimensional subspace spanned by z is an ideal, and the quotient is a 2-dimensional commutative Lie algebra.

1.5. **Some matrix Lie groups.** Lie algebras arise most naturally as the tangent space at the identity to a Lie group. This is easiest to understand in the case of *matrix* Lie groups; i.e. groups of $n \times n$ real or complex matrices which are smooth submanifolds of $\mathbb{R}^{n^2}$ or $\mathbb{C}^{n^2}$ (with coordinates given by the matrix entries).

Example 1.15 (Examples of matrix Lie groups). The most commonly encountered examples of real and complex matrix Lie groups are:

(1) $G = \mathrm{GL}(n)$, the group of invertible $n \times n$ matrices.
(2) $G = \mathrm{SL}(n)$, the group of invertible $n \times n$ matrices with determinant 1.
(3) $G = \mathrm{O}(n)$, the group of invertible $n \times n$ matrices satisfying $A^T = A^{-1}$.
(4) $G = \mathrm{Sp}(2n)$, the group of invertible $2n \times 2n$ matrices satisfying $A^T J A = J$ where $J := \left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$.
(5) $G = \mathrm{U}(n)$, the group of invertible $n \times n$ *complex* matrices satisfying $A^* = A^{-1}$ (where A^* denotes the complex conjugate of the transpose).

If G is a matrix Lie group, the Lie algebra $\mathfrak{g}$ can be obtained as the vector space of matrices of the form $\gamma'(0)$ where $\gamma : [0, 1] \to G$ is a smooth family of matrices in G, and $\gamma(0) = \mathrm{Id}$.

1.6. **Manifolds and charts.** Let U be an open subset of $\mathbb{R}^n$. A function $F : U \to \mathbb{R}^m$ is *smooth* if the coordinate functions $F_i : U \to \mathbb{R}$ are infinitely differentiable as a function of the coordinate functions x_j on the domain. A homeomorphism $F : U \to V$ between open subsets of $\mathbb{R}^n$ is a *diffeomorphism* if it is smooth and has a smooth inverse.

A *smooth manifold* M is a manifold covered by open sets U_i (called *charts*) together with homeomorphisms $\varphi_i : U_i \to \mathbb{R}^n$ for some n, so that if $U_i \cap U_j$ is nonempty, the *transition function*

$$\varphi_{ij} := \varphi_j \circ \varphi_i^{-1} : \varphi_i(U_i \cap U_j) \to \varphi_j(U_i \cap U_j)$$

is a *diffeomorphism* between open subsets of $\mathbb{R}^n$.

1.7. **Vector fields.** Let M be a smooth manifold of dimension n. We suppose that we understand the tangent bundle of $\mathbb{R}^n$, and then we can define the tangent bundle TM by locally pulling back $T\mathbb{R}^n$ on coordinate charts, and using the derivative of the transition function to glue the bundle together on overlaps.

If f is a smooth function on $\mathbb{R}^n$ and v is a vector, then it makes sense to take the partial derivative of f in the direction v. If we fix coordinates x_i on $\mathbb{R}^n$, then we can write $v = \sum v_i \frac{\partial}{\partial x_i}$ and then

$$v(f) := \sum v_i \frac{\partial f}{\partial x_i}$$

(in the sequel we will often abbreviate $\frac{\partial}{\partial x_i}$ by ∂_i). Using coordinate functions on open charts gives us a way to take the derivative of a smooth function f on M along a vector field X on M. Note that at each point p the vector X_p acts as a derivation of the ring of germs of smooth functions at p; that is, $X_p(fg) = X_p(f)g + fX_p(g)$. In fact, a vector at a point can be *defined* as a linear derivation of the ring of germs of smooth functions at that point, and a vector field can be defined as a smoothly varying family of derivations.

We denote the space of smooth vector fields on M by $\mathfrak{X}(M)$. A smooth map $g : N \to M$ induces a smooth map $dg : TN \to TM$ satisfying $dg(X)(f) = X(f \circ g)$ for any vector field X on N and smooth function f on M. Note that the composition $f \circ g$ is also written g^*f and called the *pullback* of f by g.

A vector $v \in T_pM$ can also be thought of as an equivalence class of germ of smooth path $\gamma : [0,1] \to M$ with $\gamma(0) = v$, where $v(f) = (f \circ \gamma)'(0)$. In this case we write $\gamma'(0) = v$, although really we should write $d\gamma(\partial_t|_0) = v$, where we think of t as the coordinate on $[0,1]$, and $\partial_t|_0$ as the unit tangent to $[0,1]$ at 0.

Definition 1.16 (Lie bracket)**.** If $X, Y \in \mathfrak{X}(M)$ and f is a smooth function, we define the *Lie bracket* of X and Y, denoted $[X,Y]$, to be the vector field which acts on functions according to the formula

$$[X,Y](f) := X(Y(f)) - Y(X(f))$$

One can check that this is a derivation at each point, and varies smoothly if X and Y do, so it is a vector field.

The Lie bracket is antisymmetric and satisfies the Jacobi identity, so it makes $\mathfrak{X}(M)$ into a Lie algebra.

1.8. **Differential forms.**

The *cotangent bundle* T^*M is defined to be the dual bundle to TM; i.e. the bundle whose fiber at each point is the dual to the corresponding fiber of TM. Sections of T^*M are called *covectors* or 1-*forms*. In local coordinates x_i, a 1-form can be expressed as $\alpha := \sum \alpha_i dx_i$.

We define Ω^m to be the space of smooth sections of the bundle $\Lambda^m T^*M$, whose fiber at each point p is equal to $\Lambda^m T_p^*M$. An element of Ω^m is called a (smooth) m-form. We also define $\Omega^0 = C^\infty(M)$, the space of smooth functions on M (implicitly, we are using the "identity" $\Lambda^0 V = \mathbb{R}$ for a real vector space V). An m-form can be expressed in local coordinates as a sum

$$\omega = \sum_J \alpha_J dx_J$$

where J denotes a multi-index of length m, so that dx_J stands for an expression of the form $dx_J := dx_{j_1} \wedge dx_{j_2} \wedge \cdots \wedge dx_{j_m}$ for some $j_1 < j_2 < \cdots < j_m$.

Definition 1.17 (Exterior derivative)**.** There is a linear operator $d : \Omega^m \to \Omega^{m+1}$ defined in local coordinates by

$$d(\alpha_J dx_J) = \sum_i (\partial_i \alpha_J) dx_i \wedge dx_J$$

Exterior derivative satisfies a Leibniz rule

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^{\deg(\alpha)} \alpha \wedge d\beta$$

and furthermore satisfies $d(d\omega) = 0$ for any ω (the latter identity is usually expressed by saying $d^2 = 0$). It follows that d makes Ω^* into a chain complex of real vector spaces, whose homology is the *de Rham cohomology of* M, and is denoted $H^*_{dR}(M)$. Explicitly, an m-form ω is said to be *closed* if $d\omega = 0$, and to be *exact* if there is some $m-1$-form α with $d\alpha = \omega$. Then $H^m_{dR}(M)$ is defined to be the quotient of the vector space of closed m-forms by the vector subspace of exact m-forms.

Theorem 1.18 (De Rham isomorphism)**.** *There is a canonical isomorphism* $H^*_{dR}(M) = H^*(M;\mathbb{R})$ *where the right hand side denotes singular cohomology with* $\mathbb{R}$ *coefficients.*

This theorem is usually proved by methods of sheaf cohomology, where the necessary local ingredient is the *Poincaré Lemma*, which says that on any smooth convex open subset U of $\mathbb{R}^n$, a closed form is necessarily exact. Closed forms on M can be expressed locally as closed forms on $\mathbb{R}^n$ using coordinate charts, and the fact that for any $g : N \to M$ and any form ω on M, we have $d(g^*\omega) = g^*(d\omega)$.

Forms and collections of vector fields can be paired in the obvious way, so that if $\omega \in \Omega^m$, and $X_1, X_2, \cdots, X_m \in \mathfrak{X}(M)$ then $\omega(X_1, \cdots, X_m)$ is a smooth function obtained by contracting covectors with vectors and antisymmetrizing. A single vector field X may be contracted with an m-form ω to give an $(m-1)$-form; this is denoted $\iota_X\omega$; the operator ι_X is called *interior product* with X. Contraction and exterior product are related by the formula

$$\begin{aligned} d\omega(Y_0, \cdots, Y_m) &= \sum (-1)^i Y_i(\omega(Y_0, \cdots, \hat{Y}_i, \cdots, Y_m)) \\ &\quad + \sum_{i<j} (-1)^{i+j} \omega([Y_i, Y_j], Y_0, \cdots, \hat{Y}_i, \cdots, \hat{Y}_j, \cdots, Y_m) \end{aligned}$$

The special case

$$d\alpha(X,Y) = X\alpha(Y) - Y\alpha(X) - \alpha([X,Y])$$

for α a 1-form is very useful.

Algebraically, for each point p there is a maximal ideal $\mathfrak{m}_p$ in $C^\infty(M)$ consisting of smooth functions that vanish at p, and we can identify T^*_pM with the quotient $\mathfrak{m}_p/\mathfrak{m}_p^2$. If α is a 1-form and α_p is represented by the function $f \in \mathfrak{m}_p$ then for any vector $X_p \in T_pM$ we have $\alpha_p(X_p) = X_p(f)$. In fact, one has $\alpha_p = df_p$, so that the identity $X(f) = df(X)$ holds generally.

Wedge product and exterior derivative give Ω^* the structure of a coalgebra, which is dual (in a certain sense) to the Lie algebra $\mathfrak{X}$.

1.9. **Integration.**

1.10. **Lie derivative and Cartan's formula.**

Let M be a closed manifold (i.e. M is compact and without boundary). If X is a smooth vector field on M then for every p there is a unique smooth map $\gamma : \mathbb{R} \to M$ so that $d\gamma(\partial_t|_s) = X_{\gamma(s)}$ for all $s \in \mathbb{R}$. We say γ is an *integral curve* of X through p. For any t there is a diffeomorphism $\phi_t : M \to M$ defined by $\phi_t(\gamma(s)) = \gamma(s+t)$ for all integral curves γ as above.

If Y is a covariant tensor then we can define the *Lie derivative* $\mathcal{L}_XY$ by the formula

$$(\mathcal{L}_XY)(p) := \lim_{t\to 0} \frac{1}{t} \left(d\phi_{-t}(Y(\phi_t(p))) - Y(p)\right)$$

Similarly, if ξ is contravariant then define

$$(\mathcal{L}_X\xi)(p) := \lim_{t\to 0}\frac{1}{t}\left(\phi_t^*(\xi(\phi_t(p))) - \xi(p)\right)$$

Theorem 1.19 (Properties of Lie derivative)**.** *The Lie derivative satisfies the following properties:*

(1) for $f \in C^\infty(M)$ we have $\mathcal{L}_X f = X(f)$
(2) for $Y \in \mathfrak{X}(M)$ we have $\mathcal{L}_X Y = [X,Y]$
(3) for $\omega \in \Omega^*$ we have $\mathcal{L}_X\omega = \iota_X d\omega + d\iota_X\omega$
(4) for $\omega \in \Omega^m$ and $Y_1, Y_2, \cdots, Y_m \in \mathfrak{X}$ we have

$$\mathcal{L}_X(\omega(Y_1,\cdots,Y_m)) = (\mathcal{L}_X\omega)(Y_1,\cdots,Y_m) + \sum \omega(Y_1,\cdots,[X,Y_i],\cdots,Y_m)$$

1.11. **Frobenius' Theorem.** Let M be a smooth n manifold and let ξ be a smooth p-dimensional *distribution*; i.e. a smoothly varying p-dimensional subspace of TM at each point. A distribution is *integrable* if through every point there passes (locally) a smooth p-dimensional submanifold S so that $TS = \xi$. Locally, we can choose coordinates so that ξ is everywhere spanned by ∂_i for $1 \le i \le p$, and the submanifolds S can be taken to be translates of a coordinate subspace.

Theorem 1.20 (Frobenius)**.** *Let M be a smooth manifold, and ξ a smooth distribution. The distribution ξ is integrable if and only if one of the following two (equivalent) properties hold:*

(1) The set of vector fields X which are everywhere tangent to ξ are closed under Lie bracket; i.e. the form a Lie subalgebra of $\mathfrak{X}(M)$.
(2) The set of forms α which annihilate ξ (i.e. $\alpha(X_1, X_2, \cdots, X_r) = 0$ whenever the X_i are tangent to ξ) are a differential ideal*. That is, they are closed under taking wedge product with Ω^* and exterior d.*

Frobenius theorem gives another example of the duality between Lie algebras and differential graded algebras; here the subalgebra tangent to ξ is dual to the quotient differential algebra of forms modulo forms that annihilate ξ.

2. Some examples

2.1. **The sphere.**

2.2. **The hyperbolic plane.**

3. Riemannian metrics

3.1. **The metric.**

Definition 3.1 (Riemannian metric)**.** Let M be a smooth manifold. A *Riemannian metric* is a symmetric positive definite inner product $\langle\cdot,\cdot\rangle_p$ on T_pM for each $p \in M$ so that for any two smooth vector fields X, Y the function $p \to \langle X,Y\rangle_p$ is smooth. A *Riemannian manifold* is a smooth manifold with a Riemannian metric. For $v \in T_pM$ the *length* of v is $\langle v,v\rangle^{1/2}$, and is denoted $|v|$.

In local coordinates x_i, a Riemannian metric can be written as a symmetric tensor $g := g_{ij}dx_idx_j$. The notion of a Riemannian metric is supposed to capture the idea that a Riemannian manifold should look like Euclidean space "to first order".

Example 3.2. In $\mathbb{R}^n$ a choice of basis determines a positive definite inner product by declaring that the basis elements are orthonormal. The linear structure on $\mathbb{R}^n$ lets us identify $T_p\mathbb{R}^n$ with $\mathbb{R}^n$, and we can use this to give a Riemannian metric on $\mathbb{R}^n$. With this Riemannian metric, $\mathbb{R}^n$ becomes *Euclidean space* $\mathbb{E}^n$.

Example 3.3 (Smooth submanifold). Let S be a smooth submanifold of Euclidean space. For $p \in S$ the inner product on $T_p\mathbb{E}^n$ restricts to an inner product on T_pS thought of as a linear subspace of $T_p\mathbb{E}^n$. The same construction works for a smooth submanifold of any Riemannian manifold, and defines an *induced Riemannian metric* on the submanifold.

A smooth map $f : N \to M$ between Riemannian manifolds is an isometric immersion if for all $p \in N$, the map $df : T_pN \to T_{f(p)}M$ preserves inner products; i.e. if

$$\langle u, v\rangle_p = \langle df(u), df(v)\rangle_{f(p)}$$

for all vectors $u, v \in T_pN$. The tautological map taking a smooth submanifold to itself is an isometric immersion for the induced Riemannian metric.

A famous theorem of Nash says that any Riemannian manifold M of dimension n may be isometrically immersed (actually, isometrically embedded) in Euclidean $\mathbb{E}^N$ for $N \geq m(m+1)(3m+11)/2$. This theorem means that whenever it is convenient, we can reason about Riemannian manifolds by reasoning about smooth submanifolds of Euclidean space (however, it turns out that this is not always a simplification).

Some restriction on the dimension N are necessary.

Example 3.4 (Hilbert). The round sphere S^2 embeds isometrically in $\mathbb{E}^3$ (one can and usually does take this as the definition of the metric on S^2). The hyperbolic plane $\mathbb{H}^2$ does admit *local* isometric immersions into $\mathbb{E}^3$, but Hilbert showed that it admits no *global* immersion. The difference is that a positively curved surface can be locally developed as the boundary of a *convex* region, and the rigidity this imposes lets one find global isometric immersions (the best theorem in this direction by Alexandrov says that a metric 2-sphere with non-negative curvature (possibly distributional) can be isometrically realized as the boundary of a convex region in $\mathbb{E}^3$, uniquely up to isometry). If one tries to immerse a negatively curved surface in $\mathbb{E}^3$, one must "fold" it in order to fit in the excessive area; the choice of direction to fold imposes more and more constraints, and as one tries to extend the immersion the folds accumulate and cause the surface to become singular.

If $\langle\cdot,\cdot\rangle$ and $\langle\cdot,\cdot\rangle'$ are two positive definite inner products on a vector space, then so is any convex combination of the two products (i.e. the set of positive definite inner products on a vector space V is a convex subset of S^2V^*). Thus any smooth manifold may be given a Riemannian metric by taking convex combinations of inner products defined in charts, using a partition of unity.

Definition 3.5. If $\gamma : I \to M$ is a smooth path, the *length* of γ is defined to be

$$\text{length}(\gamma) := \int_0^1 |\gamma'(t)|dt$$

and the *energy* of γ is

$$\text{energy}(\gamma) := \frac{1}{2}\int_0^1 \langle \gamma'(t), \gamma'(t)\rangle_{\gamma(t)} dt$$

Energy here should be thought of the elastic energy of a string made from some flexible material. There is another physical interpretation of this quantity: the integrand can be thought of as the kinetic energy of a moving particle of unit mass, and then the integral is what is known to physicists as the *action*.

Note that the length of a path does not depend on the parameterization, but the energy does. In fact, by the Cauchy-Schwarz inequality, for a given path the parameterization of least energy is the one for which $|\gamma'|$ is constant; explicitly, the least energy for a parameterization of a path of length ℓ is $\ell^2/2$.

A path-connected Riemannian manifold is a metric space (in the usual sense), by defining the distance from p to q to be the infimum of the length of all smooth paths from p to q. A Riemannian manifold is *complete* if it is complete (in the usual sense) as a metric space.

3.2. **Connections.**

Definition 3.6. If E is a smooth bundle on M, a *connection* on E is a bilinear map

$$\nabla : \mathfrak{X}(M) \times \Gamma(E) \to \Gamma(E)$$

where we write $\nabla(X, W)$ as $\nabla_X W$, satisfying the properties

(1) (tensor): $\nabla_{fX} W = f\nabla_X W$; and
(2) (Leibniz): $\nabla_X(fW) = (Xf)W + f\nabla_X W$.

Note that since ∇ is tensorial in the first term, we can also write $\nabla W \in \Gamma(T^*M)\otimes\Gamma(E)$.

Definition 3.7. Let E be a smooth bundle, and ∇ a connection on E. If $\gamma : I \to M$ is a smooth path, a section W of E is *parallel along* γ if $\nabla_{\gamma'} W = 0$ along $\gamma(I)$.

Really we should think of the pullback bundle γ^*E over I. Usual existence and uniqueness theorems for ODEs imply that for any $W_0 \in E_{\gamma(0)}$ there is a unique extension of W_0 to a parallel section W of γ^*E over I. Thus, a connection on E gives us a canonical way to identify fibers of E along a smooth path in M. If we let e_i be a basis for E locally, then we can express any W as $W = \sum w_i e_i$. Then by the properties of a connection,

$$0 = \nabla_{\gamma'} W = \sum \gamma'(w_i)e_i + \sum w_i \nabla_{\gamma'} e_i$$

which is a system of first order linear ODEs in the variables w_i.

A connection ∇ on E determines a connection on E^* implicitly (which by abuse of notation we also write ∇) by the Leibniz formula

$$X(\alpha(W)) = (\nabla_X \alpha)W + \alpha(\nabla_X W)$$

In particular, a family of bases for E which is parallel along some path is dual to a family of bases for E^* which are parallel along the same path.

Example 3.8. If E is a trivialized bundle $E = M \times \mathbb{R}^n$ then by convention $\nabla_X = X$ on E. Hence $\nabla_X f = X(f)$ for a function f (equivalently, $\nabla f = df$ on functions).

Example 3.9 (Functorial construction of connections). There are several functorial constructions of connections. For example, if ∇_i are connections on bundles E_i for $i = 1, 2$ we can define

$$(\nabla_1 \oplus \nabla_2)(W_1 \oplus W_2) := \nabla_1 W_1 \oplus \nabla_2 W_2$$

and

$$(\nabla_1 \otimes \nabla_2)(W_1 \otimes W_2) := \nabla_1 W_1 \otimes W_2 + W_1 \otimes \nabla_2 W_2$$

In this way a connection on a bundle E induces connections on $\Lambda^n E$ and $S^n E$ etc.

Lemma 3.10. ∇ *commutes with contraction of tensors.*

Proof. If W is a section of E and V is a section of E^* then

$$\nabla_X(V \otimes W) = (\nabla_X V) \otimes W + V \otimes (\nabla_X W)$$

whereas

$$X(V(W)) = (\nabla_X V)(W) + V(\nabla_X W)$$

as can be seen by expressing V and W in terms of local dual parallel (along some path) bases and using the Leibniz rule. This proves the lemma when there are two terms. More generally, by the properties of a connection,

$$\begin{aligned}\nabla_X(V(W)A \otimes \cdots \otimes Z) &= X(V(W))(A \cdots Z) + \sum V(W)(A \cdots \nabla_X I \cdots Z) \\ &= (\nabla_X V)(W)(A \cdots Z) + V(\nabla_X W)(A \cdots Z) + \sum V(W)(A \cdots \nabla_X I \cdots Z)\end{aligned}$$

proving the formula in general. $\square$

As a special case, if $A \in E_1^* \otimes \cdots \otimes E_n^*$ and $Y_i \in E_i$ then

$$X(A(Y_1, \cdots, A_n)) = (\nabla_X A)(Y_1, \cdots, Y_n) + \sum_i A(Y_1, \cdots, \nabla_X Y_i, \cdots, Y_n)$$

Definition 3.11. If E is a bundle with a (fibrewise) inner product (i.e. a section $q \in \Gamma(S^2 E^*)$) a connection on E is *metric* (or *compatible with the inner product*) if $\nabla q = 0$; equivalently, if

$$d(q(U, V)) = q(\nabla U, V) + q(U, \nabla V)$$

Plugging in $X \in \mathfrak{X}(M)$, the metric condition is equivalent to $Xq(U, V) = q(\nabla_X U, V) + q(U, \nabla_X V)$.

If E has a (fiberwise) metric, and ∇ is a metric connection, then $|W|$ is constant along any path for which W is parallel.

3.3. **The Levi-Civita connection.** Suppose M is a Riemannian manifold. A connection ∇ on TM is metric if $X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$ for all $X, Y, Z \in \mathfrak{X}(M)$. A Riemannian manifold usually admits many different metric connections. However, there is one *distinguished* metric connection which uses the symmetry of the two copies of $\mathfrak{X}(M)$ in the definition of ∇.

Definition 3.12. Let ∇ be a connection on TM. The *torsion* of ∇ is the expression

$$\mathrm{Tor}(V, W) := \nabla_V W - \nabla_W V - [V, W]$$

A connection on TM is *torsion-free* if $\mathrm{Tor} = 0$.

The properties of a connection imply that Tor is a *tensor*.

A connection on TM induces a connection on T^*M and on S^2T^*M and so forth. The condition that ∇ is a metric connection is exactly that the metric $g \in S^2T^*M$ is parallel. The condition that ∇ is torsion-free says that for the induced connection on T^*M, the composition

$$\Gamma(T^*M) \xrightarrow{\nabla} \Gamma(T^*M \otimes T^*M) \xrightarrow{\pi} \Gamma(\Lambda^2 T^*M)$$

is equal to exterior d, where π is the quotient map from tensors to antisymmetric tensors.

Theorem 3.13 (Levi-Civita)**.** *Let M be a Riemannian manifold. There is a unique torsion-free metric connection on TM called the* Levi-Civita connection*.*

Proof. Suppose ∇ is metric and torsion-free. By the metric property, we have

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

and similarly for the other two cyclic permutations of X, Y, Z. Adding the first two permutations and subtracting the third, and using the torsion-free property to eliminate expressions of the form $\nabla_X Y - \nabla_Y X$ gives the identity

$$2\langle \nabla_X Y, Z\rangle = X\langle Y, Z\rangle + Y\langle X, Z\rangle - Z\langle X, Y\rangle + \langle [X, Y], Z\rangle - \langle [X, Z], Y\rangle - \langle [Y, Z], X\rangle$$

This shows uniqueness. Conversely, defining ∇ by this formula one can check that it satisfies the properties of a connection. □

Example 3.14. On Euclidean $\mathbb{E}^n$ with global coordinates x_i the vector fields ∂_i are parallel in the Levi-Civita connection. If Y is a vector field, X_p is a vector at p and $\gamma : [0, 1] \to \mathbb{E}^n$ is a smooth path with $\gamma(0) = p$ and $\gamma'(0) = X_p$ then $(\nabla_X Y)(p) = \frac{d}{dt} Y(\gamma(t))|_{t=0}$.

Example 3.15. If S is a smooth submanifold of $\mathbb{E}^n$, the ordinary Levi-Civita connection on $\mathbb{E}^n$ restricts to a connection on the bundle $T\mathbb{E}^n|_S$. If we think of TS as a subbundle, then at every $p \in S$ there is a natural orthogonal projection map $\pi : T_p\mathbb{E}^n \to T_pS$. We can then define a connection $\nabla^T := \pi \circ \nabla$ on TS; i.e. the "tangential part" of the ambient connection. For $X, Y, Z \in \mathfrak{X}(S)$ we have

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle = \langle \nabla^T_X Y, Z\rangle + \langle Y, \nabla^T_X Z\rangle$$

so ∇^T is a metric connection. Similarly, since $[X, Y]$ is in $\mathfrak{X}(S)$ whenever X and Y are, the perpendicular components of $\nabla_X Y$ and of $\nabla_Y X$ are equal, so ∇^T is torsion-free. It follows that ∇^T *is* the Levi-Civita connection on S.

The case of a smooth submanifold is instructive. We can imagine defining parallel transport along a path γ in S by "rolling" the tangent plane to S along γ, infinitesimally projecting it to TS as we go. Since the projection is orthogonal, the plane does not "twist" in the direction of TS as it is rolled; this is the geometric meaning of the fact that this connection is torsion-free. In the language of flight dynamics, there is *pitch* where the submanifold S is not flat and *roll* where the curve γ is not "straight", but no *yaw*; see Figure 1. By the Nash embedding theorem, the Levi-Civita connection on any Riemannian manifold can be thought of in these terms.

Suppose we choose local smooth coordinates $x_1, \cdots, x_n$ on M, and vector fields $\partial_i := \frac{\partial}{\partial x_i}$ which are a basis for TM at each point locally. To define a connection we just need to

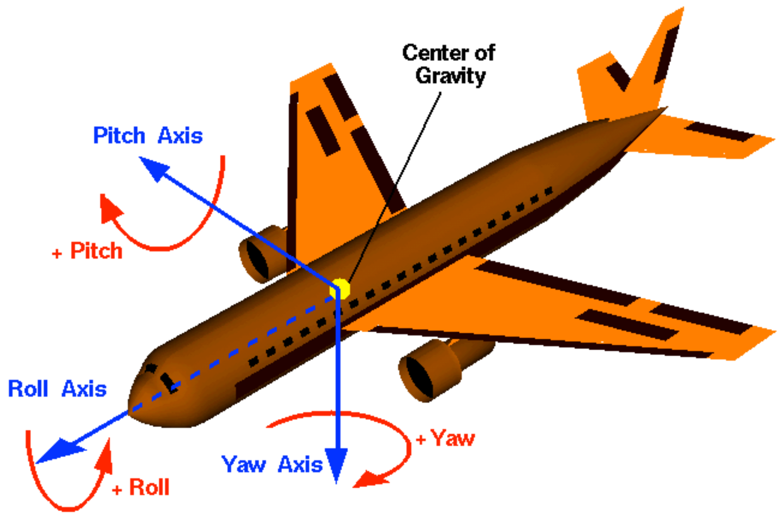


FIGURE 1. The Levi-Civita connection pitches and rolls but does not yaw (figure courtesy of NASA [8]).

give the values of $\nabla_{\partial_i}\partial_j$ for each i, j (to reduce clutter we sometimes abbreviate ∇_{∂_i} to ∇_i when the coordinates x_i are understood).

Definition 3.16 (Christoffel symbols)**.** With respect to local coordinates x_i, the *Christoffel symbols* of a connection ∇ on TM are the functions Γ^k_{ij} defined by the formula

$$\nabla_i \partial_j = \sum_k \Gamma^k_{ij} \partial_k$$

3.4. **Second fundamental form.** Suppose $N \subset M$ is a smoothly embedded submanifold. We write $\nabla^\top$ and $\nabla^\perp$ for the components of ∇ in TN and νN, the normal bundle of N in M (so that $TM|N = TN \oplus \nu N$).

Definition 3.17 (Second fundamental form)**.** For vectors $X, Y \in T_pN$ define the *second fundamental form* $\mathrm{II}(X, Y) \in \nu_p N$ by extending X and Y to vector fields on N (locally), and using the formula $\mathrm{II}(X, Y) := \nabla^\perp_X Y$.

Lemma 3.18. *The second fundamental form is tensorial (and therefore well-defined) and symmetric in its two terms. i.e. it is a section* $\mathrm{II} \in \Gamma(S^2(T^*N) \otimes \nu N)$.

Proof. Evidently $\mathrm{II}(X, Y)$ is tensorial in X, so it suffices to show that it is symmetric. We compute

$$\nabla^\perp_X Y = \nabla_X Y - \nabla^\top_X Y = \nabla_Y X - \nabla^\top_Y X = \nabla^\perp_Y X$$

where we use the fact that both ∇ and $\nabla^\top$ are torsion-free. □

Remark 3.19. The "first fundamental form" on N is simply the restriction of the inner product on M to TN; i.e. it is synonymous with the induced Riemannian metric on N.

4. Geodesics

4.1. **First variation formula.** If $\gamma : [0,1] \to M$ is a smooth curve, a *smooth variation* of γ is a map $\Gamma : [0,1] \times (-\epsilon, \epsilon) \to M$ such that $\Gamma(\cdot, 0) : [0,1] \to M$ agrees with γ. For each $s \in (-\epsilon, \epsilon)$ we denote the curve $\Gamma(\cdot, s) : [0,1] \to M$ by $\gamma_s : [0,1] \to M$ and think of this as defining a 1-parameter family of smooth curves.

Let t and s be the coordinates on the two factors of $[0,1] \times (-\epsilon, \epsilon)$. Pulling back the bundle TM to $[0,1] \times (-\epsilon, \epsilon)$ under Γ^* lets us think of the vector fields $T := \frac{\partial}{\partial t}$ and $S := \frac{\partial}{\partial s}$ as vector fields on M locally, and compute their derivatives with respect to the connection on $\Gamma^* TM$ pulled back from the Levi-Civita connection. Note that in this language, $T := \gamma_s'$. This lets us compute the derivative of $\text{length}(\gamma_s)$ with respect to s. Since length does not depend on parameterization, we are free to choose a parameterization for which $|\gamma'|$ is constant and equal to $\ell := \text{length}(\gamma)$.

Theorem 4.1 (First variation formula)**.** *Let $\gamma : [0,1] \to M$ be a smooth curve parameterized proportional to arclength with* $\text{length}(\gamma) = \ell$*, and let $\Gamma : [0,1] \times (-\epsilon, \epsilon) \to M$ be a smooth one-parameter variation. Let $\gamma_s : [0,1] \to M$ denote the restriction $\Gamma(\cdot, s) : [0,1] \to M$. Then there is a formula*

$$\frac{d}{ds}\text{length}(\gamma_s)|_{s=0} = \ell^{-1}\left(\langle S, T\rangle|_0^1 - \int_0^1 \langle S, \nabla_T T\rangle dt\right)$$

Proof. This is an exercise in the properties of the Levi-Civita connection. By definition, we have

$$\begin{aligned}\frac{d}{ds}\text{length}(\gamma_s) &= \frac{d}{ds}\int_0^1 \langle T, T\rangle^{1/2} dt = \int_0^1 S\langle T, T\rangle^{1/2} dt \\ &= \int_0^1 \langle T, T\rangle^{-1/2}\langle \nabla_S T, T\rangle dt\end{aligned}$$

where we used the fact that the Levi-Civita connection is a metric connection. Since γ is parameterized proportional to arclength, we have $\langle T, T\rangle^{-1/2} = \ell^{-1}$ at $s = 0$. Since S and T are the derivatives of coordinate functions, $[S, T] = 0$. Since the Levi-Civita connection is torsion-free, we deduce

$$\frac{d}{ds}\text{length}(\gamma_s)|_{s=0} = \ell^{-1}\int_0^1 \langle \nabla_T S, T\rangle dt = \ell^{-1}\int_0^1 T\langle S, T\rangle - \langle S, \nabla_T T\rangle dt$$

where we used again the fact that the Levi-Civita connection is a metric connection to replace $\langle \nabla_T S, T\rangle = T\langle S, T\rangle - \langle S, \nabla_T T\rangle$. Integrating out $\int_0^1 T\langle S, T\rangle dt = \langle S, T\rangle|_0^1$ we obtain the desired formula. □

4.2. **Geodesics.** It follows that if Γ is a smooth variation with endpoints fixed (i.e. with $S = 0$ at 0 and 1) then γ (parameterized proportional to arclength) is a critical point for length if and only if $\nabla_{\gamma'}\gamma' := \nabla_T T = 0$.

Definition 4.2 (Geodesic)**.** A smooth curve $\gamma : [a, b] \to M$ is a *geodesic* if it satisfies $\nabla_{\gamma'}\gamma' = 0$.

Note that $\gamma'\langle\gamma',\gamma'\rangle = 2\langle\nabla_{\gamma'}\gamma',\gamma'\rangle$ so geodesics are necessarily parameterized proportional to arclength. Recall that for any smooth curve, the unique parameterization minimizing energy is the one proportional to arclength. It follows that geodesics are also critical points for the first variation of *energy*. This is significant, because as critical points for length, geodesics are (almost always) *degenerate*, since they admit an infinite dimensional space of reparameterizations which do not affect the length. By contrast, *most* geodesics are *nondegenerate* critical points for energy, and the Hessian of the energy functional always has a finite dimensional space on which it is null or negative definite (so that the *index* of a geodesic as a critical point for energy can be defined). We return to this point when we come to derive the Second variation formula in § 5.7.

Geodesics have the following *homogeneity property*: if $\gamma : (-\epsilon,\epsilon) \to M$ is a smooth geodesic with $\gamma(0) = p$ and $\gamma'(0) = v \in T_pM$, then for any $T \neq 0$ the map $\sigma : (-\epsilon/T, \epsilon/T) \to M$ defined by $\sigma(t) = \gamma(tT)$ is also a smooth geodesic, now with $\sigma(0) = p$ and $\sigma'(0) = Tv$. In other words, at least on their maximal domains of definition, two geodesics with the same initial point and proportional derivatives at that point have the same image, and differ merely by parameterizing that image at different (constant) speeds. By taking T sufficiently small, σ may be defined on the entire interval $[0,1]$.

In local coordinates the equations for a geodesic can be expressed in terms of Christoffel symbols. Recall that we defined the Christoffel symbols in terms of local coordinates x_i by the formula

$$\nabla_i \partial_j = \sum_k \Gamma_{ij}^k \partial_k$$

We think of the coordinates x_i as functions of t implicitly by $x_i(t) := x_i(\gamma(t))$, and then $\gamma' = \sum_i x_i' \partial_i$. With this notation, the equations for a geodesic can be expressed as

$$\begin{aligned} 0 = \nabla_{\gamma'}\gamma' &= \nabla_{\sum_i x_i'\partial_i} \sum_j x_j' \partial_j \\ &= \sum_i x_i' \nabla_i \sum_j x_j' \partial_j = \sum_k x_k'' \partial_k + \sum_{i,j,k} x_i' x_j' \Gamma_{ij}^k \partial_k \end{aligned}$$

where we used the Leibniz rule for a connection, and the chain rule $\sum_i x_i' \partial_i x_k' = x_k''$. This reduces, for each k, to the equation

$$x_k'' + \sum_{i,j} x_i' x_j' \Gamma_{ij}^k = x_k''(t) + \sum_{i,j} x_i'(t) x_j'(t) \Gamma_{ij}^k(x_1(t), \cdots, x_n(t)) = 0$$

where we have stressed with our notation the fact that each Γ_{ij}^k is a (possibly complicated) function of the x_is which in turn are a function of t. This is a system of second order ODEs for the functions $x_i(t)$, and therefore there is a unique solution defined on some interval $t \in (-\epsilon,\epsilon)$ for given initial values $\gamma(0) = p$ and $\gamma'(0) = v$.

4.3. **The exponential map.** Because of local existence and uniqueness of geodesics with prescribed initial values and initial derivatives, we can make the following definition:

Definition 4.3 (Exponential map)**.** The *exponential map* exp is a map from a certain open domain U in TM (containing the zero section) to M, defined by $\exp(v) = \gamma_v(1)$, where

$v \in T_pM$, and $\gamma : [0,1] \to M$ is the unique smooth geodesic with $\gamma(0) = p$ and $\gamma'(1) = v$ (if it exists). The restriction of exp to its domain of definition in T_pM is denoted $\exp_p$.

The homogeneity property of geodesics (discussed in § 4.2) shows that $\exp_p$ is defined on an open (star-shaped) subset of T_pM containing the origin, for each p.

Lemma 4.4. *For all p there is an open subset $U \subset T_pM$ containing the origin so that the restriction $\exp_p : U \to M$ is a diffeomorphism onto its image.*

Proof. Since $\exp_p$ is smooth, it suffices to show that $d\exp_p : T_0T_pM \to T_pM$ is nonsingular. But if we identify T_0T_pM with T_pM using its linear structure, the definition of exp implies that $d\exp_p : T_pM \to T_pM$ is the identity map; in particular, it is nonsingular. □

Lemma 4.4 is a purely local statement; the map $\exp_p$, even if it is defined on all of T_pM, is typically not a global diffeomorphism, or even a covering map.

Example 4.5. On S^2, the geodesics are the arcs of great circles, parameterized at unit speed. Thus for any point p, the map $\exp_p$ is a diffeomorphism from the open unit disk of radius π in T_pM to $S^2 - q$, where q is antipodal to p. But $\exp_p$ maps the entire boundary circle identically to the point q, so $d\exp_p$ does not even have full rank there.

The exponential map lets us define certain kinds of local coordinates in which some calculations simplify considerably. Let e_i be an orthonormal basis of T_pM, and define smooth coordinates x_i on $\exp_p(U)$ (for some sufficiently small neighborhood U of 0 in T_pM so that the restriction of $\exp_p$ to this neighborhood is a diffeomorphism) by letting $x_1, \cdots, x_n$ be the coordinates of the point $\exp_p(\sum x_i e_i)$. These coordinates are called *normal coordinates*. In these coordinates, the geodesics through the origin have the form $x_i(t) = a_i t$ for some arbitrary collection of real constants a_i. Thus the geodesic equations at $t = 0$ (i.e. at the origin) reduce to $\sum_{i,j} a_i a_j \Gamma_{ij}^k|_p = 0$ for all k. Since a_i and a_j are arbitrary, it follows that we have $\Gamma_{ij}^k|_p = 0$; i.e. $\nabla_v \partial_i|_p = 0$ for any vector $v \in T_pM$.

Lemma 4.6 (Gauss' Lemma)**.** *If $\rho(t) := tv$ is a ray through the origin in T_pM and $w \in T_{\rho(t)}T_pM$ is perpendicular to $\rho'(t)$, then $d\exp_p(w)$ is perpendicular to $d\exp_p(\rho'(t))$.*

In words, Gauss' Lemma says that if we exponentiate (orthogonal) polar coordinates r, θ_i on T_pM, then the image of the radial vector field $d\exp_p(\partial_r)$ (which is tangent to the geodesics through p) is perpendicular to the level sets $\exp(r = \text{constant})$.

Proof. Use polar coordinates r, θ on the 2-dimensional subspace of T_pM spanned by w and v. The vector field $R := d\exp_p(\partial_r)$ is everywhere tangent to the geodesics through the origin, and has constant speed; i.e. $|R| = \ell$ (we scale the polar coordinates so that we are interested at $d\exp$ at the point in T_pM where $r = 1$ and $\theta = 0$). Let $T := d\exp_p(\partial_\theta)$, which vanishes at p. We want to show that $\langle R, T\rangle = 0$. We can think of R and T as tangent vector fields along a 1-parameter variation of smooth curves such that the integral curves of R with a constant value of θ are geodesics γ_θ through p parameterized at unit speed, and therefore have *constant* length. Thus by the first variation formula,

$$0 = \frac{d}{d\theta}\text{length}(\gamma_\theta) = \ell^{-1}\left(\langle T, R\rangle|_0^1 - \int_0^1 \langle T, \nabla_R R\rangle dt\right)$$

But $\nabla_R R = 0$ by the geodesic property, and $T(0) = 0$. So $\langle R, T\rangle = 0$, as claimed. □

By abuse of notation we write the function $r \circ \exp_p^{-1}$ on M just as r, and we let ∂_r denote the vector field on M obtained by pushing forward the radial vector field ∂_r on T_pM. Gauss' Lemma can be expressed by saying that ∂_r (on M) is the *gradient vector field* of the function r (also on M). By this we mean the following. Any vector field V on M can be written (at least near p) in the form $V = \sum v_i \partial_i + v_r \partial_r$ where ∂_i is short for $d\exp_p(\frac{\partial}{\partial \theta_i})$ in polar coordinates on T_pM. Gauss' Lemma is the observation that $\langle \partial_i, \partial_r \rangle = 0$ throughout the image of $\exp_p$. Then $v_r = X(r) = \langle \partial_r, X\rangle$ (in general, the gradient vector field of a function f is the unique vector field $\mathrm{grad}(f)$ defined by the property $\langle \mathrm{grad}(f), X\rangle = X(f)$ for all vectors X; it is obtained from df by using the inner product to identify $\Gamma(T^*M)$ with $\Gamma(TM)$).

Gauss' Lemma is the key to showing that geodesics are locally unique distance minimizers. That is,

Corollary 4.7. *Let $B_r(0) \subset T_pM$ be a ball of radius r on which $\exp_p$ is a diffeomorphism. Then the following are true:*

1. *For any $v \in B_r(0)$, the curve $\gamma_v : [0,1] \to M$ is the unique curve joining p to $\exp_p(v)$ of length at most $|v|$ (up to reparameterization). Thus on $\exp_p(B_r(0))$ the function $r \circ \exp_p^{-1}$ (where r is radial distance in T_pM) agrees with the function $\mathrm{dist}(p,\cdot)$.*
2. *If q is not in $\exp_p(B_r(0)) =: B_r(p)$ then there is some q' in $\partial B_r(p)$ so that*
$$\mathrm{dist}(p,q) = \mathrm{dist}(p,q') + \mathrm{dist}(q',q)$$
In particular, $\mathrm{dist}(p,q) \ge r$.

Proof. Let $\sigma : [0,1] \to M$ be a smooth curve from p to $\exp_p(v)$. When restricted to the part with image in $\exp_p(B_r(0))$, we have an inequality $|\sigma'| \ge \langle \sigma', \partial_r\rangle = \sigma'(r)$ and therefore

$$\mathrm{length}(\sigma) = \int_0^1 |\sigma'| dt \ge r(\sigma(1)) - r(\sigma(0))$$

with equality if and only if σ' is of the form $f\partial_r$ for some non-negative function f. This proves the first claim.

To prove the second part of the claim, for any smooth $\sigma : [0,1] \to M$ from p to q, there is some first point $q'(\sigma) \in \partial B_r(0)$ on the curve, and the length of the path from p to $q'(\sigma)$ is at least r. Taking a sequence of curves whose length approaches $\mathrm{dist}(p,q)$, and using the compactness of $\partial B_r(0)$, we extract a subsequence for which there is a limit q' as in the statement of the claim. □

4.4. **The Hopf-Rinow Theorem.**

Since geodesics are so important, and their short time existence and uniquess are so useful, it is important to know when they can be extended for all time. The answer is given by the Hopf-Rinow theorem.

Theorem 4.8 (Hopf-Rinow)**.** *The following are equivalent:*

1. *M is a complete metric space with respect to* dist*; or*
2. *for some $p \in M$ the map $\exp_p$ is defined on all T_pM; or*
3. *for every $p \in M$ the map $\exp_p$ is defined on all T_pM.*

Any of these conditions imply that any two points p, q of M can be joined by a geodesic γ with $\mathrm{length}(\gamma) = \mathrm{dist}(p, q)$.

Note that the first hypothesis is evidently satisfied whenever M is closed (i.e. compact and without boundary). Note too that the last conclusion is definitely weaker than the first three conditions; for example, it is satisfied by the open unit disk in $\mathbb{E}^n$, which is not complete as a metric space.

Proof. Suppose $\exp_p : T_pM \to M$ is globally defined, and let $q \in M$ be arbitrary. There is some $v \in T_pM$ with $|v| = 1$ so that $\mathrm{dist}(p, \exp_p(sv)) + \mathrm{dist}(\exp_p(sv), q) = \mathrm{dist}(p, q)$ for some $s > 0$. Let $\gamma : [0, \infty) \to M$ be the geodesic with $\gamma(0) = p$ and $\gamma'(0) = v$, so that $\gamma(t) = \exp_p(tv)$. The set of t such that $\mathrm{dist}(p, \gamma(t)) + \mathrm{dist}(\gamma(t), q) = \mathrm{dist}(p, q)$ is closed, so let t be maximal with this property. We claim $\gamma(t) = q$ and $|t| = \mathrm{dist}(p, q)$. For if not, there is some small r and some point $q' \in \partial B_r(\gamma(t))$ so that $\mathrm{dist}(\gamma(t), q') + \mathrm{dist}(q', q) = \mathrm{dist}(\gamma(t), q)$. Let $\sigma : [0, r] \to M$ be the unit speed geodesic with $\sigma(0) = \gamma(t)$ and $\sigma(r) = q'$. Then $\mathrm{dist}(p, q') = \mathrm{length}(\gamma([0, t]) \cup \sigma([0, r]))$, and therefore these two paths fit together at $\sigma(0)$ to form a smooth geodesic, contrary to the definition of t. In particular, it follows that $\exp_p$ is *surjective*, and for every q there is a geodesic of length $\mathrm{dist}(p, q)$ from p to q. Now if q_i is a Cauchy sequence, we can find $v_i \in T_pM$ with $\exp_p(v_i) = q_i$ and $|v_i| = \mathrm{dist}(p, q_i)$. By compactness, the v_i have a subsequence converging to some v, and $\exp_p(v) = q$ is a limit of the q_i. This shows that (2) implies (1).

Conversely, suppose M is complete with respect to dist. We deduce that $\exp_p$ is defined everywhere for every p. Let $v \in T_pM$ and let t be the supremum of the numbers s so that $\exp_p(sv)$ is defined. Then $\exp_p(sv)$ is a Cauchy sequence as $s \to t$, and therefore limits to some point q. It follows that $\gamma_v([0, t))$ extends continuously by adding an endpoint q. In a small ball centered around q there is some $q' = \gamma_v(s)$ and $w \in T_qM$ with $\exp_q(w) = q'$. Then defining $\sigma_w(u) = \exp_q(uw)$ for small u we get that $\sigma_w((0, u])$ agrees with $\gamma_v([t-u, t))$ (with opposite orientation) so we may take the union of $\gamma_v([0, t))$ with $\sigma_w([-u, u])$ and thereby extend γ_v. In particular, t is infinite, and $\exp_p$ is globally defined for any p. This shows that (1) implies (3). The implication (3) implies (2) is obvious.

Finally, the argument we already gave showed that (3) implies that any two points p, q can be joined by a geodesic of length $\mathrm{dist}(p, q)$. $\square$

5. Curvature

5.1. **Curvature.** The failure of holonomy transport along commuting vector fields to commute itself is measured by *curvature*. Informally, curvature measures the infinitesimal extent to which parallel transport depends on the path joining two endpoints.

Definition 5.1 (Curvature)**.** Let E be a smooth bundle with a connection ∇. The *curvature* (associated to ∇) is a trilinear map

$$R : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \Gamma(E) \to \Gamma(E)$$

which we write $R(X, Y)Z \in \Gamma(E)$, defined by the formula

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

Remark 5.2. It is an unfortunate fact that many authors use the notation $R(X,Y)Z$ to denote the *negative* of the expression for $R(X,Y)Z$ given in Definition 5.1. This is an arbitrary choice, but the choice propagates, and it makes it difficult to use published formulae without taking great care to check the conventions used. Our choice of notation is consistent with Cheeger-Ebin [2] and with Kobayashi-Nomizu [5] but is *inconsistent* with Milnor [7].

Although *a priori* it appears to depend on the second order variation of Z near each point, it turns out that the curvature is a *tensor*. The following proposition summarizes some elementary algebraic properties of R.

Proposition 5.3 (Properties of curvature)**.** *For any connection ∇ on a bundle E the curvature satisfies the following properties:*

1. *(tensor): $R(fX,gY)(hZ) = (fgh)R(X,Y)Z$ for any smooth f,g,h*
2. *(antisymmetry): $R(X,Y)Z = -R(Y,X)Z$*
3. *(metric): if ∇ is a metric connection, then $\langle R(X,Y)Z,W\rangle = -\langle R(X,Y)W,Z\rangle$*

Thus we can think of $R(\cdot,\cdot)$ as a section of $\Omega^2(M)\otimes\Gamma(\mathrm{End}(E))$ with coefficients in the Lie algebra of the orthogonal group of the fibers. If $E = TM$ and ∇ is torsion-free, then it satisfies the so-called Jacobi identity*:*

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

Consequently, the Levi-Civita connection on TM satisfies the following symmetry:

$$\langle R(X,Y)Z,W\rangle = \langle R(Z,W)X,Y\rangle$$

The Jacobi identity is sometimes also called the *first Bianchi identity*. The symmetry/antisymmetry identities, and the fact that R is a tensor, means that if we define $R(X,Y,Z,W) := \langle R(X,Y)Z,W\rangle$ then $R\in\Gamma(S^2\Lambda^2(T^*M))$.

Proof. Antisymmetry is obvious from the definition. We compute: $\nabla_{fX}\nabla_Y Z = f\nabla_X\nabla_Y Z$ whereas

$$\nabla_Y\nabla_{fX}Z = \nabla_Y(f\nabla_X Z) = f\nabla_Y\nabla_X Z + Y(f)\nabla_X Z$$

on the other hand $[fX,Y] = f[X,Y] - Y(f)X$ so

$$\nabla_{[fX,Y]}Z = f\nabla_{[X,Y]}Z - Y(f)\nabla_X Z$$

so R is tensorial in the first term. By antisymmetry it is tensorial in the second term. Finally,

$$\begin{aligned}\nabla_X\nabla_Y(fZ) &= \nabla_X f\nabla_Y Z + \nabla_X Y(f)Z \\ &= f\nabla_X\nabla_Y Z + X(f)\nabla_Y Z + Y(f)\nabla_X Z + X(Y(f))Z\end{aligned}$$

and there is a similar formula for $\nabla_Y\nabla_X(fZ)$ with X and Y reversed, whereas

$$\nabla_{[X,Y]}(fZ) = f\nabla_{[X,Y]}Z + (X(Y(f)) - Y(X(f)))\,Z$$

and we conclude that R is tensorial in the third term too. To see the metric identity, let's use the tensoriality of R to replace X and Y by commuting vector fields with the same value at some given point, and compute

$$\begin{aligned}\langle \nabla_X \nabla_Y Z, W\rangle &= X\langle \nabla_Y Z, W\rangle - \langle \nabla_Y Z, \nabla_X W\rangle \\ &= X(Y\langle Z, W\rangle) - X\langle Z, \nabla_Y W\rangle - \langle \nabla_Y Z, \nabla_X W\rangle \\ &= X(Y\langle Z, W\rangle) - \langle \nabla_X Z, \nabla_Y W\rangle - \langle \nabla_Y Z, \nabla_X W\rangle - \langle Z, \nabla_X \nabla_Y W\rangle\end{aligned}$$

Subtracting off $\langle \nabla_Y \nabla_X Z, W\rangle$ expanded similarly, the first terms cancel (since by hypothesis $[X, Y] = 0$), the second and third terms cancel identically, and we are left with $-\langle Z, R(X, Y)W\rangle$ as claimed.

To prove the Jacobi identity when ∇ on TM is torsion-free, we again use tensoriality to reduce to the case of commuting vector fields. Then the term $\nabla_X \nabla_Y Z$ in $R(X, Y)Z$ can be rewritten as $\nabla_X \nabla_Z Y$ which cancels a term in $R(Z, X)Y$ and so on.

The last symmetry (under interchanging (X, Y) with (Z, W)) follows formally from the metric property, the antisymmetry of R under interchanging X and Y, and the Jacobi identity. □

5.2. **Curvature and representation theory of** $\mathrm{O}(n)$**.** The full Riemann curvature tensor is difficult to work with directly; fortunately, there are simpler "curvature" tensors capturing some of the same information, that are easier to work with.

Definition 5.4 (Ricci curvature)**.** The *Ricci curvature tensor* Ric is the 2-tensor

$$\mathrm{Ric}(X, Y) = \text{trace of the map } Z \to R(Z, X)Y$$

If we choose an orthonormal basis e_i then $\mathrm{Ric}(X, Y) = \sum_i \langle R(e_i, X)Y, e_i\rangle$. The symmetries of the Riemann curvature tensor imply that Ric is a symmetric bilinear form on T_pM at each point p.

Definition 5.5 (Scalar curvature)**.** The *scalar curvature* s is the trace of Ric (relative to the Riemannian metric); i.e. $s = \sum_i \mathrm{Ric}(e_i, e_i)$.

The *trace-free Ricci tensor*, denoted Ric_0, is the normalization

$$\mathrm{Ric}_0 = \mathrm{Ric} - \frac{s}{n} g$$

where g denotes the metric.

The definitions of the Ricci and scalar curvatures may seem mysterious and unmotivated at first. But a little representation theory makes their meaning more clear.

We think of the curvature R as a section of $\otimes^4 T^*M$ by the formula $R(X, Y, Z, W) := \langle R(X, Y)Z, W\rangle$. For each point p, the automorphism group of T_pM is isomorphic to the orthogonal group $\mathrm{O}(n)$, and in fact T_pM and T_p^*M are isomorphic as $\mathrm{O}(n)$-modules, and both isomorphic to the *standard* representation, which we denote E for brevity, so that $R \in \otimes^4 E$. However, the symmetries of R mean that it is actually contained in $S^2\Lambda^2 E$. Furthermore the Bianchi identity shows that R is in the kernel of the $\mathrm{O}(n)$-equivariant map $b : S^2\Lambda^2 E \to S^2\Lambda^2 E$ defined by the formula

$$b(T)(X, Y, Z, W) = \frac{1}{3}\left(T(X, Y, Z, W) + T(Y, Z, X, W) + T(Z, X, Y, W)\right)$$

It can be shown that $\operatorname{Im} b = \Lambda^4 E$, and we obtain the decomposition $S^2\Lambda^2 E = \operatorname{Ker} b \oplus \operatorname{Im} b$ and we see that $R \in \operatorname{Ker} b$. Now, $\Lambda^4 E$ is irreducible as an $\mathrm{O}(n)$-module, but $\operatorname{Ker} b$ is not (at least for $n > 2$). In fact, there is a contraction $S^2\Lambda^2 E \to S^2 E$ obtained by taking a trace over the second and fourth indices, and we see that $\mathrm{Ric} \in S^2 E$ is obtained by contracting R. Finally, $S^2 E$ is not irreducible as an $\mathrm{O}(n)$ module, since it contains an $\mathrm{O}(n)$-invariant vector, namely the invariant inner product on E (and its scalar multiples). This writes $S^2 E = S^2_0 E \oplus \mathbb{R}$ where the trace $S^2 E \to \mathbb{R}$ takes Ric to s, and Ric_0 is the part in $S^2_0 E$. The part of R in the kernel of the contraction $S^2\Lambda^2 E \to S^2 E$ is called the *Weyl* curvature tensor, and is denoted W. For $n \neq 4$ these factors are all irreducible (for $n < 4$ some of them vanish). But for $n = 4$ there is a further decomposition of W into "self dual" and "anti-self dual" parts coming from the exceptional isomorphism $\mathfrak{o}(4) = \mathfrak{o}(3) \oplus \mathfrak{o}(3)$.

5.3. **Sectional curvature.** The Riemannian metric on M induces a (positive-definite) symmetric inner product on the fibers of $\Lambda^p(TM)$ for every p. For $p = 2$ we have a formula

$$\|X \wedge Y\|^2 = \langle X, X\rangle\langle Y, Y\rangle - \langle X, Y\rangle^2$$

Geometrically, $\|X \wedge Y\|$ is the area of the parallelogram spanned by X and Y in T_pM.

As observed above, the curvature R also induces a symmetric inner product on each $\Lambda^2(T_pM)$, and the ratio of the two inner products is a well-defined function on the space of rays $\mathbb{P}(\Lambda^2 T_pM)$. This leads to the following definition:

Definition 5.6 (Sectional curvature)**.** Let σ be a 2-dimensional subspace of T_pM, and let X, Y be a basis for σ. The *sectional curvature* of σ, denoted $K(\sigma)$, is the ratio

$$K(\sigma) := \frac{\langle R(X,Y)Y, X\rangle}{\|X \wedge Y\|^2}$$

Note that since both R and $\|\cdot\|$ are symmetric inner products on $\Lambda^2 T_pM$, the definition is independent of the choice of basis. Since a symmetric inner product on a vector space can be recovered from the length function it induces on vectors, it follows that the full tensor R can be recovered from its ratio with the Riemannian inner product as a function on $\mathbb{P}(\Lambda^2 T_pM)$. Since the Grassmannian of 2-planes in V is an irreducible subvariety of $\mathbb{P}(\Lambda^2 V)$ it follows that the full tensor R can be recovered from the values of the sectional curvature on all 2-planes in TM.

Remark 5.7. It might seem more natural to consider $\langle R(X,Y)X, Y\rangle$ instead in the definition of K, but this would give sectional curvature the "wrong" sign. One justification for the sign ultimately comes from the Gauss-Bonnet formula, which relates the sign of the average sectional curvature to the sign of the Euler characteristic (for a closed, oriented surface).

Authors that use the opposite definition of R (see Remark 5.2) will in fact use an expression of the form $\langle R(X,Y)X, Y\rangle$ in their formula for sectional curvature, so that the meaning of "positive sectional curvature" is unambiguous.

Let N be a smooth submanifold of M. It is instructive to compare the sectional curvature of a 2-plane σ contained in T_pN in N and in M. Choose vector fields X and Y in $\mathfrak{X}(N)$, and for convenience let's suppose $[X,Y] = 0$. Evidently $\|X \wedge Y\|^2 = \|X\|^2\|Y\|^2$ is the same whether computed in M or in N. We compute

$$K_M(\sigma)\cdot\|X\wedge Y\|^2 = K_N(\sigma)\cdot\|X\wedge Y\|^2 + \langle\nabla_X\nabla_Y^\perp Y - \nabla_Y\nabla_X^\perp Y, X\rangle$$

On the other hand, $\langle\nabla_X^\perp Y, Z\rangle = 0$ for any $X, Y, Z \in \mathfrak{X}(N)$ and therefore $\langle\nabla_X\nabla_Y^\perp Y, X\rangle = -\langle\nabla_Y^\perp Y, \nabla_X^\perp X\rangle$ (and similarly for the other term). Using the symmetry of the second fundamental form, we obtain the so-called *Gauss equation*:

$$K_N(\sigma)\cdot\|X\wedge Y\|^2 = K_M(\sigma)\cdot\|X\wedge Y\|^2 + \langle\mathrm{I\!I}(X,X),\mathrm{I\!I}(Y,Y)\rangle - \|\mathrm{I\!I}(X,Y)\|^2$$

In the special case that N is codimension one and co-orientable, the normal bundle νN may be identified with the trivial line bundle $\mathbb{R}\times N$ over N, and $\mathrm{I\!I}$ may be thought of as an ordinary symmetric inner product on N. Using the metric inner product on N, we may express $\mathrm{I\!I}$ as a symmetric matrix, by the formula $\mathrm{I\!I}(X,Y) = \langle\mathrm{I\!I}(X), Y\rangle$.

Definition 5.8 (Mean curvature)**.** Let N be a codimension one co-orientable submanifold of M. If we express $\mathrm{I\!I}$ as a symmetric matrix by using the metric inner product, the eigenvalues of $\mathrm{I\!I}$ are the *principal curvatures*, the eigenvectors of $\mathrm{I\!I}$ are the *directions of principal curvature*, and the average of the eigenvectors (i.e. $1/\dim(N)$ times the trace) is the *mean curvature*, and is denoted H.

For a surface S in $\mathbb{E}^3$, the sectional curvature can be derived in a straightforward way from the geometry of the *Gauss map*.

Definition 5.9 (Gauss map)**.** Let N be a codimension 1 co-oriented smooth submanifold of $\mathbb{E}^n$. The *Gauss map* is the smooth map $g: N \to S^{n-1}$, the unit sphere in $\mathbb{E}^n$, determined uniquely by the property that the oriented tangent space T_pN and $T_{g(p)}S^{n-1}$ are parallel for each $p \in N$.

Another way to think of the Gauss map is in terms of unit normals. If N is codimension 1 and co-oriented, the normal bundle νN is canonically identified with $\mathbb{R}\times N$ and has a section whose value at every point is the positive unit normal. On the other hand, νN is a subbundle of $T\mathbb{E}^n|N$, and the fiber at every point is canonically identified with a line through the origin in $\mathbb{E}^n$. So the unit normal section σ can be thought of as taking values in the unit sphere; the map taking a point on N to its unit normal (in S^{n-1}) *is* the Gauss map, so by abuse of notation we can write $\sigma = g$ (in Euclidean coordinates).

The Gauss map is related to the second fundamental form as follows:

Lemma 5.10. *For vectors $u, v \in T_pN$ we have $\mathrm{I\!I}(u,v) = -\langle dg(u), v\rangle$.*

Proof. Extend u, v to vector fields U, V on N near p. Then $\langle\sigma, V\rangle = 0$ where σ is the unit normal field, so

$$\langle\nabla_U\sigma, V\rangle + \langle\sigma, \nabla_U V\rangle = 0$$

Now, $\langle\sigma, \nabla_U V\rangle = \nabla_U^\perp V = \mathrm{I\!I}(U,V)$ after identifying νN with $\mathbb{R}\times N$. Furthermore, $\nabla_U\sigma = d\sigma(U) = dg(U)$, and the lemma is proved. □

Corollary 5.11. *For a smooth surface S in $\mathbb{E}^3$ the form $K\cdot d\mathrm{area} = g^*d\mathrm{area}$; i.e. the pullback of the area form on S^2 under g^* is K times the area form on S, where K is the sectional curvature (thought of as a function on S).*

Proof. At each point $p \in S$ we can choose an orthonormal basis e_1, e_2 for T_pS which are eigenvectors for II. If the eigenvalues (i.e. the principal curvatures) are k_1, k_2 then $dg(e_i) = -k_i e_i$ and therefore the Gauss equation implies that $K_S = k_1 k_2$ at each point. But this is the determinant of dg (thought of as a map from T_pS to $T_{g(p)}S^2 = T_pS$). □

At a point where S is convex, the principal curvatures both have the same sign, and S is positively curved. At a saddle point, the principal curvatures have opposite signs, and S is negatively curved.

5.4. **The Gauss-Bonnet Theorem.** If S is an oriented surface and $\gamma : [0,1] \to S$ is a smooth curve, we can think of the image of γ (locally) as a smooth submanifold, and compute its second fundamental form.

Definition 5.12 (Geodesic curvature)**.** Let γ be a smooth curve in S with positive unit normal field σ. The *geodesic curvature* of γ, denoted k_g, is defined by the formula

$$k_g = \frac{\langle \mathrm{II}(\gamma', \gamma'), \sigma\rangle}{\langle \gamma', \gamma' \rangle}$$

Hence if γ is parameterized by arclength, $|k_g| = \|\nabla_{\gamma'}\gamma'\|$.

The following theorem was first proved for Gauss for closed surfaces, and extended by Bonnet to surfaces with boundary. We give a proof for surfaces in $\mathbb{E}^3$ to emphasize the relationship of this theorem to the geometry of the Gauss map. The proof of the general case will be deferred until we discuss characteristic classes in § 7.

Theorem 5.13 (Gauss-Bonnet for surfaces in $\mathbb{E}^3$)**.** *Let S be a smooth oriented surface in $\mathbb{E}^3$ with smooth boundary ∂S. Then*

$$\int_S K d\text{area} + \int_{\partial S} k_\gamma d\text{length} = 2\pi\chi(S)$$

Proof. We first prove this theorem for a smooth embedded disk in S^2.

Let D be a smooth embedded disk in S^2 with oriented boundary γ, and suppose the north pole is in the interior of D and the south pole is in the exterior. Let (θ, ϕ) be polar coordinates on S^2, where $\phi = 0$ is the "north pole", and θ is longitude. The sectional curvature K is identically equal to 1, and therefore

$$K d\text{area} = \sin(\phi) d\phi \wedge d\theta = d(-\cos(\phi) d\theta)$$

at least away from the north pole, where $d\theta$ is well-defined. If C is a small negatively-oriented circle around the north pole, the integral of the 1-form $-\cos(\phi)d\theta$ around C is 2π. Therefore by Stokes' theorem,

$$\int_D K d\text{area} + \int_\gamma \cos(\phi) d\theta = 2\pi$$

Let's parameterize γ by arclength, so that $|\gamma'| = 1$. We would like to show that

$$\int_\gamma \cos(\phi) d\theta = \int_\gamma k_g d\text{length}$$

Geometrically, k_g measures the infinitesimal rate at which parallel transport around γ rotates relative to γ', so $\int_\gamma k_g d\text{length}$ is just the total angle through which $T_\gamma(0)S^2$ rotates after parallel transport around the loop γ. On the other hand, $\cos(\phi)d\theta(\gamma')$ is the infinitesimal rate at which parallel transport around γ rotates relative to ∂_θ. Since ∂_θ and γ' are homotopic as nonzero sections of $TS^2|_\gamma$, it follows that these integrals are the same. Thus we have proved the Gauss-Bonnet theorem

$$\int_D K d\text{area} + \int_{\partial D} k_g d\text{length} = 2\pi$$

for a smooth embedded disk in S^2.

On the other hand, we have already shown that for any smooth surface S in $\mathbb{E}^3$ the Gauss map pulls back $K d$area on S^2 to $K d$area on S. Furthermore, the pullback of the Gauss map commutes with parallel transport, so the integral of k_g along a component of ∂S is equal to the integral of k_g along the image of ∂S under the Gauss map. We can write $K d$area as exterior d applied to the pullback of $-\cos(\phi)d\theta$ away from small neighborhoods of the preimage of the north and south poles. The signed number of preimages of these points is the index of the vector field on S obtained by pulling back ∂_θ. By the Poincaré-Hopf formula, this index is equal to $\chi(S)$. Hence

$$\int_S K d\text{area} + \int_{\partial S} k_g d\text{length} = 2\pi\chi(S)$$

for any smooth compact surface in $\mathbb{E}^3$ (possibly with boundary). □

Example 5.14 (Foucault's pendulum). Even the case of a disk $D \subset S^2$ is interesting; it says that the area of the disk is equal to the total angle through which parallel transport around ∂D rotates the tangent space.

Imagine a heavy pendulum swinging back and forth over some fixed location on Earth. As the Earth spins on its axis, the pendulum *precesses*, as though being parallel transported around a circle of constant latitude. The total angle the pendulum precesses in a 24 hour period is equal to 2π minus the area enclosed by the circle of latitude (in units for which the total area of Earth is 4π). Thus at the north pole, the pendulum makes a full rotation once each day, whereas at the equator, it does not precess at all. See Figure 2.

5.5. **Jacobi fields.** To get a sense of the geometric meaning of curvature it is useful to evaluate our formulas in geodesic normal coordinates. It can then be seen that the curvature measures the second order deviation of the metric from Euclidean space.

Fix some point $p \in M$ and let v, w be vectors in T_pM. For small s consider the 1-parameter family of rays through the origin in T_pM defined by

$$\rho_s(t) = (v + sw)t$$

and observe that $\exp \circ \rho_s$ is a geodesic through p with tangent vector at zero equal to $v + sw$. We can think of this as a 2-parameter family $\Gamma : [0,1] \times (-\epsilon, \epsilon) \to M$ with $\Gamma(t,s) = \exp \circ \rho_s(t)$, and define T and V (at least locally in M) to be $d\Gamma(\partial_t)$ and $d\Gamma(\partial_s)$ respectively, thought of as vector fields along (the image of) Γ. For each fixed s the image $\Gamma : [0,1] \times s \to M$ is a radial geodesic through p, so $\nabla_T T = 0$ throughout the image. Since T and V commute, we have $[T,V] = 0$ and $\nabla_T V = \nabla_V T$ and we obtain the identity $R(T,V)T = \nabla_T \nabla_V T = \nabla_T \nabla_T V$.

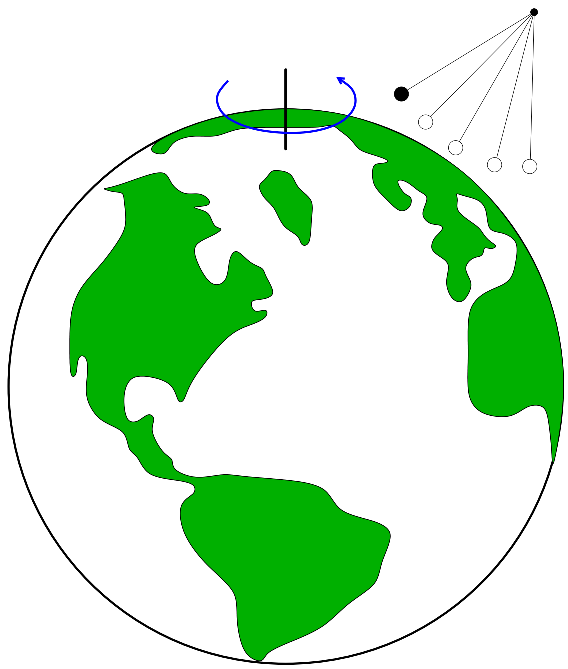


FIGURE 2. Foucault's pendulum precesses at a rate depending on the latitude.

Definition 5.15 (Jacobi equation)**.** Let V be a vector field along a geodesic γ, and let $\gamma' = T$ along γ. The *Jacobi equation* is the equation

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

for V, and a solution is called a *Jacobi field*.

If we let e_i be a parallel orthonormal frame along a geodesic γ with tangent field T, and let t parameterize γ proportional to arclength, and $V = \sum v_i e_i$, then $\nabla_T\nabla_T V = \sum_i v_i'' e_i$ while $R(T,V)T = \sum_j v_j R(T,e_j)T$ so the Jacobi equations may be expressed as a system of second order linear ODEs:

$$v_i'' = \sum_j v_j \langle R(T,e_j)T, e_i\rangle$$

and therefore there is a *unique* Jacobi field V along T with a given value of $V(0)$ and $V'(0) := \nabla_T V|_{t=0}$.

Conversely, given $V(0)$ and $V'(0)$ we choose a smooth curve $\sigma(s)$ with $\sigma'(0) = V(0)$ and extend T and $V'(0)$ to parallel vector fields along σ. Define $\Gamma : [0,1]\times(-\epsilon,\epsilon) \to M$ by $\Gamma(t,s) = \exp_{\sigma(s)}(t(T + sV'(0)))$. Then we get vector fields $U := d\Gamma(\partial_t)$ and $S := d\Gamma(\partial_s)$ such that U is tangent to the geodesics, and S gives their variation. Then $[S,U] = 0$ and $\nabla_U U = 0$ so $R(U,S)U = \nabla_U\nabla_U S$ along Γ. On the other hand, $U = T$ along γ, and $S = \sigma'$ along σ (so that $S(\sigma(0)) = V(0)$), and $\nabla_U S|_{\sigma(0)} = \nabla_S U|_{\sigma(0)} = \nabla_S(T + sV'(0))|_{|}\sigma(0) = V'(0)$. So the restriction of S to γ is the unique Jacobi field with first order part $V(0), V'(0)$. It follows that *Jacobi fields along a geodesic γ are exactly the variations of γ by geodesics*.

Consider our original one-parameter variation $\Gamma(t,s) := \exp_p((v+sw)t)$ where now we choose v and w to be orthonormal at T_pM, and let T, V be $d\Gamma(\partial_t)$ and $d\Gamma(\partial_s)$ respectively.

Note that V is a Jacobi field along $\gamma(\cdot) := \Gamma(\cdot, 0)$ with $V(0) = 0$ and $V'(0) = w$. We compute the first few terms in the Taylor series for the function $t \to \langle V(\gamma(t)), V(\gamma(t))\rangle$ at $t = 0$ (note that $V'(t) := V'(\gamma(t)) = (\nabla_T V)(\gamma(t))$ and so on).

$$\begin{aligned}
\langle V, V\rangle|_{t=0} &= 0 \\
\langle V, V\rangle'|_{t=0} &= 2\langle V, V'\rangle|_{t=0} = 0 \\
\langle V, V\rangle''|_{t=0} &= 2\langle V', V'\rangle|_{t=0} + 2\langle V'', V\rangle|_{t=0} = 2\|w\|^2 = 2 \\
\langle V, V\rangle'''|_{t=0} &= 6\langle V'', V'\rangle|_{t=0} + 2\langle V''', V\rangle|_{t=0} = 0
\end{aligned}$$

where we use the Jacobi equation to write $V'' = R(T, V)T$ which vanishes at $t = 0$ (since it is tensorial, and V vanishes at $t = 0$). On the other hand,

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

where we used the Leibniz formula for covariant derivative of the contraction of the tensor R with T, V, T, and the fact that $\nabla_T T = 0$ and $V|_{t=0} = 0$. Hence

$$\begin{aligned}
\langle V, V\rangle''''|_{t=0} &= 8\langle V''', V'\rangle|_{t=0} + 6\langle V'', V''\rangle|_{t=0} + 2\langle V'''', V\rangle|_{t=0} \\
&= 8\langle R(v, w)v, w\rangle = -8K(\sigma)
\end{aligned}$$

where σ is the 2-plane spanned by v and w. In other words,

$$\|V(t)\|^2 = t^2 - \frac{1}{3}K(\sigma)t^4 + O(t^5)$$

Thus: in 2-planes with positive sectional curvature, radial geodesic diverge *slower* than in Euclidean space, whereas in 2-planes with negative sectional curvature, radial geodesics diverge *faster* than in Euclidean space.

5.6. **Conjugate points and the Cartan-Hadamard Theorem.**

Definition 5.16 (Conjugate points)**.** Let $p \in M$, and let $v \in T_pM$. We say $q := \exp_p(v)$ is *conjugate* to p along the geodesic γ_v if $d\exp_p(v) : T_vT_pM \to T_qM$ does not have full rank.

Lemma 5.17. *Let $\gamma : [0, 1] \to M$ be a geodesic. The points $\gamma(0)$ and $\gamma(1)$ are conjugate along γ if and only if there exists a non-zero Jacobi field V along γ which vanishes at the endpoints.*

Proof. Let $w \in T_vT_pM$ be in the kernel of $d\exp_p(v)$, and by abuse of notation, use w also to denote the corresponding vector in T_pM. Define $\Gamma(s, t) := \exp_p((v+sw)t)$. Then $d\Gamma(\partial_s)$ is a Jacobi field along γ_v which vanishes at $p = \gamma_v(0)$ and $q = \gamma_v(v)$.

Conversely, suppose V is a nonzero Jacobi field along γ with $V(0) = V(1) = 0$. Then if we define $\Gamma(s, t) := \exp_{\gamma(0)}((\gamma'(0)+sV'(0))t)$, then $V = d\Gamma(\partial_s)$, and $d\exp_p(\gamma'(0))(V'(0)) = V(1) = 0$. □

It follows that the definition of conjugacy is symmetric in p and q (which is not immediate from the definition).

The Jacobi equation and Lemma 5.17 together let us use curvature to control the existence and location of conjugate points (and vice versa). One important example of this interaction is the Cartan-Hadamard Theorem:

Theorem 5.18 (Cartan-Hadamard)**.** *Let M be complete and connected, and suppose the sectional curvature satisfies $K \le 0$ everywhere. Then* exp *is nonsingular, and therefore* $\exp_p : T_pM \to M$ *is a covering map. Hence (in particular), the universal cover of M is diffeomorphic to $\mathbb{R}^n$, and $\pi_i(M) = 0$ for all $i > 1$.*

Proof. The crucial observation is that the condition $K \le 0$ implies that for V a Jacobi field along a geodesic γ, the length squared $\langle V, V\rangle$ is *convex* along γ. We compute

$$\begin{aligned}\frac{d^2}{dt^2}\langle V,V\rangle &= 2\langle V'',V\rangle + 2\langle V',V'\rangle \\ &= -2\langle R(T,V)V,T\rangle + 2\langle V',V'\rangle \\ &= -2K(\sigma)\cdot \|T\wedge V\|^2 + 2\langle V',V'\rangle \ge 0\end{aligned}$$

where σ is the 2-plane spanned by T and V. Since $\langle V,V\rangle \ge 0$, if $V(0) = 0$ but $V'(0) \ne 0$ (say), then 0 is the unique minimum of $\langle V,V\rangle$, and therefore $\gamma(0)$ is not conjugate to any other point. Hence $d\exp$ is nonsingular at every point, and $\exp_p : T_pM \to M$ is an immersion. The Riemannian metric on M pulls back to a Riemannian metric on T_pM in such a way that radial lines through the origin are geodesics. Thus, by the Hopf-Rinow Theorem (Theorem 4.8) the metric on T_pM is complete, and therefore $\exp_p$ is a covering map. □

Suppose γ is a geodesic with $\gamma' = T$, and V is a Jacobi field along γ. Then

$$\frac{d^2}{dt^2}\langle T,V\rangle = \frac{d}{dt}\langle T,V'\rangle = \langle T,V''\rangle = \langle T,R(T,V)T\rangle = 0$$

by the Jacobi equation, and the symmetries of R. Hence we obtain the formula

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

This shows that the tangential part of a Jacobi field is of the form $(at+b)\gamma'$ for some constants a,b and therefore one may as well restrict attention to normal Jacobi fields. In particular, we deduce that if a Jacobi field V vanishes at two points on γ (or more), then V and V' are everywhere perpendicular to γ.

5.7. **Second variation formula.** Let $\gamma : [a,b] \to M$ be a unit-speed geodesic, and let $\Gamma : [a,b] \times (-\epsilon,\epsilon) \times (-\delta,\delta) \to M$ be a 2-parameter variation of γ. Denote the coordinates on the three factors of the domain of Γ as t,v,w, and let $d\Gamma(\partial_t) = T$, $d\Gamma(\partial_v) = V$, $d\Gamma(\partial_w) = W$. We let $\gamma_{v,w} : [a,b] \to M$ be the restriction of Γ to the interval with constant (given) values of v and w.

Theorem 5.19 (Second variation formula)**.** *For $|v|,|w|$ small, let $\gamma_{v,w} : [a,b] \to M$ be a 2-parameter variation of a geodesic $\gamma : [a,b] \to M$. We denote $\gamma'_{v,w}$ by T, and let V and W be the vector fields tangent to the variations. Then there is a formula*

$$\begin{aligned}\frac{d^2}{dvdw}\text{length}(\gamma_{v,w})|_{v=w=0} &= \langle \nabla_W V, T\rangle|_a^b \\ &\quad + \int_a^b \langle \nabla_T V, \nabla_T W\rangle - \langle R(W,T)T,V\rangle - T\langle V,T\rangle T\langle W,T\rangle dt\end{aligned}$$

Proof. As in the derivation of the first variation formula (i.e. Theorem 4.1) we compute

$$\frac{d^2}{dvdw}\text{length}(\gamma_{v,w}) = \frac{d}{dw}\int_a^b \frac{\langle \nabla_T V, T\rangle}{\|T\|}dt$$

Differentiating under the integral, we get

$$\begin{aligned}\text{LHS} &= \int_a^b \frac{\langle \nabla_W \nabla_T V, T\rangle + \langle \nabla_T V, \nabla_W T\rangle}{\|T\|} - \frac{\langle \nabla_T V, T\rangle\langle \nabla_W T, T\rangle}{\|T\|^3}dt \\ &= \int_a^b \frac{\langle R(W,T)V, T\rangle + \langle \nabla_T \nabla_W V, T\rangle + \langle \nabla_T V, \nabla_W T\rangle}{\|T\|} - \frac{\langle \nabla_T V, T\rangle\langle \nabla_W T, T\rangle}{\|T\|^3}dt\end{aligned}$$

Evaluating this at $(0,0)$ where $\|T\| = 1$ and $\nabla_T T = 0$ we get

$$\begin{aligned}\text{LHS}|_{0,0} &= \int_a^b \langle \nabla_T V, \nabla_T W\rangle - \langle R(W,T)T, V\rangle + T\langle \nabla_W V, T\rangle - T\langle V, T\rangle T\langle W, T\rangle dt \\ &= \langle \nabla_W V, T\rangle|_a^b + \int_a^b \langle \nabla_T V, \nabla_T W\rangle - \langle R(W,T)T, V\rangle - T\langle V, T\rangle T\langle W, T\rangle dt\end{aligned}$$

as claimed. □

This formula becomes more useful if we specialize the kinds of variations we consider. Let's consider *normal* variations; i.e. those with V and W perpendicular to T along γ. Since reparameterization does not affect the length of the curve, any variation with endpoints fixed can be reparameterized to be perpendicular. If either V or W vanishes at the endpoints, the first term drops out too, and we get

$$\frac{d^2}{dvdw}\text{length}(\gamma_{v,w})|_{v=w=0} = \int_a^b \langle \nabla_T V, \nabla_T W\rangle - \langle R(W,T)T, V\rangle dt$$

Definition 5.20 (Index form)**.** Let $\mathcal{V}(\gamma)$ (or just $\mathcal{V}$ if γ is understood) denote the space of smooth vector fields along γ which are everywhere perpendicular to γ', and $\mathcal{V}_0$ the subspace of perpendicular vector fields along γ that vanish at the endpoints. The *index form* is the symmetric bilinear form I on $\mathcal{V}$ is defined by

$$I(V,W) := \int_a^b \langle \nabla_T V, \nabla_T W\rangle - \langle R(W,T)T, V\rangle dt$$

With this definition the index form is manifestly seen to be symmetric. However, it can also be re-written as follows:

$$\begin{aligned}I(V,W) &= \int_a^b T\langle \nabla_T V, W\rangle - \langle \nabla_T \nabla_T V, W\rangle - \langle R(V,T)T, W\rangle dt \\ &= \langle \nabla_T V, W\rangle|_a^b - \int_a^b \langle \nabla_T \nabla_T V - R(T,V)T, W\rangle dt\end{aligned}$$

We deduce the following corollary from the second variation formula:

Corollary 5.21. *Suppose I is positive definite on $\mathcal{V}_0$. Then γ is a unique local minimum for length among smooth curves joining p to q. More generally, the null space of I on $\mathcal{V}_0$ is exactly the set of Jacobi fields along γ which vanish at the endpoints.*

Proof. The first statement follows from the second variation formula and the definition of the index form.

If V and W vanish at the endpoints, then

$$I(V,W) = -\int_a^b \langle \nabla_T \nabla_T V - R(T,V)T, W\rangle dt$$

so V is in the null space of I (i.e. $I(V,\cdot)$ is identically zero) if and only if V is a Jacobi field. □

In particular, I has a non-trivial null space if and only if $\gamma(a)$ and $\gamma(b)$ are conjugate along γ, and the dimension of the null space of I is the dimension of the null space of $d\exp$ at the relevant point.

We have already seen (Corollary 4.7) that radial geodesics emanating from a point p are (globally!) the unique distance minimizers up to any radius r such that $\exp_p$ is a diffeomorphism when restricted to the ball of radius r. It follows (by essentially the same argument) that every geodesic is *locally* distance minimizing up to its first conjugate point. On the other hand, suppose q is conjugate to p along γ, and let r be another point on γ beyond q (so that $p = \gamma(0)$, $q = \gamma(t)$ and $r = \gamma(t')$ for some $t' > t$). Since q is conjugate to p, there is a nonzero Jacobi field V along γ which vanishes at p and q, tangent to a variation of γ (between p and q) by smooth curves γ_t which start and end at p and q, and for which $\text{length}(\gamma_t) = \text{length}(\gamma) + o(t^2)$. Since $V(q) = 0$ but V is nonzero we must have $V'(q) \neq 0$ and therefore γ_t makes a definite angle at q with γ, for small positive t. So we have a 1-parameter family of *piecewise* smooth curves from p to r, obtained by first following γ_t from p to q, and then following γ from q to r. Moreover, the length of these curves is constant to second order. But rounding the corner near q reduces the length of these curves by a term of order t^2, so we conclude that γ is *not* locally distance minimizing past its first critical point.

5.8. **Symplectic geometry of Jacobi fields.** If we fix a geodesic γ and a point p on γ, the Jacobi fields along γ admit a natural *symplectic structure*, defined by the pairing

$$\omega(U,V) := \langle U, V'\rangle - \langle U', V\rangle$$

evaluated at the point p. This form is evidently antisymmetric, and is nondegenerate in view of the identification of the space of Jacobi fields with $T_pM \times T_pM$. On the other hand, it turns out that the pairing is independent of the point p:

$$\frac{d}{dt}\omega(U,V) = \langle U, V''\rangle - \langle U'', V\rangle = \langle U, R(T,V)T\rangle - \langle R(T,U)T, V\rangle = 0$$

If we trivialize the normal bundle ν to γ as $\nu = \mathbb{R}^{n-1} \times \gamma$ by choosing a parallel orthonormal frame e_i, then the coordinates of a basis of Jacobi fields and their derivatives in terms of this frame as a function of t can be thought of as a 1-parameter family of symplectic matrices $J(t) \in \mathrm{Sp}(2n-2,\mathbb{R})$. In coordinates as a block matrix, the derivative $J'(t)$ has the form

$$J' = \begin{pmatrix} 0 & \mathrm{Id} \\ \langle R(T,e_i)T, e_j\rangle & 0 \end{pmatrix}$$

One consequence of the existence of this symplectic structure is a short proof that conjugate points along a geodesic are isolated:

Lemma 5.22 (Conjugate points are isolated)**.** *Let γ be a geodesic with initial point p. The set of points that are conjugate along γ to p is discrete.*

Proof. Suppose $q = \gamma(s)$ is a conjugate point, and let V be a Jacobi field vanishing at p and at q. Let U be any other Jacobi field vanishing at p. Then $\omega(U,V) = 0$ because both U and V vanish at p. But this implies that $\langle U(q), V'(q)\rangle = \langle U'(q), V(q)\rangle = 0$, so $U(q)$ is perpendicular to $V'(q)$. If q were not isolated as a conjugate point to p, there would be a one-parameter family of nontrivial Jacobi fields V_t with $V_t(0) = 0$ and $V_0 = V$ vanishing to first order (in t) at $\gamma(s+t)$. But if $V_t = V + tU + o(t^2)$ then

$$\frac{d}{dt}\|V_t(s+t)\||_{t=0} = \frac{1}{2}\frac{d^2}{dt^2}\langle V_t(s+t), V_t(s+t)\rangle|_{t=0} = \|V'(q)\|^2 + \|U(q)\|^2 > 0$$

where the first equality follows from L'Hôpital's rule, and the second follows from the fact (derived above) that V' and U are perpendicular at q. $\square$

The following Lemma shows that Jacobi fields are the "most efficient" variations with given boundary data, at least on geodesic segments without conjugate points.

Lemma 5.23 (Index inequality)**.** *Let γ be a geodesic from p to q with no conjugate points along it, and let W be a section of the normal bundle along γ with $W(p) = 0$. Let V be the unique Jacobi field with $V(p) = W(p) = 0$ and $V(q) = W(q)$. Then $I(V,V) \le I(W,W)$ with equality if and only if $V = W$.*

Proof. For simplicity, let $p = \gamma(0)$ and $q = \gamma(1)$. Let V_i be a basis of Jacobi fields along γ vanishing at p. Since W also vanishes at p and since there are no conjugate points along γ (so that the V_i are a basis throughout the interior of γ) we can write $W = \sum_i f_i V_i$, and $V = \sum_i f_i(1)V_i$. Then

$$\begin{aligned}
I(W,W) &= \int_0^1 \langle W', W'\rangle + \langle R(T,W)T, W\rangle dt \\
&= \int_0^1 T\langle W, W'\rangle - \langle W, W''\rangle + \langle R(T,W)T,W\rangle dt \\
&= \int_0^1 T\langle W, W'\rangle - \langle W, \sum f_i'' V_i + 2\sum f_i' V_i'\rangle dt \\
&= \int_0^1 T\langle W, W'\rangle - T\langle W, \sum f_i' V_i\rangle + \langle W', \sum f_i' V_i\rangle - \langle W, \sum f_i' V_i'\rangle dt
\end{aligned}$$

where going from line 2 to line 3 we used the Jacobi equation $\sum f_i V_i'' = \sum f_i R(T, V_i)T$. Now

$$\int_0^1 T\langle W, W' - \sum f_i' V_i\rangle dt = \langle W(1), \sum f_i V_i'(1)\rangle = \langle V(1), V'(1)\rangle = I(V, V)$$

On the other hand, $\langle V_i, V_j'\rangle = \langle V_i', V_j\rangle$ for any i, j by the symplectic identity, and the fact that the V_i all vanish at 0. So

$$\begin{aligned}\langle W', \sum f_i' V_i\rangle - \langle W, \sum f_i' V_i'\rangle &= \langle \sum f_i' V_i + \sum f_i V_i', \sum f_i' V_i\rangle - \langle W, \sum f_i' V_i'\rangle \\ &= \langle \sum f_i' V_i, \sum f_i' V_i\rangle \geq 0\end{aligned}$$

Integrating, we get $I(V, V) \leq I(W, W)$ with equality if and only if $f_i' = 0$. □

An elegant corollary of the index inequality is the following theorem of Myers, generalizing a theorem of Bonnet:

Theorem 5.24 (Myers–Bonnet). *Let M be a complete Riemannian manifold. Suppose there is a positive constant H so that $\mathrm{Ric}(v, v) \geq (n-1)H$ for all unit vectors v. Then every geodesic of length $\geq \pi/\sqrt{H}$ has conjugate points. Hence the diameter of M is at most $\pi/\sqrt{H}$, and M is compact and $\pi_1(M)$ is finite.*

Proof. Let $\gamma : [0, \ell] \to M$ be a unit-speed geodesic, and e_i an orthonormal basis of perpendicular parallel fields along γ. Define vector fields $W_i := \sin(\pi t/\ell)e_i$ along M. Then we compute

$$\begin{aligned}\sum I(W_i, W_i) &= -\sum \int_0^\ell \langle W_i, W_i'' + R(W_i, T)T\rangle dt \\ &= \int_0^\ell (\sin(\pi t/\ell))^2 \left((n-1)\pi^2/\ell^2 - \mathrm{Ric}(T, T)\right) dt\end{aligned}$$

so if $\mathrm{Ric}(T, T) > (n-1)H$ and $\ell \geq \pi/\sqrt{H}$ then $\sum I(W_i, W_i) < 0$. It follows that some W_i has $I(W_i, W_i) < 0$. If γ had no conjugate points on $[0, \ell + \epsilon]$ then by Lemma 5.23 we could find a nonzero Jacobi field V_ϵ with $V_\epsilon(0) = 0$ and $V_\epsilon(\ell+\epsilon) = W + \epsilon$ and $I(V_\epsilon, V_\epsilon) < 0$. Taking the limit as $\epsilon \to 0$ we obtain a Jacobi field V with $I(V, V) < 0$ and $V(0) = V(\ell) = 0$, which is absurd. Thus γ has a conjugate point on $[0, \ell]$.

Since geodesics fail to (even locally) minimize distance past their first conjugate points, it follows that the diameter of M is at most $\pi/\sqrt{H}$, so M is compact. Passing to the universal cover does not affect the uniform lower bound on Ric, so we deduce that the universal cover is compact too, and with the same diameter bound; hence $\pi_1(M)$ is finite. □

5.9. **Spectrum of the index form.** It is convenient to take the completion of $\mathcal{V}_0$ with respect to the pairing

$$(V, W) = \int_a^b \langle V, W\rangle dt$$

This completion is a Hilbert space H, and the operator $-\nabla_T\nabla_T \cdot + R(T, \cdot)T : H \to H$ is (at least formally, where defined) self-adjoint (this is equivalent to the symmetry of the index form). We denote the operator by $\mathcal{L}$, and call it the *stability operator*. If we choose a parallel orthonormal basis of the normal bundle ν along γ, then we can think of H as the

L^2 completion of the space of smooth functions on $[a,b]$ (vanishing at the endpoints) taking values in $\mathbb{R}^m$ (where $m = n-1$). In these coordinates, we can write $\mathcal{L} = \Delta + F$ where F is a (fixed smooth) function on $[a,b]$ with values in symmetric $m \times m$ matrices acting on $\mathbb{R}^m$ in the usual way. Note that we are using the "geometer's Laplacian" $\Delta = -\frac{d^2}{dt^2}$ which differs from the more usual (algebraic) Laplacian by a sign.

For simplicity, let's first consider the case $m = 1$, so that $\mathcal{L} = \Delta + f$ where f is just some smooth function, and let's consider the spectrum of this operator restricted to functions which vanish at the endpoints. The main theorem of Sturm-Liouville theory says that the eigenvalues λ_i of $\mathcal{L}$ are real, and can be ordered so that $\lambda_1 < \lambda_2 < \cdots$ and so that there are only *finitely many* eigenvalues in $(-\infty, s]$ for any s, and the corresponding eigenvectors ξ_i form an orthonormal basis for H. Moreover, the eigenvalues and eigenvectors (normalized to have L^2 norm equal to 1) vary continuously as a function of f and of the endpoints a, b. For fixed f, and for b sufficiently close to a, the spectrum is strictly positive. As b is increased, finitely many eigenvalues might become negative; the *index* is the number of negative eigenvalues. At a discrete set of values of b, there is a zero in the spectrum; we say that such values of b are *conjugate* to a. Evidently, at a non-conjugate value b, the index of $\mathcal{L}$ on the interval $[a,b]$ is equal to the number of points in the interior conjugate to a.

The picture is similar for general m. The eigenvalues of $\Delta + F$ are real, and there are only finitely many (counted with multiplicity) in any interval of the form $(-\infty, s]$. The eigenfunctions corresponding to different eigenvalues are orthogonal, and (suitably normalized) they form a complete orthonormal basis for H. If b is not conjugate to a along γ, the index of $\mathcal{L}$ (i.e. the number of negative eigenvalues, counted with multiplicity) is equal to the number of conjugate points (also counted with multiplicity) to a along γ in the interval $[a,b]$.

This index has another interpretation which can be stated quite simply in the language of symplectic geometry, though first we must explain the rudiments of this language. First consider $\mathbb{C}^m$ with its standard Hermitian form. The imaginary part of this form is a symplectic form ω on $\mathbb{R}^{2m} = \mathbb{C}^m$, and the Lagrangian subspaces of $\mathbb{R}^{2m}$ are exactly the totally real subspaces of $\mathbb{C}^m$. The unitary group $\mathrm{U}(m)$ acts transitively on the set of totally real subspaces of $\mathbb{C}^m$, and the stabilizer of a subspace is conjugate to $\mathrm{O}(m)$. Thus, the space of Lagrangian subspaces of $\mathbb{R}^{2m}$ can be obtained as the symmetric space $\Lambda_m := \mathrm{U}(m)/\mathrm{O}(m)$ (this space is also sometimes called the *Shilov boundary* of $\mathrm{Sp}(2m)$). The group $\mathrm{Sp}(2m)$ acts on Λ_m on the left; at the level of matrices, every symplectic matrix M has a unique (polar) factorization as $M = PQ$ where P is self-adjoint and positive definite, and Q is in $\mathrm{U}(m)$. This defines a projection $\mathrm{Sp}(2m) \to \mathrm{U}(m)$ (which is a homotopy equivalence) and thereby an orbit map $\mathrm{Sp}(2m) \to \Lambda_m$. Note that $\dim(\Lambda_m) = m(m+1)/2$.

Example 5.25. When $m = 1$ then $\mathrm{U}(1) = S^1$ and $\mathrm{O}(1) = \mathbb{Z}/2\mathbb{Z}$, so $\Lambda_1 = \mathbb{RP}^1$. When $m = 2$ then there is a map $\det^2 : \mathrm{U}(2)/\mathrm{O}(2) \to S^1$ whose fiber is $\mathrm{SU}(2)/\mathrm{SO}(2) = S^2$. So Λ_2 is an S^2 bundle over S^1 whose holonomy is the antipodal map.

Let $\pi \in \Lambda_m$ be a Lagrangian subspace of $\mathbb{R}^{2m}$. The *train* of π, denoted $\Sigma_\pi \subset \Lambda_m$, is the set of Lagrangian subspaces of $\mathbb{R}^{2m}$ which are not transverse to π. The train is a codimension one real analytic subvariety, and has a well-defined co-orientation, even though neither it nor Λ_m are orientable in general. If we fix another Lagrangian π' transverse to π, then we can decompose $\mathbb{R}^{2m} = \pi \oplus \pi'$ and for every other σ the symplectic form determines a *symmetric* quadratic form $I_{\pi,\pi'}$ on σ, as follows: if $z_i \in \sigma$ decompose as $z_i = x_i + y_i$ with $x_i \in \pi, y_i \in \pi'$ then $I_{\pi,\pi'}(z_1, z_2) = \omega(x_1, y_2)$. This is symmetric, since

$$\omega(x_1, y_2) - \omega(x_2, y_1) = \omega(x_1, y_2) + \omega(y_1, x_2) = \omega(x_1 + y_1, x_2 + y_2) = 0$$

The tangent cone to the train $T_\pi\Sigma_\pi$ decomposes $T_\pi\Lambda_m$ into chambers, corresponding to σ near π on which $I_{\pi,\pi'}$ has a particular signature; the *positive side* of $T_\pi\Sigma_\pi$ consists of those σ on which $I_{\pi,\pi'}$ is positive definite.

There is a *cone field* C on Λ_m which assigns to every $\pi \in \Lambda_m$ the positive side of $T_\pi\Sigma_\pi$. The cone field points to one side all along any train Σ_π and gives it its canonical co-orientation.

Definition 5.26 (Maslov index)**.** If $\gamma : [0, 1] \to \Lambda_m$ is a 1-parameter family of Lagrangian subspaces, the *Maslov index* of γ, denoted $\mu(\gamma)$, is the algebraic intersection number of γ with the train of $\gamma(0)$.

Now let's return to Riemannian geometry. Fix a geodesic γ and a basis of orthonormal parallel vector fields $e_1, \cdots, e_m$ orthogonal to γ, and let $\mathbb{R}^{2m}$ denote the symplectic vector space of normal Jacobi fields along γ, whose value and derivative at some basepoint $p \in \gamma$, expressed in terms of the e_i, give the $2m$ coordinates. The space of Jacobi fields vanishing at $p = \gamma(0)$ is a Lagrangian subspace π of $\mathbb{R}^{2m}$, and the coordinates of these Jacobi fields at points $\gamma(t)$ defines a 1-parameter family of Lagrangian subspaces in Λ_m (which by abuse of notation we also denote γ). If $\gamma(t)$ is not conjugate to $\gamma(0)$ along γ, the index form I is nondegenerate on L^2 normal vector fields on $\gamma([0, t])$, vanishing at the endpoints. With this notation, the index of I is just the Maslov index of the path $\gamma([0, t])$ in Λ_m. It counts the number of conjugate points to $\gamma(0)$ between $\gamma(0)$ and $\gamma(t)$, with multiplicity; this is precisely the algebraic intersection number of γ (in Λ_m) with the train Σ_π. The only nontrivial point is the fact that the sign of each intersection point is *positive*, which is a restatement of the content of Lemma 5.22 in this language. This explains why the (Maslov) index can only increase along γ, and never decrease (as might happen for an arbitrary path of Lagrangians).

Geometrically, a vector X in the Lie algebra $\mathfrak{sp}(2m, \mathbb{R})$ determines a vector field V_X on Λ_m (by differentiating the orbits of a 1-parameter group of symplectomorphisms tangent to X). The vector field V_X points to the positive side of the train Σ_π at a Lagrangian σ if $\omega(Xv, v) > 0$ for all nonzero v in $\sigma \cap \pi$ (here we think of X acting on $\mathbb{R}^{2m}$ by ordinary matrix multiplication). The set of X in $\mathfrak{sp}(2m, \mathbb{R})$ such that $\omega(Xv, v) \geq 0$ for *all* $v \in \mathbb{R}^{2m}$ is a *conjugation invariant cone*. In particular, a matrix of the form $X_F := \left(\begin{smallmatrix} 0 & \mathrm{Id} \\ F & 0 \end{smallmatrix}\right)$ is in this cone if and only if F is negative definite. If $X(t)$ is in the positive cone for all t, a path in Λ_m obtained by integrating the vector field $V_{X(t)}$ has only positive intersections with every train. This has the following consequence, first observed by Arnol'd [1], generalizing the classical Sturm comparison theorem.

Theorem 5.27 (Comparison Theorem)**.** *Let $F(t)$ and $G(t)$ be symmetric $m \times m$ matrices for each t, and suppose for each t that $F(t) \leq G(t)$ (in the sense that $F(t) - G(t)$ has all eigenvalues nonpositive). Then the index of $\Delta + F$ is at least as large as the index of $\Delta + G$ on any interval.*

Proof. Starting at some fixed $\pi \in \Lambda_m$, the matrices $F(t)$ and $G(t)$ determine one-parameter families of Lagrangian subspaces $\gamma(t)$ and $\delta(t)$ as integral curves of the (time-dependent) vector fields V_F and V_G on Λ_m respectively. By hypothesis, for each t the elements $W(t) := X_F(t) - X_G(t)$ of the Lie algebra are contained in the positive cone. The endpoint $\gamma(1)$ is obtained from $\delta(1)$ by integrating the time dependent vector field corresponding to the path $S(t)W(t)S(t)^{-1}$ in $\mathfrak{sp}(2m, \mathbb{R})$, where $S(t)$ is a suitable 1-parameter family of symplectic matrices. Since the positive cone is conjugation invariant, it follows that γ is homotopic rel. endpoints to a path obtained from δ by concatenating it with a path tangent to the cone field; such a path can only have non-negative intersection with any train, so the index of γ is at least as big as that of δ. $\square$

6. Lie groups and homogeneous spaces

6.1. **Abstract Lie Groups.** We have met several concrete (matrix) Lie groups already; in fact, it would be hard to develop much of the theory of smooth manifolds (let alone Riemannian manifolds) without at least an implicit understanding of certain matrix Lie groups. On the other hand, Lie groups themselves (and spaces obtained from them) are among the most important examples of manifolds.

Definition 6.1 (Lie groups and algebras)**.** Let G admit both the structure of a group and a smooth manifold. G is a *Lie group* if the multiplication map $G \times G \to G$ and the inverse map $G \to G$ are smooth. The *Lie algebra* $\mathfrak{g}$ is the tangent space to G at the origin.

Example 6.2 (Myers–Steenrod). If M is any Riemannian manifold, Myers–Steenrod [6] showed that the group of isometries $\text{Isom}(M)$ is a Lie group. One way to see this is to observe (e.g. by using the exponential map) that if M is connected, and ξ is any orthonormal frame at any point $p \in M$, an isometry of M is determined by the image of ξ. So if we fix ξ, we can identify $\text{Isom}(M)$ with a subset of the frame bundle of M, and see that this gives it the structure of a smooth manifold.

We often denote the identity element of a Lie group by $e \in G$, so that $\mathfrak{g} = T_eG$.

For every $g \in G$ there are diffeomorphisms $L_g : G \to G$ and $R_g : G \to G$ called (respectively) *left* and *right* multiplication, defined by $L_g(h) = gh$ and $R_g(h) = hg$ for $h \in G$. Note that $L_g^{-1} = L_{g^{-1}}$ and $R_g^{-1} = R_{g^{-1}}$. The maps $g \to L_g$ and $g \to R_{g^{-1}}$ are homomorphisms from G to $\text{Diff}(G)$.

A vector field X on G is said to be *left invariant* if $dL_g(X) = X$ for all $g \in G$. Since G acts transitively on itself with trivial stabilizer, the left invariant vector fields are in bijection with elements of the Lie algebra, where $X(e) \in \mathfrak{g} = T_eG$ determines a left-invariant vector field X by $X(g) = dL_gX(e)$ for all g, and conversely a left-invariant vector field restricts to a vector in T_eG. So we may (and frequently do) identify $\mathfrak{g}$ with the space of left-invariant vector fields on G. If X and Y are left-invariant vector fields, then so is $[X, Y]$, since for *any* smooth map ϕ between manifolds, $d\phi([X, Y]) = [d\phi(X), d\phi(Y)]$. Thus Lie bracket of vector fields on G induces a Lie bracket on $\mathfrak{g}$, satisfying the properties in Definition 1.10 (in particular, it satisfies the Jacobi identity).

Definition 6.3 (1-parameter subgroup)**.** A smooth map $\gamma : \mathbb{R} \to G$ is a *1-parameter subgroup* if it is a homomorphism; i.e. if $\gamma(s + t) = \gamma(s)\gamma(t)$ for all $s, t \in \mathbb{R}$.

Suppose $\gamma : \mathbb{R} \to G$ is a 1-parameter subgroup, and let $X(e) = \gamma'(0) \in T_eG = \mathfrak{g}$. Let X be the corresponding left-invariant vector field on G. Differentiating the defining equation of a 1-parameter subgroup with respect to t at $t = 0$, we get

$$\gamma'(s) = d\gamma(s)(X(e)) = X(\gamma(s))$$

so that γ is obtained as the integral curve through e of the left-invariant vector field X. Conversely, if X is a left-invariant vector field, and γ is an integral curve of X through the origin, then γ is a 1-parameter subgroup. Thus we see that every $X(e) \in \mathfrak{g}$ arises as the tangent vector at e to a *unique* 1-parameter subgroup. Note that left multiplication permutes the integral curves of any left-invariant vector field; thus every left-invariant vector field X on G is complete (i.e. an integral curve through any point can be extended to $(-\infty, \infty)$).

A vector field X on any manifold M determines a 1-parameter family of (at least locally defined) diffeomorphisms $\phi_t : M \to M$ by $\phi_t'(p) = X(\phi_t(p))$. Formally, we can let e^X denote the infinite series

$$e^X := \mathrm{Id} + X + \frac{X^2}{2!} + \frac{X^3}{3!} + \cdots$$

which we interpret (where it converges) as a differential operator on the smooth functions on M. Let f be a smooth function on M, and if we pick some point p we can write $f(t) = f(\phi_t(p))$. If M, f and X are real analytic, we can express $f(t)$ as a power series in t with a positive radius of convergence, and observe that with notation, $X^n f(t) = f^{(n)}(t)$. Thus Taylor's theorem gives rise to the identity

$$e^X f(p) = f(\phi_1(p))$$

at least on the domain of definition of ϕ_1. This formula suggests an abuse of notation, using the expression e^X to denote the (partially defined) diffeomorphism $\phi_1 \in \mathrm{Diff}(M)$.

If $X \in \mathfrak{g}$ and $\gamma : \mathbb{R} \to G$ is the unique 1-parameter subgroup with $\gamma'(0) = X$, then $\phi_t(g) = g\gamma(t)$ for any t and any g, so we suggestively write $e^X = \gamma(1)$, and think of "exponentiation" as defining a map $\mathfrak{g} \to G$ (note that the integral curves of the left invariant vector field X are obtained from a singular integral curve $\gamma(t)$ by left translation by elements $g \in G$). Note that the derivative of this map at 0 is the identity map $\mathfrak{g} \to \mathfrak{g}$, and therefore exponentiation is a diffeomorphism from some neighborhood of 0 in $\mathfrak{g}$ to some neighborhood of e in G, although it is not typically globally surjective.

Remark 6.4. If G is given a Riemannian metric, then there is an exponential map $\exp_e : \mathfrak{g} \to G$ in the usual sense of Riemannian geometry. As we shall see, this is closely related to exponentiation (as defined above), but the two maps are different in general, and we use different notation $\exp(X)$ and e^X to distinguish the two maps.

Suppose G acts smoothly on a manifold M; i.e. there is given a smooth map $G \times M \to M$ so that for any $g, h \in G$ and $p \in M$ we have $g(h(p)) = (gh)(p)$. We can think of an action as a homomorphism $\rho : G \to \mathrm{Diff}(M)$. For each $X \in \mathfrak{g}$ and associated 1-parameter subgroup $\gamma : \mathbb{R} \to G$ with $\gamma'(0) = X$ we get a 1-parameter family of diffeomorphisms $\phi_t := \rho \circ \gamma(t)$ on M. Define $d\rho(X) \in \mathfrak{X}(M)$ to be the vector field tangent to ϕ_t; i.e. $d\rho(X) := \frac{d}{dt}\phi_t|_{t=0}$. Then $d\rho([X, Y]) = [d\rho(X), d\rho(Y)]$. Said another way, the map $d\rho : \mathfrak{g} \to \mathfrak{X}(M)$ is a homomorphism of Lie algebras. By exponentiating, we get the identity $\rho(e^X) = e^{d\rho(X)}$.

Exponentiation satisfies the formula $e^{sX} = \gamma(s)$ so that $e^{sX}e^{tX} = e^{(s+t)X}$ for any $s, t \in \mathbb{R}$. Moreover, if $[X, Y] = 0$ then e^{sX} and e^{tY} commute for any s and t, by Frobenius' theorem, and $e^{X+Y} = e^X e^Y = e^Y e^X$ in this case. We have already observed that exponentiation defines a diffeomorphism from a neighborhood of 0 in $\mathfrak{g}$ to a neighborhood of e in G; we denote the inverse by log. Formally, if we use the power series for log to define

$$\log(g) = (g-e) - \frac{1}{2}(g-e)^2 + \frac{1}{3}(g-e)^3 + \cdots + \frac{(-1)^m}{m}(g-e)^m + \cdots$$

and make it operate on functions f on G by $(gf)(h) = f(hg)$ and extend by linearity and take limits, then $\log(e^X) = X$ (as operators on smooth functions) for X in a small enough neighborhood of 0 in $\mathfrak{g}$. We can therefore write

$$\log(e^X e^Y) = \sum_k \frac{(-1)^{k-1}}{k} \sum_h \frac{X^{p_1}Y^{q_1}\cdots X^{p_h}Y^{q_h}}{p_1!q_1!\cdots p_h!q_h!}$$

where the sum is taken over all expessions with $p_i + q_i > 0$ for each i. It turns out that the terms with the $p_1 + \cdots + p_k + q_1 + \cdots + q_k = n$ can be grouped together as a formal rational sum of *Lie polynomials* in X and Y; i.e. expressions using only the Lie bracket operation (rather than composition). Thus one obtains the following:

Theorem 6.5 (Campbell-Baker-Hausdorff Formula). *For $X, Y \in \mathfrak{g}$ sufficiently close to 0, if we define $e^X e^Y = e^Z$ then there is a convergent series expansion for Z:*

$$Z = X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] - \frac{1}{24}[Y,[X,[X,Y]]] - \cdots$$

An explicit closed formula for the terms involving n-fold brackets was obtained by Dynkin. Note that if $\mathfrak{g}$ is a *nilpotent* Lie algebra — i.e. if there is a uniform n for which any n-fold bracket vanishes — then the CBH formula becomes a *polynomial*, which converges everywhere. The CBH formula shows that the group operation of a Lie group can be reconstructed (at least on a neighborhood of the identity) from its Lie algebra.

Definition 6.6 (Adjoint action). The group G acts on itself by conjugation; i.e. there is a map $G \to \mathrm{Aut}(G)$ sending $g \to L_g \circ R_{g^{-1}}$. Conjugation fixes e. The *adjoint action* of G on $\mathfrak{g}$ is the derivative of the conjugation automorphism at e; i.e. the map $\mathrm{Ad} : G \to \mathrm{Aut}(\mathfrak{g})$ defined by $\mathrm{Ad}(g)(Y) = d(L_g \circ R_{g^{-1}})Y$.

The *adjoint action* of $\mathfrak{g}$ on $\mathfrak{g}$ is the map $\mathrm{ad} : \mathfrak{g} \to \mathrm{End}(\mathfrak{g})$ defined by $\mathrm{ad}(X)(Y) = \frac{d}{dt}\mathrm{Ad}(e^{tX})(Y)|_{t=0}$.

If we think of $\mathfrak{g}$ as a smooth manifold, the adjoint action is a homomorphism $\mathrm{Ad} : G \to \mathrm{Diff}(\mathfrak{g})$ and its derivative is a homomorphism of Lie algebras $\mathrm{ad} : \mathfrak{g} \to \mathfrak{X}(\mathfrak{g})$. Thus we obtain the identity $e^{\mathrm{ad}(X)} = \mathrm{Ad}(e^X)$. Since all maps and manifolds under consideration are real analytic, this identity makes sense when interpreted as power series expansions of operators.

Let $X \in \mathfrak{g}$ denote both a vector in T_eG and the corresponding left-invariant vector field. Let ϕ_t denote the flow associated to X, so that $\phi_t(g) = ge^{tX}$. For any other vector field Y on G (left invariant or not) recall that the Lie derivative is defined by the formula

$$(\mathcal{L}_X(Y))(p) := \lim_{t\to 0} \frac{d\phi_{-t}(Y(\phi_t(p))) - Y(p)}{t} = [X,Y]$$

Now suppose Y is left-invariant. If we write $Y(e) = \frac{d}{ds}e^{sY}|_{s=0}$ then we get the formula

$$(\mathcal{L}_X(Y))(e) = \frac{d}{dt}d\phi_{-t}(Y(\phi_t(e)))|_{t=0} = \frac{\partial}{\partial t}\frac{\partial}{\partial s}\phi_{-t}(e^{tX}e^{sY})|_{t=0,s=0} = \frac{\partial}{\partial t}\frac{\partial}{\partial s}e^{tX}e^{sY}e^{-tX}|_{t=0,s=0}$$

where we use the fact that $\phi_t(e) = e^{tX}$, and the fact that the left-invariant vector field Y at e^{tX} is tangent to $e^{tX}e^{sY}$. Thus we see from the definitions that $\mathrm{ad}(X)(Y) = [X, Y]$.

This is consistent with expanding the first few terms of the power series to derive the following estimate:

$$e^{tX}Ye^{-tX} = Y + [tX, Y] + O(t^2)$$

If we write similarly $\mathrm{ad}(X)^n(Y) = [X, [X, [\cdots, [X, Y]\cdots]]]$ then the identity $e^{\mathrm{ad}(X)} = \mathrm{Ad}(e^X)$ becomes the formula

$$e^XYe^{-X} = \sum_{n=0}^{\infty}\frac{\mathrm{ad}(X)^n(Y)}{n!}$$

which can be seen directly by computing

$$e^XYe^{-X} = \sum_{n=0}^{\infty}\sum_{m=0}^{n}\frac{(-1)^{n-m}X^mYX^{n-m}}{m!(n-m)!}$$

and using the formula $\mathrm{ad}(X)^n(Y) = \sum_{m=0}^n \binom{n}{m}(-1)^{n-m}X^mYX^{n-m}$ which can be proved by induction.

Note that the Jacobi identity for Lie bracket on $\mathfrak{g}$ is equivalent to the fact that $\mathfrak{g}$ acts on $\mathfrak{g}$ by derivations under the adjoint representation; i.e.

$$\mathrm{ad}(X)([Y, Z]) = [\mathrm{ad}(X)(Y), Z] + [Y, \mathrm{ad}(X)(Z)]$$

Proposition 6.7 (Singularities of exponentiation)**.** *For any $X, Y \in \mathfrak{g}$, there is a formula*

$$e^{X+tY}e^{-X} = e^{t\frac{e^{\mathrm{ad}(X)}-\mathrm{Id}}{\mathrm{ad}(X)}Y+O(t^2)}$$

It follows that the map $\mathfrak{g} \to G$ sending $X \to e^X$ is singular at X if and only if $\mathrm{ad}(X)$ has an eigenvalue of the form $2m\pi i$ for some nonzero integer m.

Here the expression $\frac{e^{\mathrm{ad}(X)}-\mathrm{Id}}{\mathrm{ad}(X)}$ means the formal power series $\sum_{n=0}^{\infty}\mathrm{ad}(X)^n/(n+1)!$

Proof. Let's define for each s the vector $B(s) \in \mathfrak{g} = T_eG$ by

$$B(s) := \frac{d}{dt}e^{s(X+tY)}e^{-sX}|_{t=0}$$

so that we are interested in computing $B(1)$. We claim $B'(s) = \mathrm{Ad}(e^{sX})Y$. To see this, compute

$$\begin{aligned} B(s+\Delta s) - B(s) &= \frac{d}{dt}\left(e^{sX}e^{-s(X+tY)}e^{(s+\Delta s)(X+tY)}e^{-(s+\Delta s)X}\right)|_{t=0} + O((\Delta s)^2) \\ &= \frac{d}{dt}\left(e^{sX}e^{(\Delta s)(X+tY)}e^{-(s+\Delta s)X}\right)|_{t=0} + O((\Delta s)^2) \end{aligned}$$

Therefore

$$B'(s) = \mathrm{Ad}(e^{sX})\lim_{\Delta s\to 0}\frac{\frac{d}{dt}e^{(\Delta s)(X+tY)}|_{t=0}}{\Delta s} = \mathrm{Ad}(e^{sX})(Y)$$

Writing $\mathrm{Ad}(e^{sX}) = e^{s\,\mathrm{ad}(X)}$ and integrating we obtain

$$\int_0^1 \mathrm{Ad}(e^{sX})ds = \frac{e^{\mathrm{ad}(X)} - \mathrm{Id}}{\mathrm{ad}(X)}$$

and the proposition is proved. □

Example 6.8 ($\mathrm{SL}(2,\mathbb{R})$). Unlike the exponential map on complete Riemannian manifolds, exponentiation is *not* typically surjective for noncompact Lie groups. The upper half-space model of hyperbolic 2-space $\mathbb{H}^2$ consists of the subset of $z \in \mathbb{C}$ with $\mathrm{Im}(z) > 0$. The group $\mathrm{SL}(2,\mathbb{R})$ acts on $\mathbb{H}^2$ by Möbius transformations

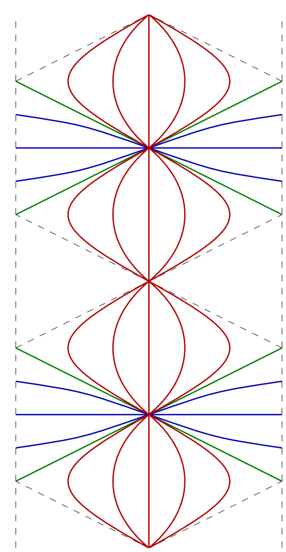
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}$$

The kernel consists of the center $\pm\mathrm{Id}$, and the image is the group $\mathrm{PSL}(2,\mathbb{R})$ which acts transitively and faithfully on $\mathbb{H}^2$ by isometries. Every 1-parameter family of isometries in $\mathbb{H}^2$ is one of three kinds:

(1) Elliptic subgroups; these are conjugate to $\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$ which fixes $i \in \mathbb{C}$ and acts on the unit tangent circle by rotation through angle 2θ;
(2) Parabolic subgroups; these are conjugate to $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ which fixes ∞ and acts by translation by t;
(3) Hyperbolic subgroups; these are conjugate to $\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$ which fixes 0 and ∞ and acts by dilation by e^{2t}.

Since $\mathrm{SL}(2,\mathbb{R})$ double-covers $\mathrm{PSL}(2,\mathbb{R})$, they have isomorphic Lie algebras, and there is a bijection between 1-parameter subgroups. In particular, any matrix in $\mathrm{SL}(2,\mathbb{R})$ with trace in $(-\infty,-2)$ is not in the image of exponentiation.

Identifying $\mathrm{PSL}(2,\mathbb{R})$ with the unit tangent bundle of $\mathbb{H}^2$ under the orbit map shows that it is diffeomorphic to an open $S^1 \times D^2$, and $\mathrm{SL}(2,\mathbb{R})$ is the connected double-cover (which is also diffeomorphic to $S^1 \times D^2$).



In the (adjacent) figure, two fundamental domains for $\mathrm{SL}(2,\mathbb{R})$ are indicated, together with the image of exponentiation. Elliptic subgroups are indicated in red, parabolic subgroups in green, and hyperbolic subgroups in blue. The dotted vertical lines are "at infinity". The white gaps are matrices with trace < -2 and the slanted dotted lines are matrices with trace -2 which are not in the image of exponentiation.

6.2. **Homogeneous spaces.** The left and right actions of G on itself induce actions on the various tensor bundles associated to G as a smooth manifold, so it makes sense to say that a volume form, or a metric (or some other structure) is left-invariant, right-invariant, or bi-invariant.

Since G acts on itself transitively with trivial point stabilizers, a left-invariant tensor field is determined by its value at e, and conversely any value of the field at e can be transported around by the G action to produce a unique left-invariant field with the given value (a similar statement holds for right-invariant tensor fields). The left and right actions commute, giving an action of $G\times G$ on itself; but now, the point stabilizers are conjugates of the (anti-)diagonal copy of G, acting by the adjoint representation. Thus: the bi-invariant tensor fields are in bijection with the tensors at e fixed by the adjoint representation.

The adjoint representation $\mathrm{Ad} : G \to \mathrm{Aut}(\mathfrak{g})$ is an example of a *linear representation*. Given a group G and a linear representation $\rho : G \to \mathrm{GL}(V)$, any natural operation on V gives rise to new representations of G. Hence there are natural actions of G on V^*, on all the tensor powers of V and V^*, on the symmetric or alternating tensor powers, and so on. Fixed vectors for the action form an invariant subspace V^G. The derivative of the linear representation determines a map $d\rho : \mathfrak{g} \to \mathfrak{gl}(V)$, and from the definition we see that V^G is in the kernel of $d\rho(\mathfrak{g})$. Conversely, for *connected* $\rho(G)$ we have the converse:

Lemma 6.9 (Fixed subspace)**.** *Let $\rho : G \to \mathrm{GL}(V)$ be a linear representation with fixed subspace V^G. If $\rho(G)$ is connected, then V^G is precisely the kernel of $d\rho(\mathfrak{g})$; i.e. $V^G = \cap_{X\in\mathfrak{g}} \ker(d\rho(X))$.*

Proof. We have already seen that $\ker(d\rho(X))$ contains V^G. Conversely, for any $v \in \ker(d\rho(X))$ and any $X \in \mathfrak{g}$, we have $\rho(e^X)v = e^{d\rho(X)}v = v$ so v is fixed by every element of a 1-parameter subgroup of $\rho(G)$. Since $\rho(G)$ is a Lie group, the 1-parameter subgroups fill out a neighborhood of the origin, so $\mathrm{fix}(v)$ is open. But an open subgroup of a connected topological group contains the connected component of the identity. Since $\rho(G)$ is connected by assumption, we are done. □

Example 6.10 (Invariant bilinear form)*.* Suppose G is a Lie group and $\rho : G \to \mathrm{GL}(V)$ has connected image. If β is a bilinear form on V (i.e. an element of $\otimes^2 V^*$) then β is invariant under $\rho(G)$ if and only if

$$(d\rho(X)\beta)(u, v) := \beta(d\rho(X)u, v) + \beta(u, d\rho(X)v) = 0$$

for all $u, v \in V$ and $X \in \mathfrak{g}$.

Suppose $\mathfrak{h}$ is a subspace of $\mathfrak{g}$ which is closed under Lie bracket; i.e. $\mathfrak{h}$ is a Lie subalgebra of $\mathfrak{g}$. Thinking of the elements of $\mathfrak{h}$ as left-invariant vector fields on G, Frobenius' theorem (i.e. Theorem 1.20) says that the distribution spanned by $\mathfrak{h}$ at each point is integrable, and tangent to a left-invariant foliation of G, of dimension equal to $\dim(\mathfrak{h})$. The leaf of this foliation through e is itself a Lie subgroup H of G (with Lie algebra $\mathfrak{h}$) when given its induced *path* topology, but it will *not* in general be closed in G.

Example 6.11 (Irrational foliation on a torus)*.* Let G be the flat torus $\mathbb{R}^2/\mathbb{Z}^2$ with addition as the group law. Then $\mathfrak{g} = \mathbb{R}^2$ with the trivial bracket. Any nonzero $v \in \mathfrak{g}$ spans a 1-dimensional Lie algebra $\mathfrak{h}$, and the integral curves of the associated left-invariant vector field on the torus are the lines of constant slope. If the slope is irrational, these lines are dense in G.

Conversely, if $H \subset G$ is a subgroup which is also an immersed submanifold, the tangent space $\mathfrak{h}$ to H at e is a Lie subalgebra of $\mathfrak{g}$. A subalgebra $\mathfrak{h}$ for which $[\mathfrak{g}, \mathfrak{h}] \subset \mathfrak{h}$ is called an *ideal*. The group H associated to $\mathfrak{h}$ by exponentiating is normal if and only if $\mathfrak{h}$ is an ideal.

If H is closed, the quotient space G/H is naturally a manifold with a transitive G action (coming from left multiplication of G on itself), and the map $G \to G/H$ has the structure of a principal H-bundle coming from the H action. Since $\mathfrak{h}$ is a Lie subalgebra, $\mathrm{ad}(v)$ preserves $\mathfrak{h}$ for all $v \in \mathfrak{h}$, and similarly $\mathrm{Ad}(h)$ preserves $\mathfrak{h}$ for all $h \in H$. So the adjoint representation $\mathrm{Ad} : H \to \mathrm{Aut}(\mathfrak{g})$ descends to $\mathrm{Ad} : H \to \mathrm{Aut}(\mathfrak{g}/\mathfrak{h})$, where $\mathfrak{g}/\mathfrak{h}$ is the tangent space at the image of e to G/H.

A smooth manifold M admitting a transitive (smooth) G action for some Lie group G is said to be a *homogeneous space* for G. If we pick a basepoint $p \in M$ the orbit map $G \to M$ sending $g \to g(p)$ is a fibration of G over M with fibers the conjugates of the point stabilizers, which are closed subgroups H. Thus we see that homogeneous spaces for G are simply spaces of the form G/H for closed Lie subgroups H of G.

An action of G on a homogeneous space $M = G/H$ is *effective* if the map $G \to \mathrm{Diff}(G/H)$ has trivial kernel. It is immediate from the definition that the kernel is precisely equal to the normal subgroup $H_0 := \cap_g gHg^{-1}$, which may be characterized as the biggest normal subgroup of G contained in H. If H_0 is nontrivial, then we may define $G' = G/H_0$ and $H' = H/H_0$, and then $G/H = G'/H'$ is a homogeneous space for G', and the action of G on G/H factors through an action of G'. Thus when considering homogeneous spaces one may always restrict attention to homogeneous spaces with effective actions.

Proposition 6.12 (Invariant metrics on homogeneous spaces)**.** *Let G be a Lie group and let H be a closed Lie subgroup with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$ respectively.*

(1) The G-invariant tensors on the homogeneous space G/H are naturally isomorphic with the $\mathrm{Ad}(H)$ *invariant tensors on $\mathfrak{g}/\mathfrak{h}$.*
(2) Suppose G acts effectively on G/H. Then G/H admits a G-invariant metric if and only if the closure of $\mathrm{Ad}(H)$ *in* $\mathrm{Aut}(\mathfrak{g})$ *is compact.*
(3) If G/H admits a G-invariant metric, and G acts effectively on G/H, then G admits a left-invariant metric which is also right-invariant under H, and its restriction to H is bi-invariant.
(4) If G is compact, then G admits a bi-invariant metric.

Proof. (1): Any G-invariant tensor on G/H may be restricted to $T_H G/H = \mathfrak{g}/\mathfrak{h}$ whose stabilizer is H acting by a suitable representation of $\mathrm{Ad}(H)$. Conversely, any $\mathrm{Ad}(H)$-invariant tensor on $\mathfrak{g}/\mathfrak{h}$ can be transported around G/H by the left G action by choosing coset representatives.

(2): If G acts effectively on G/H, then for any left-invariant metric on G the group G embeds into the isometry group G^* and H embeds into the the isotropy group H^*, the subgroup of G^* fixing the basepoint $H \in G/H$. By Myers-Steenrod (see Example 6.2) G^* and H^* are Lie groups, with Lie algebras $\mathfrak{g}^*$ and $\mathfrak{h}^*$, and since G is effective, the natural maps $\mathfrak{g} \to \mathfrak{g}^*$ and $\mathfrak{h} \to \mathfrak{h}^*$ are inclusions. Since H^* is a closed subgroup of an orthogonal group of some dimension, it is compact, and therefore so is its image $\mathrm{Ad}(H^*) \in \mathrm{Aut}(\mathfrak{g}^*)$.

On any compact group, a right-invariant metric gives rise to a right-invariant volume form which can be rescaled to have total volume 1. Let ω be such a right-invariant volume form on $\mathrm{Ad}(H^*)$. For any inner product $\langle \cdot, \cdot \rangle$ on $\mathfrak{g}^*$ define a new inner product $\langle \cdot, \cdot \rangle'$ by

$$\langle X, Y \rangle' := \int_{\mathrm{Ad}(H^*)} \langle \mathrm{Ad}(h)(X), \mathrm{Ad}(h)(Y) \rangle \omega(h)$$

Note that $\langle \cdot, \cdot \rangle'$ is positive definite if $\langle \cdot, \cdot \rangle$ is. Then for any $z \in H^*$ we have

$$\begin{aligned}\langle \mathrm{Ad}(z)(X), \mathrm{Ad}(z)(Y)\rangle' &= \int_{\mathrm{Ad}(H^*)} \langle \mathrm{Ad}(hz)(X), \mathrm{Ad}(hz)(Y)\rangle \omega(h) \\ &= \int_{\mathrm{Ad}(H^*)} \langle \mathrm{Ad}(h)(Z), \mathrm{Ad}(h)(Y)\rangle R^*_{z^{-1}}\omega(hz^{-1}) = \langle X, Y\rangle'\end{aligned}$$

by the right-invariance of ω. Thus $\mathrm{Ad}(H^*)$ (and therefore $\mathrm{Ad}(H)$) acts by isometries on $\mathfrak{g}^*$ for some positive definite inner-product, and therefore the restriction of $\mathrm{Ad}(H)$ preserves a positive definite inner-product on $\mathfrak{g}$. Hence $\mathrm{Ad}(H)$ is contained in the orthogonal group of this inner-product, which is compact, and therefore the closure of $\mathrm{Ad}(H)$ is compact.

Conversely, if the closure of $\mathrm{Ad}(H)$ is compact, by averaging any metric under a right-invariant volume form on $\mathrm{Ad}(H)$ as above we obtain an $\mathrm{Ad}(H)$-invariant metric on $\mathfrak{g}$. Let $\mathfrak{p}$ be the orthogonal complement $\mathfrak{p} = \mathfrak{h}^\perp$ of $\mathfrak{h}$ in this $\mathrm{Ad}(H)$-invariant metric. Then $\mathrm{Ad}(H)$ fixes $\mathfrak{p}$ and preserves its inner metric. Identifying $\mathfrak{p} = \mathfrak{g}/\mathfrak{h}$ we get an $\mathrm{Ad}(H)$-invariant metric on $\mathfrak{g}/\mathfrak{h}$ and a G-invariant metric on G/H.

(3): If G acts effectively on G/H and G/H admits a G-invariant metric, then by (2), $\mathrm{Ad}(H)$ has compact closure in $\mathrm{Aut}(\mathfrak{g})$, and preserves a positive-definite inner product on $\mathfrak{g}$. This inner product defines a left-invariant Riemannian metric on G as in (1), and its restriction to H is $\mathrm{Ad}(H)$-invariant, and is therefore bi-invariant, since the stabilizer of a point in H under the $H \times H$ action coming from left- and right- multiplication is $\mathrm{Ad}(H)$.

(4): Since G is compact, so is $\mathrm{Ad}(G)$. Thus $\mathrm{Ad}(G)$ admits a right-invariant volume form, and by averaging any positive-definite inner product on $\mathfrak{g}$ under the $\mathrm{Ad}(G)$ action (with respect to this volume form) we get an $\mathrm{Ad}(G)$-invariant metric on $\mathfrak{g}$, and a bi-invariant metric on G. □

Example 6.13 (Killing form). The *Killing form* is the 2-form β on $\mathfrak{g}$ defined by

$$\beta(X, Y) = \mathrm{tr}(\mathrm{ad}(X)\mathrm{ad}(Y))$$

Since the trace of a product is invariant under cyclic permutation of the factors, β is symmetric. Furthermore, for any Z,

$$\begin{aligned}\beta(\mathrm{ad}(Z)(X), Y) &= \mathrm{tr}(\mathrm{ad}([Z, X])\mathrm{ad}(Y)) = \mathrm{tr}([\mathrm{ad}(Z), \mathrm{ad}(X)]\mathrm{ad}(Y)) \\ &= \mathrm{tr}(\mathrm{ad}(Z)\mathrm{ad}(X)\mathrm{ad}(Y) - \mathrm{ad}(X)\mathrm{ad}(Z)\mathrm{ad}(Y)) \\ &= -\mathrm{tr}(\mathrm{ad}(X)\mathrm{ad}(Z)\mathrm{ad}(Y) - \mathrm{ad}(X)\mathrm{ad}(Y)\mathrm{ad}(Z)) \\ &= -\mathrm{tr}(\mathrm{ad}(X)[\mathrm{ad}(Z), \mathrm{ad}(Y)]) = -\mathrm{tr}(\mathrm{ad}(X), \mathrm{ad}([Z, Y])) \\ &= -\beta(X, \mathrm{ad}(Z)(Y))\end{aligned}$$

Hence by Example 6.10 the form β is $\mathrm{Ad}(G)$-invariant, and therefore determines a bi-invariant symmetric 2-form on G (which by abuse of notation we also denote β). A Lie algebra $\mathfrak{g}$ is *semisimple* if β is nondegenerate. A Lie algebra $\mathfrak{g}$ is said to be *reductive* if it admits *some* nondegenerate $\mathrm{Ad}(G)$-invariant symmetric 2-form; hence every semisimple Lie algebra is reductive by definition. Not every reductive Lie algebra is semisimple: for example, every abelian Lie algebra is reductive.

Let $\mathfrak{h}$ be the maximal subspace on which $\beta(\mathfrak{h}, \cdot) = 0$. Then $\mathfrak{h}$ is an ideal, since $\beta(\mathrm{ad}(Z)(X), Y) = -\beta(X, \mathrm{ad}(Z)(Y))$.

6.3. Formulae for left- and bi-invariant metrics.

Proposition 6.14 (Left-invariant metric). *Let $\langle\cdot,\cdot\rangle$ be a left-invariant metric on G, and let X,Y,Z,W be left-invariant vector fields corresponding to vectors in $\mathfrak{g}$. Let ∇ be the Levi-Civita connection on G associated to the metric.*

(1) $\nabla_X Y = \frac{1}{2}\left([X,Y] - \mathrm{ad}^*(X)(Y) - \mathrm{ad}^*(Y)(X)\right)$ *where* $\mathrm{ad}^*(W)$ *denotes the adjoint of the operator* $\mathrm{ad}(W)$ *with respect to the inner product, for any* $W \in \mathfrak{g}$*;*
(2) $\langle R(X,Y)Z,W\rangle = \langle \nabla_X Z, \nabla_Y W\rangle - \langle \nabla_Y Z, \nabla_X W\rangle - \langle \nabla_{[X,Y]}Z, W\rangle$*.*

Proof. We know that the Levi-Civita connection satisfies

$$\langle \nabla_X Y, Z\rangle = \frac{1}{2}\left(\langle [X,Y],Z\rangle - \langle Y,[X,Z]\rangle - \langle X,[Y,Z]\rangle\right)$$

proving (1). To prove (2) first note that left-invariance implies $X\langle \nabla_Y Z, W\rangle = 0$ for any X,Y,Z,W and therefore $\langle \nabla_X \nabla_Y Z, W\rangle = -\langle \nabla_Y Z, \nabla_X W\rangle$. The formula follows immediately from this. □

Corollary 6.15 (Bi-invariant metric). *Let $\langle\cdot,\cdot\rangle$ be a bi-invariant metric on G, and let X,Y,Z,W be left-invariant vector fields corresponding to vectors in $\mathfrak{g}$. Let ∇ be the Levi-Civita connection on G associated to the metric.*

(1) $\nabla_X Y = \frac{1}{2}[X,Y]$*;*
(2) $\langle R(X,Y)Z,W\rangle = \frac{1}{4}\left(\langle [X,W],[Y,Z]\rangle - \langle [X,Z],[Y,W]\rangle\right)$*;*
(3) $\langle R(X,Y)Y,X\rangle = \frac{1}{4}\|[X,Y]\|^2$*;*
(4) 1-parameter subgroups are geodesics; hence exponentiation agrees with the exponential map (in the metric) on $\mathfrak{g}$*.*

Proof. By bullet (1) from Proposition 6.12 we see $\mathrm{ad}^*(X) = -\mathrm{ad}(X)$ for any X and for any bi-invariant metric. Thus (1) follows from the bullet (1) from Proposition 6.14. Similarly, (2) follows from bullet (2) from Proposition 6.14, and (3) follows from (2).

If $\gamma(t) = e^{tX}$ is a 1-parameter subgroup, then $\gamma' = X$ and $\nabla_{\gamma'}\gamma' = \nabla_X X = [X,X] = 0$. This proves (4). □

Bullet (3) says that every bi-invariant metric on any Lie group is non-negatively curved, and is flat only on abelian subgroups and their translates. In fact, for any nonzero $X \in \mathfrak{g}$ we have $\mathrm{Ric}(X) > 0$ unless $\mathrm{ad}(X) = 0$; that is, unless X is in the center of $\mathfrak{g}$. For any Lie algebra $\mathfrak{g}$, we let $\mathfrak{z}$ denote the *center* of $\mathfrak{g}$; i.e.

$$\mathfrak{z} = \{X \in \mathfrak{g} \text{ such that } [X,Y] = 0 \text{ for all } Y \in \mathfrak{g}\}$$

Note that $\mathfrak{z}$ is an ideal, and therefore exponentiates to a *normal* subgroup Z of G. On the other hand, if G admits a bi-invariant metric, and I is an ideal in $\mathfrak{g}$, then $I^\perp$ is also an ideal, since for any $Y \in I$ and $Z \in I^\perp$,

$$0 = \langle [X,Y],Z\rangle = \langle Y,[X,Z]\rangle$$

so $[X,Z] \in I^\perp$ for all $X \in \mathfrak{g}$. Thus we have a natural splitting as Lie algebras $\mathfrak{g} = \mathfrak{z} \oplus \mathfrak{h}$. Note that since the metric is bi-invariant, exponentiation agrees with the exponential map and is surjective.

Now suppose G is simply-connected. The splitting $\mathfrak{g} = \mathfrak{z} \oplus \mathfrak{h}$ exponentiates to $G = Z \times H$ and we observe that G is isometric to the product of Z with H. Since $\mathfrak{z}$ is abelian, Z is

Euclidean. Since $\mathfrak{h}$ is simple and admits a bi-invariant metric, the Ricci curvature of H is bounded below by a uniform positive constant, so by Myers–Bonnet (Theorem 5.24), H is compact.

6.4. **Riemannian submersions.**

Definition 6.16 (Riemannian submersion)**.** Let $\pi : M \to N$ be a submersion; i.e. a smooth map such that $d\pi$ has rank equal to the dimension of N at each point. Let $V = \ker(d\pi)$. Suppose M and N are Riemannian manifolds, and $H = V^\perp$. Then π is a *Riemannian submersion* if $d\pi|_H$ is an isometry.

Note that $TM = V \oplus H$ as vector bundles. The bundle V is known as the *vertical* bundle, and H is the *horizontal* bundle. Note that V is integrable, by Frobenius' theorem, but H will typically not be. In fact, the failure of H to be integrable at a point p measures the "difference" between the metric on N near $\pi(p)$ and the metric on M in the direction of H near p. A precise statement of this is O'Neill's Theorem, to be proved below.

Let $\pi : M \to N$ be a Riemannian submersion. For any vector field X on N there is a unique vector field $\bar{X}$ on M contained in H such that $d\pi(\bar{X}) = X$. If $\gamma : [0,1] \to N$ is a smooth curve, and $q \in M$ is a point with $\pi(q) = \gamma(0)$ then there is a unique *horizontal lift* $\bar{\gamma} : [0,1] \to M$ with $\pi \circ \bar{\gamma} = \gamma$ and $\bar{\gamma}'$ horizontal. Informally, we can think of H as defining a connection on the "bundle" over N whose fibers are the point preimages of π.

Given a vector field Y on M we denote by Y^V and Y^H the *vertical part* and the *horizontal part* of Y; i.e. the orthogonal projection of Y to V and H respectively.

Lemma 6.17. *Let X and Y be vector fields on N with lifts $\bar{X}$ and $\bar{Y}$.*

1. *$[\bar{X},\bar{Y}]^H = \overline{[X,Y]}$; in fact, for any two vector fields $\tilde{X},\tilde{Y}$ on M with $d\pi(\tilde{X}) = X$ and $d\pi(\tilde{Y}) = Y$, we have $d\pi([\tilde{X},\tilde{Y}]) = [X,Y]$.*
2. *$[\bar{X},\bar{Y}]^V$ is tensorial in X and Y (i.e. its value at a point q only depends on the values of X and Y at $\pi(q)$).*

Proof. (1) is proved by computing how the two vector fields operate on smooth functions on N. Let f be a smooth function on N, so that $\tilde{X}(f \circ \pi) = (X(f)) \circ \pi$ and similarly for Y. Then

$$d\pi([\tilde{X},\tilde{Y}])(f) = (\tilde{X}\tilde{Y} - \tilde{Y}\tilde{X})(f \circ \pi) = ((XY - YX)f) \circ \pi = ([X,Y]f) \circ \pi$$

(2) is proved by calculation. Let T be a vertical vector field; i.e. a section of V. Then if we denote the Levi-Civita connection on M by $\bar{\nabla}$, we compute

$$\begin{aligned}\langle [\bar{X},\bar{Y}], T\rangle &= \langle \bar{\nabla}_{\bar{X}}\bar{Y} - \bar{\nabla}_{\bar{Y}}\bar{X}, T\rangle \\ &= \bar{X}\langle \bar{Y}, T\rangle - \langle \bar{Y}, \bar{\nabla}_{\bar{X}}T\rangle - \bar{Y}\langle \bar{X}, T\rangle + \langle \bar{X}, \bar{\nabla}_{\bar{Y}}T\rangle \\ &= \langle \bar{X}, \bar{\nabla}_{\bar{Y}}T\rangle - \langle \bar{Y}, \bar{\nabla}_{\bar{X}}T\rangle\end{aligned}$$

which is tensorial in X and Y. $\square$

Compare the tensoriality of $[\bar{X},\bar{Y}]^V$ to the tensoriality of the second fundamental form of a submanifold.

We now prove O'Neill's Theorem, which precisely relates the curvature of N to the curvature of M and the failure of integrability of H.

Theorem 6.18 (O'Neill)**.** *Let $\pi : M \to N$ be a Riemannian submersion. For any vector fields X and Y on N, there is a formula*

$$K(X,Y) = K(\bar{X},\bar{Y}) + \frac{3}{4}\|[\bar{X},\bar{Y}]^V\|^2$$

where the left hand side is evaluated at a point $p \in N$, and the right hand side is evaluated at any point $q \in M$ with $\pi(q) = p$.

Proof. By bullet (1) of Lemma 6.17, for any vector fields X, Y, Z on N we have $\langle[\bar{X},\bar{Y}],\bar{Z}\rangle = \langle[X,Y],Z\rangle$. Similarly, by the definition of a Riemannian submersion we have $\bar{X}\langle\bar{Y},\bar{Z}\rangle = X\langle Y,Z\rangle$ for any $X,Y,Z \in \mathfrak{X}(N)$. As in the proof of Theorem 3.13, $\langle\bar{\nabla}_{\bar{X}}\bar{Y},\bar{Z}\rangle$ can be expressed as a linear combination of expressions of the form $\bar{X}\langle\bar{Y},\bar{Z}\rangle$ and $\langle[\bar{X},\bar{Y}],\bar{Z}\rangle$ and therefore $\langle\bar{\nabla}_{\bar{X}}\bar{Y},\bar{Z}\rangle = \langle\nabla_X Y, Z\rangle$ and consequently $\bar{X}\langle\bar{\nabla}_{\bar{Y}}\bar{Z},\bar{W}\rangle = X\langle\nabla_Y Z, W\rangle$.

If T is a vertical vector field, bullet (1) of Lemma 6.17 gives $\langle[\bar{X},T],\bar{Y}\rangle = 0$. As in Theorem 3.13 we compute

$$\begin{aligned}\langle\bar{\nabla}_{\bar{X}}\bar{Y},T\rangle &= \frac{1}{2}\{\bar{X}\langle\bar{Y},T\rangle + \bar{Y}\langle\bar{X},T\rangle - T\langle\bar{X},\bar{Y}\rangle + \langle[\bar{X},\bar{Y}],T\rangle - \langle[\bar{X},T]\bar{Y}\rangle - \langle[\bar{Y},T],\bar{X}\rangle\} \\ &= \frac{1}{2}\langle[\bar{X},\bar{Y}],T\rangle\end{aligned}$$

and therefore we obtain $\bar{\nabla}_{\bar{X}}\bar{Y} = \overline{\nabla_X Y} + \frac{1}{2}[\bar{X},\bar{Y}]^V$. Similarly,

$$\langle\bar{\nabla}_T\bar{X},\bar{Y}\rangle = \langle\bar{\nabla}_{\bar{X}}T,\bar{Y}\rangle = -\langle\bar{\nabla}_{\bar{X}}\bar{Y},T\rangle = -\frac{1}{2}\langle[\bar{X},\bar{Y}]^V,T\rangle$$

Hence we compute

$$\langle\bar{\nabla}_{\bar{X}}\bar{\nabla}_{\bar{Y}}\bar{Z},\bar{W}\rangle = \langle\nabla_X\nabla_Y Z, W\rangle - \frac{1}{4}\langle[\bar{Y},\bar{Z}]^V,[\bar{X},\bar{W}]^V\rangle$$

and likewise

$$\langle\bar{\nabla}_{[\bar{X},\bar{Y}]}\bar{Z},\bar{W}\rangle = \langle\nabla_{[X,Y]}Z,W\rangle - \frac{1}{2}\langle[\bar{Z},\bar{W}]^V,[\bar{X},\bar{Y}]^V\rangle$$

Setting $\bar{Z} = \bar{Y}$ and $\bar{X} = \bar{W}$, the theorem follows. □

Now let G be a Lie group, and H a closed subgroup. There is a submersion $G \to G/H$. If G/H admits a G-invariant metric, then by bullet (3) from Proposition 6.12, G admits a left-invariant metric which is bi-invariant under H. The associated inner product on $\mathfrak{g}$ splits as $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{h}^\perp$ in an $\mathrm{Ad}(H)$ invariant way, and we see that $G \to G/H$ is a Riemannian submersion for these metrics. If we denote $\mathfrak{h}^\perp = \mathfrak{p}$ then $\mathfrak{p}$ is tangent to the horizontal distribution, and $\mathfrak{h}$ is tangent to the vertical distribution on G.

So let X, Y be orthonormal vectors in $\mathfrak{p}$ with respect to the $\mathrm{Ad}(H)$ invariant metric. From bullet (2) from Proposition 6.14, and from Theorem 6.18 we can compute

$$\begin{aligned}K(d\pi(X),d\pi(Y)) &= \|\mathrm{ad}^*(X)(Y) + \mathrm{ad}^*(Y)(X)\|^2 - \langle\mathrm{ad}^*(X)(X),\mathrm{ad}^*(Y)(Y)\rangle \\ &\quad - \frac{3}{4}\|[X,Y]_{\mathfrak{p}}\|^2 - \frac{1}{2}\langle[[X,Y],Y],X\rangle - \frac{1}{2}\langle[[Y,X],X],Y\rangle\end{aligned}$$

6.5. **Locally symmetric spaces.**

Definition 6.19 (Locally symmetric)**.** A Riemannian manifold M is *locally symmetric* if for each $p \in M$ there is some positive number r so that the map $\exp_p(v) \to \exp_p(-v)$ for $|v| \le r$ is an isometry of the ball $B_r(p)$ to itself.

7. Characteristic classes

7.1. **Moving frames.** Let E be a smooth (real or complex) bundle over M with structure group G, and let ∇ be a connection on M for which parallel transport respects the structure group.

We can choose local coordinates on E of the form $E|_{U_i} = U_i \times \mathbb{R}^n$ or $E|_{U_i} = U_i \times \mathbb{C}^n$ so that on the overlaps $U_i \cap U_j$ the transition maps between fibers take values in G acting in the given way on $\mathbb{R}^n$ or $\mathbb{C}^n$.

In a local chart, let $s_1, \cdots, s_n$ be a basis of sections of E. Then the covariant derivative can be expressed in the form

$$\nabla \sum_i f_i s_i = \sum_i df_i \otimes s_i + \sum_{i,j} f_i \omega_{ij} \otimes s_j$$

where ω is a matrix of 1-forms. If we identify G with its image in $\mathrm{Aut}(\mathbb{R}^n)$ or $\mathrm{Aut}(\mathbb{C}^n)$, then we can think of $\mathfrak{g}$ as a Lie algebra of matrices, and then we can think of ω as a 1-form with coefficients in $\mathfrak{g}$. Informally, we can write $\nabla = d + \omega$. Covariant differentiation extends to $\Omega^* \otimes \Gamma(E)$ by the Leibniz formula

$$\nabla(\alpha \otimes s) = d\alpha \otimes s + (-1)^{\deg(\alpha)} \alpha \wedge (ds + \omega \otimes s)$$

In the same local coordinates, we can express the curvature as follows. Apply one further covariant derivative to get the formula

$$\begin{aligned}\nabla(\nabla s) &= d(ds + \omega \otimes s) + \omega \wedge ds - \omega \wedge \omega \otimes s \\ &= d\omega \otimes s - \omega \wedge \omega \otimes s\end{aligned}$$

In other words, in local coordinates we can express the curvature of the connection as a 2-form Ω taking values in $\mathfrak{g}$ by the formula $\Omega := \nabla^2 = d\omega - \omega \wedge \omega$. We interpret this formula as follows: if u, v are vectors in T_pM then

$$R(u,v)s_i = \sum_j \Omega_{ij}(u,v) s_j$$

If G is a subgroup of the orthogonal group O or the unitary group U then we can choose a basis of *orthonormal* sections s_i, we can simply write

$$\langle R(u,v)s_i, s_j \rangle = \Omega_{ij}(u,v)$$

7.2. **The Gauss-Bonnet Theorem.** We use this formalism to give a short proof of the Gauss-Bonnet Theorem, valid for arbitrary surfaces (i.e. those not necessarily embedded in $\mathbb{E}^3$).

Let S be an oriented surface with a Riemannian metric. We can think of the tangent bundle TS as a $\mathbb{C}$ bundle with structure group $\mathrm{U}(1)$, by means of the exceptional isomorphism $\mathrm{U}(1) = \mathrm{SO}(2)$. If we let s be a unit length section of TS (locally) then in terms of s we can write the Levi-Civita connection as $d + \omega$ where ω is a 1-form with coefficients in the Lie algebra of U(1); i.e. in $i\mathbb{R}$. Similarly we have $\Omega = d\omega - \omega \wedge \omega = d\omega$ (because the Lie algebra of U(1) is abelian). The sectional curvature can be expressed in the form

$$K = \frac{\langle R(X,Y)s, is\rangle}{\|X \wedge Y\|}$$

and therefore we have simply $K d\text{area} = i d\omega$.

Stokes theorem gives

$$\int_S K d\text{area} = \int_{\partial S} i\omega + \lim_{\epsilon \to 0} \int_{C_i(\epsilon)} i\omega$$

where each $C_i(\epsilon)$ is a small circle of radius ϵ around one of the singularities of s. The Poincaré-Hopf formula says that if we make s tangent along the boundary, the number of singularities (counted with sign) is equal to $\chi(S)$. Each singularity contributes 2π to the integral (as can be seen by integrating ω over a small loop around a singularity). On the other hand, $\int_{\partial S} i\omega$ is the *negative* of the amount of twisting of TS under parallel transport around ∂S, relative to $s|_{\partial S}$. If we choose s tangent to ∂S this is just (the negative of) the integral of geodesic curvature on the boundary, and we obtain:

Theorem 7.1 (Gauss-Bonnet)**.** *Let S be a compact Riemannian surface (possibly with boundary). Then*

$$\int_S K d\text{area} + \int_{\partial S} k_g d\text{length} = 2\pi\chi(S)$$

7.3. **Chern classes.** Once we have seen this formalism, it is natural to consider the (real) 2-form $\frac{i}{2\pi}\Omega$ associated to any $U(1)$-bundle E. For a general G-bundle the matrix-valued 2-form Ω depends on a choice of local section, but transforming the section s by $s \to gs$ where g is a section of G, transforms Ω by $\Omega \to g\Omega g^{-1}$. In the case of U(1) this implies that Ω is a well-defined 2-form. Since in every local coordinate it is exact, it is actually a *closed* 2-form, and therefore represents some class $[\frac{i}{2\pi}\Omega] \in H^2_{dR}(M;\mathbb{R})$. We denote this class by c_1.

To understand the meaning of c_1, we must pair it with a 2-dimensional homology class. This means that we take a closed oriented surface S and map it to M by $f : S \to M$, and then consider $c_1(f_*[S])$. Equivalently, we can pull E back to f^*E over S and compute $f^*c_1([S])$. This means just integrating $\frac{i}{2\pi}\Omega$ over S, where now Ω is the curvature 2-form of the (pulled back) connection on f^*E.

Exactly as we argued in § 7.2, if we choose a generic section s of f^*E then each singularity of s contributes 1 or -1 to the integral, and we see that $f^*c_1([S])$ is the *obstruction to finding a nonzero section of f^*E*. In particular, it is an *integer*, and does not depend on the choice of connection. Thus, we have shown that $\frac{i}{2\pi}\Omega$ determines a class $c_1(E) \in H^2(M;\mathbb{Z})$ called the *first Chern class*, which depends only on the smooth bundle E (as a $U(1)$ bundle), and *not* on the choice of connection.

8. Hodge theory

8.1. **The Hodge star.** Let V be a vector space with a positive definite symmetric inner product. As we have observed, the inner product on V determines an inner product on all natural vector spaces obtained from V. For instance, on $\Lambda^p(V)$ it determines an inner product, defined on primitive vectors by

$$\langle v_1 \wedge \cdots \wedge v_p, w_1 \wedge \cdots \wedge w_p \rangle := \det \langle v_i, w_j \rangle$$

Thus, if $e_1, \cdots, e_n$ is an orthonormal basis for V, the vectors of the form $e_{i_1} \wedge \cdots \wedge e_{i_p}$ with $i_1 < \cdots < i_p$ are an orthonormal basis for $\Lambda^p(V)$.

Definition 8.1 (Orientation)**.** Suppose V is n-dimensional, so that $\Lambda^n(V)$ is isomorphic to $\mathbb{R}$ (though not canonically). An *orientation* on V is a choice of connected component of $\Lambda^n(V) - 0$ whose elements are said to be *positively oriented* (with respect to the orientation).

Definition 8.2 (Hodge star)**.** If V is an oriented inner product space, there is a linear map $* : \Lambda^p(V) \to \Lambda^{n-p}(V)$ for each p, defined with respect to any orthonormal basis $e_1, \cdots, e_n$ by the formula

$$*(e_1 \wedge \cdots \wedge e_p) = \pm e_{p+1} \wedge \cdots \wedge e_n$$

where the sign is $+$ if $e_1 \wedge \cdots \wedge e_n$ is positively oriented, and $-$ otherwise.

Applying Hodge star twice defines a map $*^2 : \Lambda^p(V) \to \Lambda^p(V)$ which (by looking at the effect on each basis vector) is just multiplication by $(-1)^{p(n-p)}$. Moreover, the inner product may be expressed simply in terms of Hodge star by

$$\langle v, w \rangle = *(w \wedge *v) = *(v \wedge *w)$$

for any $v, w \in \Lambda^p(V)$.

8.2. **Hodge star on differential forms.** Suppose that M is a smooth manifold, and $\Omega^p(M)$ denotes the space of smooth p-forms. A Riemannian metric on M determines a positive definite inner product on T_xM and thereby on T_x^*M and $\Lambda^p(T_x^*M)$ for each $x \in M$. An orientation on M determines an orientation on $\Lambda^n(T_x^*M)$ for each x. Then applying Hodge star to p-forms fiber by fiber defines a linear operator $* : \Omega^p(M) \to \Omega^{n-p}(M)$. With this notation, the volume form of M is just $*1$; i.e. Hodge star applied to the function 1, thought of as a section of $\Omega^0(M)$.

The pointwise pairing on $\Lambda^p(T_x^*M)$ can be integrated against the volume form $*1$. Since $*^2 = 1$ on functions, this gives the following symmetric positive definite bilinear pairing on $\Omega^p(M)$:

$$\langle \alpha, \beta \rangle := \int_M \alpha \wedge *\beta$$

On a Riemannian manifold, it is sometimes convenient to introduce notation for the canonical isomorphisms between vector fields and 1-forms coming from the metric. We denote the map $\Omega^1(M) \to \mathfrak{X}(M)$ by $\sharp$ and the inverse by $\flat$. Thus (for example) the gradient $\text{grad}(f)$ of a smooth function is $(df)^\sharp$.

Definition 8.3. If X is a vector field on an oriented Riemannian manifold M, the *divergence* of X, denoted $\text{div}(X)$, is the function defined by $\text{div}(X) := *d * X^\flat$.

Note if $\omega := *1$ is the oriented volume form, then $\mathcal{L}_X\omega = d\iota_X(\omega) = \text{div}(X)\omega$. That is to say, a vector field is divergence free if and only if the flow it generates is volume-preserving.

Definition 8.4. The *Laplacian on functions* is the operator $\Delta := -\text{div} \circ \text{grad}$. In terms of Hodge star, this is the operator $\Delta = - * d * d$.

More generally, if we define $\delta : \Omega^p(M) \to \Omega^{p-1}(M)$ by $\delta = (-1)^{n(p+1)+1} * d*$ then the *Laplacian on p-forms* is the operator $\Delta := \delta d + d\delta$.

Lemma 8.5. *Let M be a closed and oriented Riemannian manifold. The operator δ is the adjoint of d on $\Omega^p(M)$. That is, for any $\alpha \in \Omega^{p-1}(M)$ and $\beta \in \Omega^p(M)$ we have*

$$\langle d\alpha, \beta \rangle = \langle \alpha, \delta\beta \rangle$$

Consequently Δ is self-adjoint; i.e. $\langle \Delta\alpha, \beta \rangle = \langle \alpha, \Delta\beta \rangle$ for any $\alpha, \beta \in \Omega^p(M)$.

Proof. We compute

$$d(\alpha \wedge *\beta) = d\alpha \wedge *\beta + (-1)^{p-1}\alpha \wedge d * \beta = d\alpha \wedge *\beta - \alpha \wedge *\delta\beta$$

because $(p-1)(n-p-1) + (p-1)$ is odd. By Stokes' theorem, $\int_M d(\alpha \wedge *\beta) = 0$, so δ is the adjoint of d. That Δ is self-adjoint is immediate. □

Lemma 8.6. $\Delta\alpha = 0$ *if and only if* $d\alpha = 0$ *and* $\delta\alpha = 0$.

Proof. One direction is obvious. To see the other direction,

$$\langle \Delta\alpha, \alpha \rangle = \langle (d\delta + \delta d)\alpha, \alpha \rangle = \langle \delta\alpha, \delta\alpha \rangle + \langle d\alpha, d\alpha \rangle = \|d\alpha\|^2 + \|\delta\alpha\|^2$$

So $\Delta\alpha = 0$ implies $d\alpha = 0$ and $\delta\alpha = 0$. □

A form α is *closed* if $d\alpha = 0$. We say it is *co-closed* if $\delta\alpha = 0$ and *harmonic* if $\Delta\alpha = 0$. Thus, Lemma 8.6 says that α is harmonic if and only if it is closed and co-closed.

Denote the space of harmonic p-forms by $\mathcal{H}^p$.

Lemma 8.7. *A p-form is closed if and only if it is orthogonal to $\delta\Omega^{p+1}$. Similarly, a p-form is co-closed if and only if it is orthogonal to $d\Omega^{p-1}$. Hence a p-form is harmonic if and only if it is orthogonal to both $\delta\Omega^{p+1}$ and $d\Omega^{p-1}$, which are themselves orthogonal.*

Proof. Since $\langle d\alpha, \beta \rangle = \langle \alpha, \delta\beta \rangle$, we have $d\alpha = 0$ if and only if α is orthogonal to $\delta\beta$ for all β. The second statement is proved similarly. Finally, $\langle d\alpha, \delta\beta \rangle = \langle d^2\alpha, \beta \rangle = 0$. □

Suppose $C_0 \xrightarrow{d} C_1 \xrightarrow{d} \cdots \xrightarrow{d} C_n$ is a chain complex of *finite dimensional* vector spaces. Pick a positive definite inner product on each vector space, and define δ to be the adjoint of d. Then as above, an element of C_p is in $\ker(d)$ if and only if it is orthogonal to δC_{p+1}, and similarly an element of C_p is in $\ker(\delta)$ if and only if it is orthogonal to dC_{p-1}. Since C_p is finite dimensional, this implies that $\ker(d) = (\delta C_{p+1})^\perp$. Since $dC_{p-1} \subset \ker(d)$ we can identify the homology group H_p with the orthogonal complement of dC_{p-1} in $\ker(d)$; i.e. $H_p = (dC_{p-1})^\perp \cap \ker(d) = \ker(d) \cap \ker(\delta)$. Thus we have an orthogonal decomposition

$$C_p = H_p \oplus dC_{p-1} \oplus \delta C_{p+1}$$

where $H_p = \ker(d) \cap \ker(\delta)$.

For infinite dimensional vector spaces, we cannot take orthogonal complements. However, the *Hodge theorem* says that the terms in the complex of smooth forms on a closed, oriented manifold admit such a decomposition, and therefore we may identify the harmonic p-forms $\mathcal{H}^p$ with the de Rham cohomology groups H^p_{dR}. Said another way, each (real) cohomology class on a closed, oriented Riemannian manifold admits a *unique* harmonic representative.

In fact, once one knows that a cohomology class admits a harmonic representative α, that representative is easily seen to be unique. For, any two cohomologous forms differ by an exact form, so if $\alpha' = \alpha + d\beta$ is harmonic, then $d\delta d\beta = 0$ so $\delta d\beta$ is closed. But the closed forms are orthogonal to the co-exact forms, so since $\delta d\beta$ is co-exact, it is orthogonal to itself; i.e. $\delta d\beta = 0$. Again, since the co-closed forms are orthogonal to the exact forms, since $d\beta$ is exact, it is orthogonal to itself, and therefore $d\beta = 0$, proving uniqueness.

Another immediate corollary of the definition is that a harmonic p-form α (if it exists) is the *unique* minimizer of $\|\alpha\|^2$ in its cohomology class. For, since harmonic forms are orthogonal to exact forms, $\|\alpha + d\beta\|^2 = \|\alpha\|^2 + \|d\beta\|^2 \ge \|\alpha\|^2$ with equality iff $d\beta = 0$.

8.3. **The Hodge Theorem.**

Theorem 8.8 (Hodge)**.** *Let M be a closed, oriented Riemannian manifold. Then there are orthogonal decompositions*

$$\Omega^p(M) = \mathcal{H}^p \oplus \Delta\Omega^p = \mathcal{H}^p \oplus d\delta\Omega^p \oplus \delta d\Omega^p = \mathcal{H}^p \oplus d\Omega^{p-1} \oplus \delta\Omega^{p+1}$$

Furthermore, $\mathcal{H}^p$ is finite dimensional*, and is isomorphic to the de Rham cohomology group H^p_{dR}.*

We do not give a complete proof of this theorem, referring the reader to Warner [11], § 6.8. But it is possible to explain some of the main ideas, and how the proof fits into a bigger picture.

The main analytic principle underlying the proof is the phenomenon of *elliptic regularity*. To state this principle we must first define an elliptic linear differential operator.

Let Ω be an open domain in $\mathbb{R}^n$. For a vector $\alpha := (\alpha_1, \cdots, \alpha_n)$ of non-negative integers, we introduce the following notation: $|\alpha| := \alpha_1 + \cdots + \alpha_n$, $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, and $\partial^\alpha := \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n}$.

Definition 8.9. A *linear differential operator* on $\mathbb{R}^n$ of order k is an operator of the form

$$Lf := \sum_{|\alpha| \le k} a_\alpha(x) \partial^\alpha f$$

If k is even, the operator L is (strongly) *elliptic* if there is an estimate of the form

$$(-1)^{k/2} \sum_{|\alpha| = k} a_\alpha(x) \xi^\alpha \ge C|\xi|^k$$

for some positive constant C, for all $x \in \Omega$ and all $\xi \in T_x^*\mathbb{R}^n = \mathbb{R}^n$.

The function σ_x on $T_x^*\mathbb{R}^n$ defined by $\sigma_x(\xi) := \sum_{|a|=k} a_\alpha(x)\xi^\alpha$ is called the *symbol* of L at x. For a linear differential operator on sections of some smooth bundle on a manifold, the symbol is a tensor field. For L a linear operator on functions of order 2, L is elliptic if and only if the symbol is a negative definite symmetric inner product on T^*M.

Algebraically, the symbol can be expressed in the following terms. For smooth bundles E and F over M, let $\mathrm{Diff}_k(E, F)$ denote the space of differential operators of order $\le k$

from $\Gamma(E)$ to $\Gamma(F)$. There are inclusions $\mathrm{Diff}_l(E,F) \to \mathrm{Diff}_k(E,F)$ for any $l \le k$. Then the symbol map fits into an exact sequence

$$0 \to \mathrm{Diff}_{k-1}(E,F) \to \mathrm{Diff}_k(E,F) \xrightarrow{\sigma} \mathrm{Hom}(S^k(T^*M)\otimes E, F) \to 0$$

Example 8.10. We compute the Laplacian $\Delta = -\mathrm{div}\circ\mathrm{grad}$ on functions on a closed, oriented Riemannian manifold. In local coordinates x_i we express the metric as $\sum g_{ij} dx^i \otimes dx^j$ and define g^{ij} to be the coefficients of the inverse matrix; i.e. $\sum_k g^{ik} g_{kj} = \delta_{ij}$ where δ_{ij} is the Kronecker delta, which equals 1 if $i = j$ and 0 otherwise. In these coordinates, the volume form ω is $\sqrt{\det(g_{ij})}dx^1 \wedge \cdots \wedge dx^n$, and the $\sharp$ and $\flat$ isomorphisms are $\sharp : dx^i \to \sum_j g^{ij}\partial_j$ and $\flat : \partial_i \to \sum_j g_{ij} dx^j$.

In coordinates, $\mathrm{grad} f = (df)^\sharp = (\sum_i \partial^i f dx^i)^\sharp = \sum g^{ij}\partial_i f \partial_j$. The formula for the divergence can be derived by using the fact that it is the adjoint of the gradient; hence for any vector field X we have

$$\begin{aligned}
\langle X, \mathrm{grad} f\rangle &= \int_M \langle X^i\partial_i, g^{kj}\partial_k f \partial_j\rangle \sqrt{\det(g)}dx^1 \wedge \cdots \wedge dx^n \\
&= \int_M X^i(\partial_k f) g^{kj} g_{ij} \sqrt{\det(g)} dx^1 \wedge \cdots dx^n \\
&= \int_M X^i(\partial_i f)\sqrt{\det(g)}dx^1 \wedge \cdots dx^n \\
&= -\int_M f \cdot \partial_i (X^i \sqrt{\det(g)}) dx^1 \wedge \cdots dx^n \\
&= \langle f, -\frac{1}{\sqrt{\det(g)}}\partial_i(X^i\sqrt{\det(g)})\rangle
\end{aligned}$$

where we sum over repeated indices (repressed in the notation to reduce clutter). Note that the penultimate step was integration by parts; these local formulae must therefore be interpreted chart by chart where the coordinates are defined. In particular, we obtain the formula

$$\mathrm{div} X = -\sum_i \frac{1}{\sqrt{\det(g)}}\partial_i(X^i\sqrt{\det(g)})$$

and therefore

$$\Delta f = -\sum_{i,j} \frac{1}{\sqrt{\det(g)}}\partial_j(g^{ij}\sqrt{\det(g)}\partial_i f) = -g^{ij}\partial_j\partial_i f + \text{ lower order terms}$$

We therefore see that the symbol $\sigma(\Delta)$ is a quadratic form on T^*M which is the negative of the inner product coming from the metric. In particular, this quadratic form is negative definite, so Δ is elliptic.

Here in words is the crucial analytic property of ellipticity. Let L be an elliptic operator of order k (on functions, for simplicity). Suppose we want to solve an equation of the form $Lu = f$ for some fixed (smooth) function f. Ellipticity lets us bound the kth order derivatives of a solution u in terms of lower order derivatives of u and in terms of f. Inductively, *all* the derivatives of a solution may be estimated from the first few derivatives of u. This process, known as *elliptic bootstrapping*, shows that a weak (L^2) solution u actually has enough regularity that it is smooth.

In our context, elliptic regularity (for the Laplacian acting on p-forms) takes the following form.

Proposition 8.11. *(1) Let ℓ be a weak (i.e. L^2) solution to $\Delta\omega = \alpha$; that is, a bounded linear functional ℓ on Ω^p (in the L^2 norm) so that $\ell(\Delta^*\varphi) = \langle\alpha, \varphi\rangle$ for all $\varphi \in \Omega^p$. Then there is $\omega \in \Omega^p$ such that $\ell(\beta) = \langle\omega, \beta\rangle$. In particular, $\Delta\omega = \alpha$.*

(2) Let α_i be a sequence in Ω^p such that $\|\alpha_i\|$ and $\|\Delta\alpha_i\|$ are uniformly bounded. Then some subsequence converges in L^2.

Note that $\Delta^* = \Delta$, so that a weak solution ℓ to $\Delta\omega = \alpha$ is one satisfying $\ell(\Delta\varphi) = \langle\alpha, \varphi\rangle$ for all $\varphi \in \Omega^p$. Assuming this proposition, we can prove the Hodge theorem as follows.

Proof. If we can show $\mathcal{H}^p$ is finite dimensional, the latter two isomorphisms follow formally from the first, by what we have already shown. Thus we just need to show that $\mathcal{H}^p$ is finite dimensional, and $\Omega^p = \mathcal{H}^p \oplus \Delta\Omega^p$.

Suppose $\mathcal{H}^p$ is infinite dimensional, and let φ_i be an infinite orthonormal sequence. Bullet (2) from Proposition 8.11 says that φ_i contains a subsequence which converges in L^2, which is a contradiction. Thus $\mathcal{H}^p$ is finite dimensional, with an orthonormal basis $\omega_1, \cdots, \omega_m$.

If $\alpha \in \Omega^p$ is arbitrary, we can write

$$\alpha = \beta + \sum \langle\alpha, \omega_i\rangle\omega_i$$

where $\beta \in (\mathcal{H}^p)^\perp$. Now, $\Delta\Omega^p$ is contained in $(\mathcal{H}^p)^\perp$, since Δ is self-adjoint. So it suffices to show that for every $\alpha \in (\mathcal{H}^p)^\perp$ the equation $\Delta\omega = \alpha$ has a weak solution, since then by bullet (1) from Proposition 8.11 it has a smooth solution.

So suppose $\alpha \in (\mathcal{H}^p)^\perp$ and define ℓ on $\Delta\Omega^p$ by

$$\ell(\Delta\varphi) = \langle\alpha, \varphi\rangle$$

Note that ℓ is well-defined, since if $\Delta\varphi_1 = \Delta\varphi_2$ then $\varphi_1 - \varphi_2 \in \mathcal{H}^p$ which is orthogonal to α. We must show that ℓ is bounded.

Let $\varphi \in \Omega^p$ and write $\varphi = \psi + \sum\langle\varphi, \omega_i\rangle\omega_i$. Then

$$|\ell(\Delta\varphi)| = |\ell(\Delta\psi)| = |\langle\alpha, \psi\rangle| \le \|\alpha\|\|\psi\|$$

Now, we claim that there is some uniform constant C so that $\|\psi\| \le C\|\Delta\psi\|$ for all $\psi \in (\mathcal{H}^p)^\perp$. For otherwise we can take some sequence $\psi_i \in (\mathcal{H}^p)^\perp$ with $\|\psi_i\| = 1$ and $\|\Delta\psi_i\| \to 0$, and by bullet (2) of Proposition 8.11 some subsequence converges in L^2. But if we define the L^2 limit to be the weak operator $\ell'(\phi) = \lim_{i\to\infty}\langle\psi_i, \phi\rangle$ then $\ell'(\Delta\phi) = \lim_{i\to\infty}\langle\psi_i, \Delta\phi\rangle = \lim_{i\to\infty}\langle\Delta\psi_i, \phi\rangle = 0$ so ℓ' is weakly harmonic (i.e. a weak solution to $\Delta\omega = 0$), so there is some $\psi_\infty \in \mathcal{H}^p$ with $\langle\psi_i, \beta\rangle \to \langle\psi_\infty, \beta\rangle$. But then $\|\psi_\infty\| = 1$ and $\psi_\infty \in (\mathcal{H}^p)^\perp$ which is a contradiction, thus proving the claim.

So in conclusion,

$$|\ell(\Delta\varphi)| = \|\alpha\|\|\psi\| \le C\|\alpha\|\|\Delta\psi\| = C\|\alpha\|\|\Delta\varphi\|$$

so ℓ is a *bounded* linear operator on $\Delta\Omega^p$, and therefore by the Hahn-Banach theorem extends to a bounded linear operator on Ω^p. Thus ℓ is a weak solution to $\Delta\omega = \alpha$ and therefore a strong solution ω exists. So $(\mathcal{H}^p)^\perp = \Delta\Omega^p$ as claimed. □

We now explain the main steps in the proof of Proposition 8.11, for the more general case of an elliptic operator L of order k. For more details, see e.g. Warner [11], pp. 227–251, Rosenberg [10], pp. 14–39 or Gilkey [4], Ch. 1.

If we work in local coordinates, the analysis on M reduces to understanding analogous operators on the space $C_c^\infty(\Omega)$ of compactly supported smooth functions in an open domain $\Omega \subset \mathbb{R}^n$, and its completions in various norms. For each s define the *sth Sobolev norm* on $C_c^\infty(\Omega)$ to be

$$\|f\|_s := \Big(\sum_{|\alpha|\le s} \|\partial^\alpha f\|^2\Big)^{1/2}$$

In other words, a sequence of functions f_i converges in $\|\cdot\|_s$ if the functions and their derivatives of order $\le s$ converge in L^2. The completion in the norm $\|\cdot\|_s$ is denoted $H_s(\Omega)$, and is called the *sth Sobolev space*.

The basic trick is to use Fourier transform to exchange differentiation for multiplication. If we work in local coordinates, the Fourier transform $\mathcal{F}$ on $\mathbb{R}^n$ is the operator

$$\mathcal{F}(f)(\xi) := \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x)dx$$

It is well-defined on the space of compactly supported smooth functions $C_c^\infty(\mathbb{R}^n)$, and is an isometry in the L^2 norm, and therefore extends as an isometry to $L^2(\mathbb{R}^n)$.

Fourier transform satisfies the following properties:

(1) it interchanges multiplication and convolution:

$$\mathcal{F}(f*g) = \mathcal{F}(f)\mathcal{F}(g), \quad \mathcal{F}(fg) = \mathcal{F}(f)*\mathcal{F}(g)$$

(2) its inverse is given by the transform $f(x) = \int_{\mathbb{R}^n} e^{ix\cdot\xi}\mathcal{F}(f)(\xi)d\xi$;

(3) it interchanges differentiation and coordinate multiplication:

$$\mathcal{F}(D^\alpha f) = \xi^\alpha\mathcal{F}(f), \quad \mathcal{F}(x^\alpha f) = D^\alpha\mathcal{F}(f)$$

here we use the notation D^α for $(-i)^{|\alpha|}\partial^\alpha$.

From the third property above one obtains an estimate of the Sobolev norm in terms of Fourier transform of the form

$$\|f\|_s \sim \Big(\int_{\mathbb{R}^n} |\mathcal{F}(f)|^2(1+|\xi|^2)^s d\xi\Big)^{1/2}$$

Then the definition of an elliptic operator gives rise to a fundamental inequality of the following form:

Theorem 8.12 (Gårding's Inequality). *Let L be elliptic of order k. Then for any s there is a constant C so that*

$$\|f\|_{s+k} \le C(\|Lf\|_s + \|f\|_s)$$

Proof. We have

$$\|f\|_{s+k}^2 \sim \int_{\mathbb{R}^n} |\mathcal{F}(f)(\xi)|^2(1+|\xi|^2)^{s+k} d\xi$$

When $|\xi|$ is small, $(1+|\xi|^2)^{s+k}$ is comparable to $(1+|\xi|^2)^s$. When $|\xi|$ is big, $(1+|\xi|^2)^{s+k}$ is comparable to $|\sigma(L)(\xi)|^2(1+|\xi|^2)^s$, because L is elliptic. Since Fourier transform interchanges differentiation and coordinate multiplication, $\|Lf\|_s^2$ can be approximated by an integral over $\mathbb{R}^n$, where the integrand is itself approximated (for big $|\xi|$) by $|\mathcal{F}(f)(\xi)|^2|\sigma(L)(\xi)|^2(1+|\xi|^2)^s$. The estimate follows. □

This is largely where the role of ellipticity figures into the story. The remainder of the argument depends on fundamental properties of Sobolev norms.

Theorem 8.13 (Properties of Sobolev norms)**.** *Let Ω be a domain in $\mathbb{R}^n$, and let $H_s(\Omega)$ denote the completion of $C_c^\infty(\Omega)$ in the sth Sobolev norm.*

1. *(Rellich Compactness): if $t > s$ the inclusion $H_t(\Omega) \to H_s(\Omega)$ is compact.*
2. *(Sobolev Embedding): if $f \in H_k(\Omega)$ then $f \in C^s(\overline{\Omega})$ for all $s < k - \frac{n}{2}$.*

Proof. To prove Rellich compactness, one first argues that for any sequence f_i with $\|f_i\|_t$ bounded, there is a subsequence for which $\mathcal{F}(f_i)$ converges uniformly on compact subsets. Once this is done, after taking Fourier transforms, we only need to control

$$\|f_i - f_j\|_s^2 \sim \int_R |\mathcal{F}(f_i) - \mathcal{F}(f_j)|^2(1+|\xi|^2)^s d\xi$$

for regions of the form $R = \{\xi \text{ such that } |\xi| > r\}$. But since $\|f_i - f_j\|_t$ is bounded, it follows that we have a uniform estimate of the form

$$\int_R |\mathcal{F}(f_i) - \mathcal{F}(f_j)|^2(1+|\xi|^2)^t d\xi < C$$

Thus

$$\begin{aligned}\int_R |\mathcal{F}(f_i) - \mathcal{F}(f_j)|^2(1+|\xi|^2)^s d\xi &\le \int_R (1+r^2)^{s-t} \int_R |\mathcal{F}(f_i) - \mathcal{F}(f_j)|^2(1+|\xi|^2)^t d\xi \\ &\le C \cdot (1+r^2)^{s-t}\end{aligned}$$

which is arbitrarily small when r is large.

The Sobolev embedding theorem is proved by first using Cauchy-Schwarz (applied to the inverse Fourier transform) to estimate the value of a function f at a point in terms of $\|f\|_k$:

$$\begin{aligned}|f(x)| &= \left|\int_{\mathbb{R}^n} e^{ix\cdot\xi}\mathcal{F}(f)(\xi)d\xi\right| \\ &= \left|\int_{\mathbb{R}^n} e^{ix\cdot\xi}(1+|\xi|^2)^{-k/2}(1+|\xi|^2)^{k/2}\mathcal{F}(f)(\xi)d\xi\right| \\ &\le \left(\int_{\mathbb{R}^n}(1+|\xi|^2)^{-k}d\xi\right)^{1/2} \cdot \left(\int_{\mathbb{R}^n}|\mathcal{F}(f)(\xi)|^2(1+|\xi|^2)^k d\xi\right)^{1/2}\end{aligned}$$

The points is that if $k > n/2$, the first term is *finite*, so $f(x) \le C \cdot \|f\|_k$ for some universal (dimension-dependent) constant C.

Since any $f \in H_k$ is the limit in the Sobolev k norm of a sequence of smooth functions f_i, this estimate implies that $f_i \to f$ uniformly, so we deduce that f is *continuous*. Higher derivatives are controlled in a similar manner. □

Now, Theorem 8.12 and Rellich Compactness (i.e. Theorem 8.13 bullet (1)) together imply that if α_i is a sequence in Ω^p with $\|\alpha_i\|$ and $\|\Delta\alpha_i\|$ bounded, then $\|\alpha_i\|_2$ is bounded, and therefore the image of the α_i in L^2 contains a convergent subsequence, thus proving Proposition 8.11, bullet (2).

Similarly, if ω is an L^2 solution to $\Delta\omega = \alpha$, then we can estimate $\|\omega\|_{2+k} \le C \cdot (\|\alpha\|_k + \|\omega\|_k)$ by Theorem 8.12, so inductively we can show $\omega \in H_s$ for all s. Then Sobolev Embedding (i.e. Theorem 8.13 bullet (2)) proves that ω in C^∞, thus proving Proposition 8.11, bullet (1). This completes the sketch of the proof of Proposition 8.11 and the Hodge Theorem.

8.4. **Weitzenböck formulae.** Once we have discovered the idea of taking the adjoint of a differential operators like d, it is natural to study adjoints of other (naturally defined) differential operators, and their properties.

If we think of ∇ as a differential operator $\nabla : \Omega^p M \to T^*M \otimes \Omega^p M$ then giving these spaces their natural inner product, we can define an adjoint ∇^* and form the *Bochner Laplacian* $\nabla^*\nabla$, which is a self-adjoint operator from Ω^p to itself.

It turns out that Δ and $\nabla^*\nabla$ have the same symbol, and therefore their difference is *a priori* a differential operator of order 1. But because both operators are "natural", the symbol of their difference should be invariant under the action of the orthogonal group of T_x^*M for each x. Of course, the only invariant such vector is zero, so the difference turns out to be of 0th order — i.e. it is a tensor field, and can be expressed in terms of the curvature operator.

In the special case of 1-forms, this simplifies considerably, and one has the *Bochner formula*:

Proposition 8.14 (Bochner formula)**.** *On 1-forms, there is an identity*

$$\Delta = \nabla^*\nabla + \mathrm{Ric}$$

Here we think of Ric as a symmetric quadratic form on T_x^*M for each x by identifying it with T_xM using the metric.

Proof. We already know that the composition of ∇ with antisymmetrization agrees with d. If we let $e_1, \cdots, e_n$ be a local orthonormal frame for TM, and let $\eta_1, \cdots, \eta_n$ denote a dual basis for T^*M, then

$$d\omega = \sum_i \eta_i \wedge \nabla_{e_i}\omega$$

and by the properties of the Hodge star in an orthonormal basis,

$$\delta\omega = -\sum_i \iota(e_i)\nabla_{e_i}\omega$$

Thus we can compute

$$\delta d\omega = -\sum_{i,j} \iota(e_j)\nabla_{e_j}(\eta_i \wedge \nabla_{e_i}\omega)$$

and

$$d\delta\omega = \sum_{i,j} \eta_i \wedge \nabla_{e_i}(-\iota(e_j)\nabla_{e_j}\omega)$$

If we work at a point p where the e_i are normal geodesic coordinates, so that $\nabla e_i|_p = 0$, then we get

$$\delta d\omega|_p = -\sum_{i,j} \iota(e_j)\eta_i \wedge \nabla_{e_j}\nabla_{e_i}\omega|_p$$

and

$$d\delta\omega|_p = -\sum_{i,j} \eta_i \wedge \iota(e_j)\nabla_{e_i}\nabla_{e_j}\omega|_p$$

□

9. Minimal surfaces

10. Acknowledgments

Danny Calegari was supported by NSF grant DMS 1005246.

References

[1] V. Arnol'd, *Sturm theorems and symplectic geometry*, Funct. Anal. Appl. **19** (1985), no. 4, 251–259
[2] J. Cheeger and D. Ebin, *Comparison theorems in Riemannian geometry*, Revised reprint of the 1975 original. AMS Chelsea Publishing, Providence, RI, 2008
[3] T. Colding and W. Minicozzi, *A course in minimal surfaces*, Graduate Studies in Mathematics, 121. American Mathematical Society, Providence, RI, 2011
[4] P. Gilkey, *Invariance theory, the Heat equation, and the Atiyah-Singer Index Theorem*, Second Edition. Stud. Adv. Math. CRC Press, Boca Raton, FL, 1995
[5] S. Kobayashi and K. Nomizu, *Foundations of differential geometry, Vols. 1,2*, Interscience Publishers, Wiley, New York-London, 1963
[6] S. Myers and N. Steenrod, *The group of isometries of a Riemannian manifold*, Ann. of Math. (2) **40** (1939), 400–416
[7] J. Milnor, *Morse theory*, Ann. of Math. Studies, Princeton University Press, Princeton, NJ, 1963
[8] NASA, *Aircraft Rotations*, `http://www.grc.nasa.gov/WWW/K-12/airplane/rotations.html`
[9] M. Postnikov, *Lectures in Geometry Semester V: Lie Groups and Lie Algebras*, Mir Publishers, Moscow, 1986
[10] S. Rosenberg, *The Laplacian on a Riemannian manifold*, LMS Student Texts **31**, Cambridge University Press, 1997
[11] F. Warner, *Foundations of differentiable manifolds and Lie groups*, Corrected reprint of the 1971 edition. Graduate Texts in Mathematics, 94. Springer-Verlag, New York-Berlin, 1983.

University of Chicago, Chicago, Ill 60637 USA
E-mail address: dannyc@math.uchicago.edu