Riemann–Finsler Geometry
MSRI Publications
Volume **50**, 2004

Nonreversible Finsler Metrics of Positive Flag Curvature

HANS-BERT RADEMACHER

Contents

1. Introduction

Finsler geometry is an essential extension of Riemannian geometry. Instead of an inner product on every tangent space one considers Minkowski norms on every tangent space. For a Finsler metric the unit sphere in each tangent space is a strictly convex hypersurface. One obtains for every nonzero tangent vector an inner product, arising from Minkowski norm; in the Riemannian case these inner products all coincide on a fixed tangent space. The length of a smooth curve is well-defined. Geodesics — locally length-minimizing curves parametrized with constant speed — are uniquely defined for a given initial direction. From the viewpoint of the calculus of variations Finsler metrics are a suitable generalization of Riemannian metrics such that the variational problem for the length of curves between two fixed points is positive and positive regular. In terms of physics a Finsler metric describes a Lagrangian system without a potential; a Riemannian metric can be viewed as the special case of quadratic kinetic energy.

But in contrast to the Riemannian case there is no canonical connection, so several connections have been used in Finsler geometry. We use here the one introduced by S.-S. Chern [Bao et al. 2000, Chapter 2], transposed to vector fields on the manifold for a fixed direction field: Given a nowhere vanishing vector field V in an open nonempty subset U, there is a uniquely determined torsionfree connection ∇^V that is almost metric. Using this connection one can define the *flag curvature*, which generalizes the sectional curvature in Riemannian geometry and controls the infinitesimal behavior of geodesics. Given a geodesic c and a nowhere vanishing geodesic vector field V in an open neighborhood of c extending the velocity field c' of the geodesic, there is a Riemannian metric g_V on U such that c is also a geodesic of the Riemannian manifold (U, g_V) and the flag curvature $K(c'; \sigma)$ for any plane σ containing c' coincides with the sectional curvature $K(\sigma)$ of g_V. In particular the Jacobi fields of the Finsler metric and of g_V coincide along the geodesic c. Thus the flag curvature can be introduced without selecting a connection. The flag curvature does not completely determine the metric. For example, in contrast to the Riemannian case, there are Finsler metrics of constant positive flag curvature on spheres that are not isometric to the standard Riemannian metric; see [Bao et al. 2003; Bryant 2002; Shen 2002] for their characterization.

We call a Finsler metric F *reversible* (or *symmetric*) if opposite vectors have the same length: $F(X) = F(-X)$ for all tangent vectors X. In this case the unit sphere $T_p^1 M = \{X \in T_pM \mid F(X) = 1\}$ is symmetric under reflection through the origin. But this assumption excludes many interesting examples, for example *Randers metrics*, which are Finsler metrics defined by adding a one-form to the norm induced by a Riemannian metric: $F(X) = \sqrt{g(X, X)} + \alpha(X)$, where g is a Riemannian metric and α is a one-form.

In Riemannian geometry a metric with constant positive sectional curvature on a compact simply connected manifold is isometric to the standard sphere of the same curvature. The now classical *Sphere Theorem* states that a compact, simply connected manifold of dimension n with sectional curvature K such that $\frac{1}{4} < K \leq 1$ everywhere is homeomorphic to the n-sphere [Klingenberg 1995, § 2.8; Abresch and Meyer 1997]. In this form the result is contained in [Klingenberg 1961]; earlier contributions are due to M. Berger, H. Rauch and V. A. Toponogov [Berger 1998, I A 2]. The proof uses an estimate for the injectivity radius and the Toponogov comparison theorem for geodesic triangles. In [Klingenberg 1963] it is shown that one can prove the Sphere Theorem without making use of Toponogov's comparison result for geodesic triangles. Instead one uses Morse theory of the energy functional on the space of curves between two fixed points and on the space of loops. The injectivity radius is bounded from below by π, so geodesic loops have length at least 2π and their Morse index is bounded below by $n-1$. This implies that the loop space is $(n-2)$-connected and therefore the manifold is homotopy equivalent to the n-sphere.

Though in Finsler geometry the condition of constant positive flag curvature no longer determines the metric up to isometry, one can show using the exponential map that a simply connected Finsler manifold of constant positive flag curvature is homeomorphic to the n-sphere. P. Dazord [1968a; 1968b] remarked that in the case of a *reversible* Finsler metric one can carry over the Morse-theoretic proof of the Sphere Theorem found in [Klingenberg 1963]. The original proof of the Sphere Theorem does not carry over, since the triangle comparison result cannot be extended to the Finsler case.

The main topic of this article is to show in detail how the estimates for the injectivity radius, the length of a nonminimal geodesic between two fixed points, the length of a nonconstant geodesic loop and the length of a nonconstant closed geodesic can be extended to the case of a *nonreversible* Finsler metric by introducing the notion of *reversibility* $\lambda := \sup\{F(-X) \mid F(X) = 1\} \geq 1$. We will derive from a length estimate (Proposition 9.9) and from Theorem 9.10 the following *Sphere Theorem for nonreversible Finsler metrics*:

THEOREM 9.11. *A simply connected and compact Finsler manifold of dimension $n \geq 3$ with reversibility λ and with flag curvature K satisfying $\left(1 - \frac{1}{1+\lambda}\right)^2 < K \leq 1$ is homotopy equivalent to the n-sphere.*

A proof appears in [Rademacher 2004]. In this article we will present this result in detail, adopt a slightly different approach at places. The examples due to A. Katok of nonreversible Finsler metrics on S^2 with only two geometrically distinct closed geodesics are of great importance in the theory of closed geodesic as a test case for several statements [Rademacher 1992, § 5.3]. It was pointed out in [Rademacher 2004, Chapter 5] that the Finsler metric of Katok's example on S^2 coincides with the Finsler metric of constant flag curvature constructed in [Shen 2002]. These examples show that the length estimate for a shortest geodesic given in Theorem 9.10 is sharp. Using the Legendre transformation, we see that Katok's examples describe Finsler metrics of *Randers type*.

It remains an open problem whether one can improve the Sphere Theorem in the nonreversible case by choosing the lower curvature bound $\frac{1}{4}$ as in the reversible case.

2. Conventions

We consider metric structures on a differentiable manifold $M = M^n$ of dimension n. If not otherwise stated, differentiable means C^∞-differentiable. The *tangent bundle* of M is denoted by TM, with projection $\tau : TM \to M$, and $T_xM := \tau^{-1}(x)$ for $x \in M$. We denote by $\mathcal{V}M$ the *vector space of smooth vector fields* on M, that is, the space $\Gamma(TM)$ of smooth sections of the tangent bundle. The zero section T^0M of TM is the union of the zero vectors $0_x \in T_xM$; it can be identified with M. The *cotangent bundle* of M is denoted by T^*M.

If $(x^1, \dots, x^n)$ are coordinates on M, the coordinate vector fields $(\partial_1, \dots, \partial_n)$ defined by $\partial_i(x) = (\partial/\partial x^i)(x)$ form a basis for the tangent space T_xM. For this set of coordinates, the tangent bundle can be given *canonical coordinates* by associating $(x, y) = (x^1, \dots, x^n, y^1, \dots, y^n)$ to the tangent vector $\sum_{i=1}^n y^i \partial_i(x) \in T_xM$. A vector field V can be written as $V(x) = \sum_{i=1}^n v^i(x)\partial_i(x)$.

The real vector space of differentiable functions $f : M \to \mathbb{R}$ is denoted by $\mathcal{F}M = C^\infty(M)$. A multilinear map

$$A : \underbrace{\mathcal{V}M \times \cdots \times \mathcal{V}M}_{k} \to \mathcal{F}M$$

is called a $(0, k)$-*tensor field* on M if it is linear in each argument with respect to the vector space $\mathcal{F}M$. A multilinear map $A : \mathcal{V}M \times \cdots \times \mathcal{V}M \mapsto \mathcal{V}M$ is called a $(1, k)$-*tensor field* on M if it satisfies the same condition. A $(0, k)$-tensor field A on M is *symmetric* if for any $x \in M$ the induced k-linear map $A_x : T_xM \times \dots \times T_xM \to \mathbb{R}$ is symmetric, that is, satisfies $A_x(X_{\sigma(1)}, \dots, X_{\sigma(k)}) = A_x(X_1, \dots, X_k)$ for all $X_1, \dots, X_k \in T_xM$ and all permutations $\sigma \in \mathfrak{S}_k$. Symmetric $(1, k)$-tensor fields are defined similarly.

Let X be a vector field on M and let $f \in \mathcal{F}M$. For $p \in M$ and $\gamma : (-\varepsilon, \varepsilon) \to M$ a smooth curve with $\gamma(0) = p$ and $\gamma'(0) = X(p)$, the quantity $(d/dt)|_{t=0} f(\gamma(t))$ does not depend on the choice of γ. As p varies, this defines a new function on M, called the *Lie derivative* of f in the direction of X, and written Xf or $X(f)$. We also write $df(p)X$ for $Xf(p)$.

The projection $\tau : TM \to M$ induces by differentiation the *double tangent bundle* $\tau_* : TTM \to TM$; for any $X \in T_xM$, the space $T_X(TM)$ is called the *double tangent space* at $X \in TM$ and has dimension $2n$, where n is the dimension of M. The tangent vectors $Y'(0)$ of *vertical curves* $Y : (-\varepsilon, \varepsilon) \to T_xM \subset TM$ with $Y(0) = X$ span a distinguished n-dimensional subspace of $T_X(TM)$, called the *vertical tangent space* and denoted by $T_X^v(TM) = T_X(T_xM)$. Hence $T_X^v(TM) = \ker\big(d\tau : T_X(TM) \to T_xM\big)$. Together the vertical subspaces form the *vertical subbundle* $T^vTM \subset TTM$.

Given a tangent vector $Y \in T_xM$, we define a map $\overline{Y} : T_xM \to T(T_xM) = T_xM \times T_xM$ by setting $\overline{Y}(X) = (X, Y)$. Then from any vector field on M we obtain an associated vertical vector field on TM (that is, a section of T^vTM): its value at $X \in T_xM$ is $\overline{Y}(X)$, where Y is the value at x of the given vector field on M. All of this is independent of coordinates.

If Y is a vector field on M with associated vertical vector field $\overline{Y}$ on TM, the Lie derivative of a function $F : U \subset TM \to \mathbb{R}$ with respect to $\overline{Y}$ is given by

$$\overline{Y}F(V) = (d/dt)|_{t=0} F(V + tY).$$

In terms of a set of canonical coordinates $(x, y) = (x^1, \dots, x^n, y^1, \dots, y^n)$ on TM, we have $\overline{\partial/\partial x^i}F = \partial/\partial y^i F$ for F equal to each coordinate function; hence

$$\overline{\partial/\partial x^i} = \partial/\partial y^i.$$

3. Finsler Metrics

DEFINITION 3.1. A *Finsler manifold* (M, F) is a differentiable manifold M equipped with a Finsler metric F. A *Finsler metric* on M is a continuous map, $F : TM \to \mathbb{R}$ differentiable outside the zero section T^0M and satisfying three conditions:

(1) F is *positively homogeneous*, that is, $F(\mu X) = \mu F(X)$ for all positive $\mu \in \mathbb{R}$ and all tangent vectors $X \in TM$.
(2) If $F(X) = 0$ then $X = 0$.
(3) The *Legendre condition* or *strong convexity condition*: for any nonzero $V \in T_xM$, the symmetric bilinear form $g_V : T_xM \times T_xM \to \mathbb{R}$ given by

$$g_V(X, Y) = \langle X, Y \rangle_V := \tfrac{1}{2}\overline{X}\,\overline{Y} F^2(V) = \frac{1}{2}\frac{\partial^2}{\partial s \partial t}\bigg|_{\substack{s=0\\t=0}} F^2(V + sX + tY)$$

is positive definite.

REMARK 3.2. (a) In terms of a set of canonical coordinates $(x, y) = (x^1, \ldots, x^n, y^1, \ldots, y^n)$ on TM, and setting

$$g_{ij}(x, y) := g_{(x,y)}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \frac{1}{2}\frac{\partial^2 F^2}{\partial y^i \partial y^j}(x, y),$$

the Legendre condition states that the symmetric matrix $\big(g_{ij}(x, y)\big)_{1 \le i,j \le n}$ is positive definite whenever $y \neq 0$.

(b) Since $F(\mu X) = \mu F(X)$ for all $\mu > 0$, we have

$$\langle V, V \rangle_V = \frac{1}{2}\frac{d^2}{dt^2}\bigg|_{t=0} F^2(V + tV) = \frac{1}{2}\frac{d^2}{dt^2}\bigg|_{t=0} (1 + t)^2 F^2(V) = F^2(V)$$

In coordinates,

$$\frac{1}{2} y^i y^j \frac{\partial^2}{\partial y^i \partial y^j} F^2(x, y) = F^2(x, y).$$

(c) *Euler's Theorem* states that for a positively homogeneous function $f : V \to \mathbb{R}$ of order k on a vector space V (meaning that $f(\mu X) = \mu^k f(X)$ for $X \in V$ and $\mu > 0$), the radial derivative coincides up to the factor k with f itself:

$$\sum_{i=1}^{n} y^i \frac{\partial}{\partial y^i} f(y) = k f(y).$$

See [Bao et al. 2000, Theorem 1.2.1].

DEFINITION 3.3. The *Legendre transformation* on a Finsler manifold (M, F) is the map $\mathcal{L}_F : TM \to T^*M$ defined by

$$\mathcal{L}_F(V)(W) = g_V(V, W).$$

One can view $\mathcal{L}_F(V)$ as the 1-form dual to V with respect to the metric g_V.

LEMMA 3.4. *If F is a Finsler metric, $F(X+Y) \leq F(X) + F(Y)$ for all $X, Y \in T_xM$. Equality holds only if $Y = \mu X$ for some $\mu \geq 0$.*

This implies that $\langle X, Y\rangle_Y \leq F(X)F(Y)$ for all $X, Y \in T_xM$, with equality if and only if $Y = \mu X$ for some $\mu \geq 0$ [Shen 2001a, § 1.2; Bao et al. 2000, p. 10].

LEMMA 3.5 [Bao et al. 2000, Proposition 14.8.1; Shen 2001a, Lemma 1.2.4]. *Let $V, W \in T_xM$ be nonzero. If $\langle X, V\rangle_V = \langle X, W\rangle_W$ for all $X \in T_xM$, then $V = W$.*

Consequently, the Legendre transformation is an isomorphism.

A *Riemannian metric* g on a manifold M is symmetric $(2,0)$-tensor field $g : \mathcal{V}M \times \mathcal{V}M \to \mathcal{F}M$ such that for every $x \in M$ the bilinear map $g_x : T_xM \times T_xM \to \mathbb{R}$ is positive definite. The associated Finsler metric is defined by $F(X) = \sqrt{g(X,X)}$; forming the metric g_V for any nonzero V we obtain g for every nonzero V.

EXAMPLE 3.6 (RANDERS METRICS). Suppose given a Riemannian metric α and a differential 1-form β. There is a vector field ζ satisfying $\beta(X) = \alpha(X, \zeta)$ for all X; we say ζ is *dual* to β with respect to α. Define $\|\beta\| := \|\zeta\| = \sqrt{\alpha(\zeta,\zeta)}$. If $\|\beta\| < 1$ everywhere,

$$F(X) := \sqrt{\alpha(X,X)} + \beta(X) = \sqrt{\alpha(X,X)} + \alpha(X,\zeta)$$

defines a Finsler metric. This type of Finsler metric is called a *Randers metric*.

Since $X \neq 0$ implies $\|X\| = \sqrt{\alpha(X,X)} > 0$, we have

$$F(X) = \|X\| \left(1 + \alpha\left(\frac{X}{\|X\|}, \zeta\right)\right) \geq \|X\| \left(1 - \|\zeta\|\right) > 0,$$

showing that F satisfies condition (2) of Definition 3.1. Condition (1) is obvious. For the proof of (3) we refer to [Bao et al. 2000, Chapter 11].

A Randers metric $F(X) = \sqrt{\alpha(X,X)} + \alpha(X,\zeta)$ is only positively homogeneous. If $\alpha(X,\zeta) \neq 0$ then $F(-X) \neq F(X)$. This motivates the following notion:

DEFINITION 3.7. On a Finsler manifold (M, F) the *reversibility function* $\lambda : M \to \mathbb{R}^+$ is defined by

$$\lambda(x) := \sup \bigl\{ F(-X) \mid X \in T_xM,\ F(X) = 1 \bigr\}.$$

The number $\lambda = \lambda(M,F) = \sup \bigl\{ \lambda(x) \mid x \in M \bigr\}$ is called the *reversibility* of the Finsler manifold (M, F), if it exists — for example, if M is compact.

One has to show that the function λ is continuous, which is done using a standard argument: The subspace $T^1_xM = \{X \in T_xM \mid F(X) = 1\}$ is called the *unit sphere* or *indicatrix* at the point x. It is a compact space diffeomorphic to the sphere S^{n-1}. The subspaces T^1_xM, $x \in M$, form a sphere bundle over M. The function $X \in T^1_xM \mapsto F(-X) \in \mathbb{R}$ is continuous, therefore the supremum is actually the maximum of this function. The sphere bundle $T^1M \to M$ is locally trivial, that is, for small open sets $U \subset M$ the restriction $T^1U \to U$ can be identified via a fiber-preserving diffeomorphism with the canonical projection $U \times S^{n-1} \to U$ on the first factor. The function $T^1U \cong U \times S^{n-1} \mapsto F(-X) \in \mathbb{R}$ is continuous. A proof by contradiction, using the sequential compactness of S^{n-1}, then shows that the reversibility function $\lambda : M \to \mathbb{R}$ is continuous. If M is compact, the function is bounded and the supremum is actually a maximum.

Since F is positively homogeneous we could also write

$$\lambda(x) = \max\left\{ \frac{F(-X)}{F(X)} \;\Big|\; X \in T_xM,\ X \neq 0 \right\}.$$

We call a Finsler metric *reversible* if $F(-X) = F(X)$ for all $X \in TM$. Then obviously $\lambda = 1$. If there is a tangent vector X such that $F(-X) \neq F(X)$ then

$$\lambda \geq \max\left\{ \frac{F(-X)}{F(X)}, \frac{F(X)}{F(-X)} \right\} > 1.$$

Hence a Finsler metric is reversible if and only if $\lambda = 1$. In this case the *indicatrix* T_p^1M is symmetric with respect to reflection $X \mapsto -X$. Sometimes a reversible metric is also called *symmetric*, but this terminology conflicts with other notions such as symmetric quadratic forms and symmetric spaces.

For an arbitrary Finsler metric on a compact manifold M we obtain

$$\lambda^{-1}F(X) \leq F(-X) \leq \lambda F(X).$$

If $\gamma : [0,1] \to M$ is a smooth curve on M, we define the length of γ as $L(\gamma) = \int_0^1 F(\gamma'(t))\,dt$. We also introduce $\gamma^{-1} : [0,1] \to M$, the curve γ run in reverse: $\gamma^{-1}(t) = \gamma(1-t)$. The lengths of γ and γ^{-1} satisfy

$$\frac{1}{\lambda}L(\gamma) \leq L(\gamma^{-1}) \leq \lambda L(\gamma). \tag{3–1}$$

EXAMPLE 3.8. Let $F(X) = \sqrt{\alpha(X,X)} + \alpha(X,\zeta)$ be a Randers metric, with vector field ζ. For fixed $x \in M$, we will find $\lambda(x)$ by looking at the quotient $F(-X)/F(X)$ on the unit ball $\{X \in T_xM \mid \alpha(X,X) = 1\}$ of α:

$$X \mapsto \frac{F(-X)}{F(X)} = \frac{1 - \alpha(X,\zeta(x))}{1 + \alpha(X,\zeta(x))}.$$

This quotient attains its maximum for $X = -\zeta(x)/\|\zeta(x)\|$, and we obtain

$$\lambda(x) = \frac{1 + \|\zeta(x)\|}{1 - \|\zeta(x)\|}.$$

For a nonzero tangent vector $V \in T_xM$, we define on T_xM the trilinear form

$$\begin{aligned}\langle X_1, X_2, X_3\rangle_V &:= \tfrac{1}{4}\overline{X}_1\overline{X}_2\overline{X}_3F^2(V)\\ &= \frac{1}{4}\frac{\partial^3}{\partial s_1 \partial s_2 \partial s_3}\Big|_{(s_1,s_2,s_3)=(0,0,0)} F^2\Big(V + \sum_{i=1}^{3} s_iX_i\Big).\end{aligned}$$

For a given everywhere nonzero vector field V defined on an open subset $U \subset M$, we obtain a symmetric $(0,3)$-tensor, called the *Cartan tensor*; its coefficients are usually denoted C_{ijk} (that is, the functions C_{ijk} at $V \in T_xM$ express the trilinear form $\langle\ ,\ ,\ \rangle_V$ in a given system of canonical coordinates).

The Cartan tensor vanishes if and only if the Finsler metric comes from a Riemannian metric g (meaning that $F^2(X) = g(X,X)$). Euler's theorem implies that

$$\langle V,X,Y\rangle_V = \langle X,V,Y\rangle_V = \langle X,Y,V\rangle_V = 0 \tag{3–2}$$

for all vector fields X, Y.

A *distance function* on a differentiable manifold M is a smooth function $\theta : M \times M \to [0,\infty)$ such that $\theta(p,q) = 0$ if and only if $p = q$, and such that the triangle inequality is satisfied:

$$\theta(p,q) \le \theta(p,r) + \theta(r,q) \quad \text{for all } p,q,r \in M.$$

Given a connected Finsler manifold (M,F), the *induced distance* $\theta : M\times M \to \mathbb{R}$ is defined by

$$\theta(p,q) := \inf\{L(c) \mid c : [0,1] \to M \text{ piecewise smooth, } c(0) = p,\ c(1) = q\}.$$

(*Piecewise smooth* means that c is continuous and there is a finite partition of the interval $[0,1]$ such the restriction of c to each closed subinterval is smooth.) It is easy to check that θ is a distance function. If the Finsler metric is nonreversible, the induced metric is not symmetric: there are points p, q with $\theta(p,q) \ne \theta(q,p)$.

Lemma 3.9. *The induced distance of a Finsler manifold (M,F) with reversibility $\lambda \ge 1$ satisfies*

$$\frac{1}{\lambda}\theta(p,q) \le \theta(q,p) \le \lambda\theta(p,q) \tag{3–3}$$

Proof. For every $k \in \mathbb{N}$, let $\gamma_k : [0,1] \to M$ be a piecewise smooth curve with $p = \gamma_k(0)$, $q = \gamma_k(1)$ and $L(\gamma_k) \le \theta(p,q) + 1/k$. Then (3–1) gives $L(\gamma_k^{-1}) \le \lambda\theta(p,q) + \lambda/k$ for all k. Thus $\theta(q,p) \le L(\gamma_k) \le \lambda\big(\theta(p,q) + 1/k\big)$ for all k. □

Given a Finsler manifold (M,F), the *symmetrized distance* $d : M \times M \to \mathbb{R}$ is defined by $d(p,q) = \frac{1}{2}\big(\theta(p,q) + \theta(q,p)\big)$. The distance functions θ and d of a Finsler manifold coincide if and only if the Finsler metric is reversible.

For U an open subset of a manifold M, recall that $\mathcal{V}U$ is the space of smooth vector fields on U, and let $\mathcal{V}U^+ \subset \mathcal{V}U$ be the subset of nowhere vanishing vector fields. For the next theorem we recall the definition of an *affine connection*: a map $\nabla^V : (X,Y) \in \mathcal{V}U \times \mathcal{V}U \mapsto \nabla^V_X Y \in \mathcal{V}U$, linear in Y and satisfying

$$\nabla^V_X(fY) = f\nabla^V_X Y + X(f)Y \quad\text{and}\quad \nabla^V_{fX}Y = f\nabla_X Y \quad \text{for all } f \in \mathcal{F}U,\ X,Y \in \mathcal{V}U.$$

Theorem 3.10. *Let (M,F) be a Finsler manifold and $U \subset X$ an open subset. There is a map*

$$\nabla : (V,X,Y) \in \mathcal{V}U^+ \times \mathcal{V}U \times \mathcal{V}U \mapsto \nabla^V_X Y \in \mathcal{V}U$$

with the following properties:

(a) *for every* $V \in \mathcal{V}U^+$, *the map* $\nabla^V : (X,Y) \in \mathcal{V}U \times \mathcal{V}U \mapsto \nabla^V_X Y \in \mathcal{V}U$ *is an affine connection.*

(b) ∇^V *is torsionfree*, *that is*,

$$\nabla^V_X Y - \nabla^V_Y X = [X,Y] \quad \text{for all } X, Y \in \mathcal{V}U. \tag{3–4}$$

(c) ∇^V *is almost metric*, *that is*,

$$X\big(\langle Y,Z\rangle_V\big) = \langle \nabla^V_X Y, Z\rangle_V + \langle Y, \nabla^V_X Z\rangle_V + 2\langle \nabla^V_X V, Y, Z\rangle_V. \tag{3–5}$$

Moreover we have

$$\begin{aligned} 2\langle \nabla^V_X Y, Z\rangle_V = {} & X\big(\langle Y,Z\rangle_V\big) + Y\big(\langle Z,X\rangle_V\big) - Z\big(\langle X,Y\rangle_V\big) \\ & + \big\langle [X,Y],Z\big\rangle_V - \big\langle [Y,Z],X\big\rangle_V + \big\langle [Z,X],Y\big\rangle_V \\ & - 2\langle \nabla^V_X V, Y, Z\rangle_V - 2\langle \nabla^V_Y V, Z, X\rangle_V + 2\langle \nabla^V_Z V, X, Y\rangle_V \end{aligned} \tag{3–6}$$

for all vector fields $X, Y, Z \in \mathcal{V}U$, *and this equation*, *called the generalized Koszul formula*, *uniquely determines* ∇.

Sketch of proof. From the requirements (3–4) and (3–5) we obtain, through straightforward calculations, the generalized Koszul formula (3–6). Equations (3–5), (3–6) and (3–2) then imply that

$$\langle \nabla^V_V V, Z\rangle_V = 2V\big(\langle V,Z\rangle_V\big) - Z\big(\langle V,V\rangle_V\big) + 2\big\langle [Z,V],V\big\rangle_V$$

and

$$\begin{aligned} 2\langle \nabla^V_X V, Z\rangle_V = {} & X\big(\langle V,Z\rangle_V\big) + V\big(\langle Z,X\rangle_V\big) - Z\big(\langle X,V\rangle_V\big) \\ & + \big\langle [X,V],Z\big\rangle_V - \big\langle [V,Z],X\big\rangle_V + \big\langle [Z,X],V\big\rangle_V - 2\langle \nabla^V_V V, Z, X\rangle_V. \end{aligned} \tag{3–7}$$

Thus the right-hand side of (3–6) can be expanded into an expression devoid of any reference to ∇, showing that $\nabla^V_X Y$ is uniquely determined. Then one has to check that the ∇^V thus defined is in fact an affine connection, torsionfree and almost metric. □

Remark 3.11. (a) In the Riemannian case the connection $\nabla^V = \nabla$ is independent of $V \in \mathcal{V}U$, it is *metric*, meaning that $X\big(\langle Y,Z\rangle\big) = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$, and it is determined by the *Koszul formula*:

$$\begin{aligned} 2\langle \nabla_X Y, Z\rangle = {} & X\langle Y,Z\rangle + Y\langle Z,X\rangle - Z\langle X,Y\rangle \\ & + \big\langle [X,Y],Z\big\rangle - \big\langle [Y,Z],X\big\rangle + \big\langle [Z,X],Y\big\rangle. \end{aligned} \tag{3–8}$$

∇ is called the *Levi-Civita connection* or *canonical connection*.

(b) We point out the correspondence between the development adopted above (which one can find in [Matthias 1980, Chapter 2]) and the description given in [Bao et al. 2000, Chapter 2] and [Shen 2001a, § 5.2]. In canonical coordinates

$(x,y)=(x^1,\dots,x^n,y^1,\dots,y^n)$ of the tangent bundle TU of an open set $U\subset M$, we obtain the functions

$$g_{ij}:(x,y)\in TU\mapsto g_{ij}(x,y)=g(x,y)\Big(\frac{\partial}{\partial x^i},\frac{\partial}{\partial x^j}\Big)=\frac{1}{2}\frac{\partial^2}{\partial y^i\partial y^j}F^2(x,y).$$

The coefficients $C_{ijk}=C_{ijk}(x,y)$ of the Cartan tensor are

$$C_{ijk}(x,y)=\frac{1}{4}\frac{\partial^2}{\partial y^i\partial y^j}F^2(x,y)=\Big\langle\frac{\partial}{\partial x^i},\frac{\partial}{\partial x^j},\frac{\partial}{\partial x^k}\Big\rangle.$$

We also define $g^{ij}(x,y)$ as the coefficients of the inverse matrix of $g_{ij}(x,y)$. Then one can define *formal Christoffel symbols* $\gamma^i_{jk}:TU\to\mathbb{R}$:

$$\gamma^i_{jk}(x,y)=\tfrac{1}{2}g^{il}(x,y)\Big(\frac{\partial g_{lj}}{\partial x^k}(x,y)-\frac{\partial g_{jk}}{\partial x^l}(x,y)+\frac{\partial g_{kl}}{\partial x^j}(x,y)\Big).$$

We can raise and lower indices by contracting with the coefficients g^{ij} and g_{ij}; for example, $C^i_{jk}=g^{il}C_{ljk}$. Here we use the Einstein summation convention. Then we define the quantities

$$N^i_j=N^i_j(x,y)=\gamma^i_{jk}(x,y)y^k-C^i_{jk}(x,y)\gamma^k_{rs}y^ry^s.$$

It turns out that the coefficients of the Chern connection are

$$\Gamma^i_{jk}(x,y)=\gamma^i_{jk}-g^{li}\,(C_{ijr}N^r_k-C_{jkr}N^r_i+C_{kir}N^r_j),$$

(see [Bao et al. 2000, (2.4.9)]), the Chern connection is given by

$$\nabla_{\partial/\partial x^i}\frac{\partial}{\partial x^j}(x,y)=\Gamma^k_{ij}(x,y)\frac{\partial}{\partial x^k},$$

and one shows that

$$N^j_i(x,y)=\Gamma^j_{ik}(x,y)y^k.$$

The Chern connection is torsionfree, which implies that $\Gamma^k_{ij}=\Gamma^i_{jk}$. It is also almost metric, which implies (see [Shen 2001a, (5.22), (5.29)])

$$\begin{aligned}\frac{\partial g_{jl}}{\partial x^m}(x,y)&=g_{kl}\Gamma^i_{jm}+g_{kj}\Gamma^k_{lm}+2C_{jkl}N^k_m\\&=g_{kl}\Gamma^i_{jm}+g_{kj}\Gamma^k_{lm}+2C_{jkl}\Gamma^k_{mr}y^r.\end{aligned}$$

In particular:

LEMMA 3.12. *For two vector fields $V,W\in\mathcal{V}U^+$ and a point $p\in U$ with $V(p)=W(p)$ and for arbitrary vector fields $X,Y\in\mathcal{V}U$ we have*:

$$\nabla^V_XY(p)=\nabla^W_XY(p).$$

Using the connection ∇^V we introduce the *covariant derivative* ∇^V/dt along a curve $c:[a,b]\to M$. For a vector field X along the curve c with tangent vector field c' define $(\nabla^V/dt)X(t)=\nabla^V_{c'}X(t)$, where on the right-hand side one has to take extensions of the vector fields V,X,c' onto an open subset containing the curve. This expression is independent of the chosen extensions. If the vector fields V,c' coincide, we also write simply $(\nabla^V/dt)X=(\nabla/dt)X$.

For a differentiable map $H:[0,1]\times[0,1]\to M$ and a vector field $X(s,t)$ along F (meaning that $X(s,t)\in T_{H(s,t)}M$), we define $(\nabla^V/\partial t)X(t)$ as a vector field along the curve $t\mapsto H(s_1,t)$ for a fixed s_1 and $(\nabla^V/\partial s)X(t)$ as a vector field along the curve $s\mapsto H(s,t_1)$ for a fixed t_1. Then we obtain the following rule for exchanging the order of differentiation:

$$\frac{\nabla^V}{\partial t}\frac{\partial H}{\partial s}=\frac{\nabla^V}{\partial s}\frac{\partial H}{\partial t}$$

This rule follows since the connection ∇^V is torsionfree.

4. First Variation of the Energy Functional

In the Morse-theoretic proof of the Sphere Theorem we use the energy functional E on a suitable space of curves as the Morse function. For a smooth curve $c:[0,1]\to M$, the *energy* is defined as

$$E(c)=\frac{1}{2}\int_0^1 F^2\big(c'(t)\big)\,dt.$$

For a variation c_s in the first variation formula one studies the first derivative $(d/ds)|_{s=0}E(c_s)$:

LEMMA 4.1 (FIRST VARIATIONAL FORMULA). *If $c_s:[a,b]\to M$, for $s\in(-\varepsilon,\varepsilon)$, is a smooth variation of the curve $c=c_0$ with variation vector field $V(t)=(\partial/\partial s)|_{s=0}c_s(t)$, then*

$$\frac{d}{ds}\Big|_{s=0}E(c_s)=\big\langle c'(b),V(b)\big\rangle_{c'(b)}-\big\langle c'(a),V(a)\big\rangle_{c'(a)}-\int_a^b\Big\langle\frac{\nabla}{dt}c',V\Big\rangle_{c'}\,dt. \tag{4–1}$$

PROOF.

$$\frac{1}{2}\frac{\partial}{\partial s}\langle c_s',c_s'\rangle_{c_s'}=\Big\langle\frac{\nabla^{c_s'}}{\partial s}c_s',c_s'\Big\rangle_{c_s'}+\underbrace{\Big\langle\frac{\nabla^{c_s'}}{\partial s}c_s',c_s',c_s'\Big\rangle_{c_s'}}_{=0}=\Big\langle(\nabla/\partial t)\frac{\partial c_s}{\partial s},c_s'\Big\rangle_{c_s'}$$

$$=\frac{\partial}{\partial t}\Big\langle\frac{\partial c_s}{\partial s},c_s'\Big\rangle_{c_s'}-\Big\langle\frac{\partial c_s}{\partial s},(\nabla/\partial t)c_s'\Big\rangle_{c_s'}-\underbrace{\Big\langle\frac{\nabla^{c_s'}}{\partial s}c_s',\frac{\partial c_s}{\partial s},c_s'\Big\rangle_{c_s'}}_{=0}$$

Hence we conclude that

$$\frac{d}{ds}\Big|_{s=0}\int_a^b\langle c_s',c_s'\rangle_{c_s'}\,dt=\big\langle V(t),c'(t)\big\rangle_{c'}\Big|_a^b-\int_a^b\big\langle V(t),(\nabla/dt)c'\big\rangle_{c'}\,dt. \qquad\square$$

Corollary 4.2. *If $c : [0,1] \to M$ is a piecewise smooth curve such that no other piecewise smooth curve joining $p = c(0)$ and $q = c(1)$ is shorter, then c is a geodesic, that is, c is smooth and $(\nabla/dt)c' = 0$.*

Proof. Let c be smooth when restricted to each subinterval $[t_j, t_{j+1}]$ of a partition $0 = t_0 < t_1 < \cdots < t_k = 1$ of $[0,1]$. We first want to prove that these restrictions are geodesics, so that c is a *broken geodesic* (also known as a *geodesic polygon*). If $(\nabla/dt)c'(s) \neq 0$ for some $s \in (t_j, t_{j+1})$, we choose a vector field $V(t)$ along the image of c as follows:

$$V(t) = \phi(t)\frac{\nabla}{dt}c'(t),$$

where $\phi : [0,1] \to [0,1]$ is a smooth function with $\phi(s) = 1$ and $\phi(t) = 0$ for $t \notin (t_j, t_{j+1})$. Then we take a smooth variation $c_s : [0,1] \to M$ of the piecewise smooth curve c with variation vector field $V(t) = (\partial/\partial s)|_{s=0} c_s(t)$. (Saying that the family c_s is a *smooth variation of* c is saying that, for each $j = 0, \ldots, k-1$, the restrictions $c_s|[t_j, t_{j+1}]$ form a smooth variation of $c|[t_j, t_{j+1}]$.) Then the first variation formula (Lemma 4.1) gives

$$0 = \frac{d}{ds}\Big|_{s=0} E(c_s) = \int_0^1 \Big\langle \frac{\nabla}{dt}c', V\Big\rangle_{\dot c}\, dt = \int_0^1 \phi(t)\Big\|\frac{\nabla}{dt}c'\Big\|^2 dt.$$

Since $\phi(t) \geq 0$ for all t and $\phi(s) = 1$, the right-hand side is positive, a contradiction. Hence no such s exists, and c is a broken geodesic.

Now fix $l \in \{1, 2, \ldots, k-1\}$ and choose any tangent vector $V_0 \in T_{c(t_l)}M$ and a variation vector field $V = V(t)$ along the broken geodesic c such that $V(t_l) = V_0$ and $V(t_j) = 0$ for $j \neq l$. Again, Lemma 4.1 shows that

$$0 = \frac{d}{ds}\Big|_{s=0} E(c_s) = \big\langle c'(t_l^+), V_0\big\rangle_{c'(t_l^+)} - \big\langle c'(t_l^-), V_0\big\rangle_{c'(t_l^-)}$$

for all $V_0 \in T_{c(t_l)}M$. (Here, as usual, $c'(t^{\pm}) = \lim_{\varepsilon\to 0,\, \varepsilon>0} c'(t \pm \varepsilon)$.) We conclude that $c'(t_l^+) = c'(t_l^-)$, since the Legendre transformation is an isomorphism (Lemma 3.5). Hence c is a smooth curve. □

A vector field $V \in \mathcal{V}U$ is called a *geodesic vector field* if $\nabla_V^V V = 0$, which says that the flow lines of V are geodesics of the Finsler metric. These lines can also be seen as geodesics of an associated Riemannian metric:

Lemma 4.3. *Let V be a nowhere vanishing geodesic field defined on an open subset $U \subset M$. Denote by $\overline{\nabla}$ the Levi-Civita connection of the Riemannian manifold (U, g_V). Then*

$$\nabla_X^V V = \overline{\nabla}_X V$$

for all vector fields X; in particular, the vector field V is also geodesic for the Riemannian manifold (U, g_V).

PROOF. $\overline{\overline{\nabla}}$ is uniquely determined by the Koszul formula (see (3–8), with $\overline{\overline{\nabla}}$ playing the role of ∇). Since $\nabla^V_V V = 0$ we conclude from Equation 3–7 that

$$\langle \overline{\overline{\nabla}}_X V, Z\rangle_V = \langle \nabla^V_X V, Z\rangle_V \tag{4–2}$$

for all vector fields X, Y. □

We obtain a similar statement if we restrict to vector fields along a given geodesic:

LEMMA 4.4. *Let $c : [0,1] \to M$ be a non-self-intersecting geodesic of the Finsler manifold (M,F), and $V \in \mathcal{V}U^+$ an extension of the velocity vector field c' onto an open neighborhood U of $c([0,1])$. We call the Riemannian manifold (U, g_V) an osculating Riemannian manifold, denote its Levi-Civita connection by $\overline{\overline{\nabla}}$ and the covariant derivative along c by $(\overline{\overline{\nabla}}/dt)$. Then c is also a geodesic of the osculating Riemannian metric g_V and*

$$\frac{\nabla}{dt} X(t) = \frac{\overline{\overline{\nabla}}}{dt} X(t)$$

for any vector field X along c.

PROOF. As in the proof of Lemma 4.3 we show that (4–2) holds along the given geodesic:

$$\langle \overline{\overline{\nabla}}_X V, Z\rangle_V(c(t)) = \langle \nabla^V_X V, Z\rangle_V(c(t)).$$

Then

$$\begin{aligned}(\nabla^V/dt)_{c'} X(t) &= \nabla^V_{c'} X(t) = \nabla^V_X c'(t) + [c', X] \\ &= \overline{\overline{\nabla}}_X c' + [c', X] = \nabla^V_{c'} X(t) = (\overline{\overline{\nabla}}/dt) X(t).\end{aligned}$$

For $X = c'$ it follows that c is also a geodesic of the osculating Riemannian metric. □

5. Flag curvature, Jacobi Fields and Conjugate Points

For a Riemannian manifold (M,g) with Levi-Civita connection ∇, the *Riemann curvature tensor* is a $(1,3)$-tensor defined by

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z, \quad \text{for} X, Y, Z \in \mathcal{V}M.$$

It is determined by the *Jacobi operators* or *directional curvature operators* R^X, $X \in T_xM$, given by

$$Y \in T_xM \mapsto R^X(Y) := R(Y,X)X.$$

The *sectional curvature* $K(\sigma) = K(X,Y)$ of a plane $\sigma \subset T_xM$ spanned by the tangent vectors X, Y is defined by

$$K(X,Y) = \frac{\big\langle R(X,Y)Y, X\big\rangle}{|X|^2|Y|^2 - \langle X,Y\rangle^2} = \frac{\big\langle R^X(Y), Y\big\rangle}{|X|^2|Y|^2 - \langle X,Y\rangle^2}.$$

The Finsler geometry counterparts of these entities can be introduced by considering the osculating Riemannian metric. In the next statement we make use of the notion of a *geodesic variation* of a geodesic c; this is simply a variation c_s of c such that each curve $c_s : [0,1] \to M$ is geodesic.

PROPOSITION 5.1. (a) [Shen 2001b, Lemma 8.1.1] *For every nonzero tangent vector $X \in T_xM$ on a Finsler manifold M with geodesic $c = c_X : [0,1] \to M$ and $c'_X(0) = X$, the map $R^X : T_xM \to T_xM$ given by*

$$R^X(Y(0)) = -\frac{\nabla^2}{dt^2} Y(0),$$

where $Y(t) = (\partial/\partial s)|_{s=0}\, c_s(t)$ is the variation vector field of a geodesic variation of c, is well-defined and linear. It is called the Riemann curvature (operator) of the Finsler manifold.

(b) [Shen 2001b, Proposition 8.4.1] *For every nonzero tangent vector $X \in T_xM$ with a nonzero geodesic vector field V extending $X \in T_xM$ in an open neighborhood U of x, the Riemann curvature operator R^X of the Finsler manifold coincides with the Jacobi operator $\bar{R}^X$ of the osculating Riemannian metric $\bar{g} = g^V$ defined on U.*

We can now introduce the Finsler counterpart of the sectional curvature. In contrast with the Riemannian case, the notion depends not only on the choice of a two-dimensional tangent plane but also on a direction in this plane.

DEFINITION 5.2. For a Finsler manifold (M, F) and a *flag* (X, σ) consisting of a nonzero tangent vector $X \in T_xM$ and a plane $\sigma \subset T_xM$ spanned by the tangent vectors X, Y, the *flag curvature* is defined as

$$K(X;\sigma) = K(X;Y) = \frac{\langle R^X(Y), Y\rangle_X}{|X|_X^2 |Y|_X^2 - \langle X, Y\rangle_X^2}.$$

The notation $\delta < K \le 1$, where $\delta \in (0,1)$, will be used often; it is a shorthand for the condition $\delta < K(X;\sigma) \le 1$ for all flags $(X;\sigma)$ in the tangent bundle. Given a nonzero vector field V on a Finsler manifold (M, F) with connection ∇^V (see Theorem 3.10), one can consider the curvature tensor R^V defined by

$$R^V(X,Y)Z = \nabla^V_X \nabla^V_Y Z - \nabla^V_Y \nabla^V_X Z - \nabla^V_{[X,Y]} Z.$$

If the vector field V is geodesic, it follows from the definition of the Riemannian curvature that

$$R^V(Y) = R^V(Y,V)V = -\nabla^V_V \nabla^V_Y V - \nabla^V_{[Y,V]} V.$$

As in the Riemannian case, the flag curvature geometrically controls the infinitesimal behavior of geodesics, as described by the *Jacobi fields* along a geodesic:

DEFINITION 5.3. On a Finsler manifold (M, F) we call a vector field $Y = Y(t)$ along a geodesic $c : [0,1] \to M$ a *Jacobi field* if it satisfies the differential equation

$$\frac{\nabla^2}{dt^2} Y(t) + R^{c'}(Y) = 0.$$

It follows from Proposition 5.1 that the Jacobi fields of an osculating Riemannian metric (U, g_V) along the geodesic coincide with the Jacobi fields of the Finsler metric along c. Therefore the following well-known facts of Riemannian geometry (see [Klingenberg 1995, 1.12], for example) carry over to the Finsler case immediately:

LEMMA 5.4. *Let (M, F) be a Finsler manifold and $c : [0, 1] \to M$ a geodesic.*

(a) *For any $Y_0, Y_1 \in T_{c(0)}M$, there is a uniquely determined Jacobi field Y along c with initial conditions $Y(0) = Y_0$ and $(\nabla/dt)Y(0) = Y_1$.*
(b) *If, in addition, $\langle Y_0, c'(0)\rangle_{c'(0)} = 0 = \langle Y_1, c'(0)\rangle_{c'(0)}$, the Jacobi field Y thus defined satisfies $\langle Y(t), c'(t)\rangle_{c'(t)} = 0$ for all $t \in [0, 1]$.*

The Jacobi equation of Definition 5.3 is the linearization of the geodesic equation:

LEMMA 5.5. *Let (M, F) be a Finsler manifold.*

(a) *The variation vector field $V(t) = (\partial/\partial s)|_{s=0} c_s(t)$ of a geodesic variation $c_s : [0, 1] \to M$ on M is a Jacobi field.*
(b) *For every Jacobi field $Y = Y(t)$ along the geodesic $c : [0, 1] \to M$ there is a geodesic variation $c_s : [0, 1] \to M$ whose variation vector field coincides with Y.*

The proof from the Riemannian case carries over; see [Klingenberg 1995, 1.12.4].

DEFINITION 5.6. Let (M, F) be a Finsler manifold and X a unit tangent vector in TM. Let the geodesic parametrized by arc length with initial direction X be defined (at least) in the closed interval $[0, a]$, and denote it by $c : [0, a] \to M$, so that $c'(0) = X$. Suppose that for some $s \in (0, a)$ there is a nontrivial Jacobi field $Y = Y(t)$ along c that vanishes for $t = 0$ and $t = s$. (Nontrivial means that $(\nabla/dt)Y(0) \neq 0$.) Then the point $c(s)$ is called *conjugate* to $p = c(0)$ along c. Moreover, Y can be chosen so that $\langle Y, c'\rangle_{c'} = 0$, and the set of all Y satisfying all these conditions is a vector space whose dimension is called the *multiplicity* of the conjugate point $c(s)$. We define $\mathrm{conj}_X \in (0, \infty]$ as the smallest positive number r such that $c(r)$ is conjugate to p along c. The point $c(r)$ is called the *first conjugate point* to p along c. The *conjugate locus* is the set of all first conjugate points to p. The *conjugate radius* conj_p of a point $p \in M$ is the infimum of the set $\{\mathrm{conj}_X \mid X \in T_pM,\ F(X) = 1\}$.

The conjugate locus of p consists of critical points of the exponential map (see Section 8). The function $X \in T_p^1M \mapsto \mathrm{conj}_X \in \mathbb{R}^+ \cup \{\infty\}$ is continuous. We denote by $\mathrm{conj} := \inf\{\mathrm{conj}_p \mid p \in M\}$ the *conjugate radius of M*. If M is compact this is a positive real number or ∞.

REMARK 5.7. In the case of constant flag curvature one can describe Jacobi fields explicitly. Let $c : \mathbb{R} \to M$ be a geodesic parametrized by arc length on a Finsler manifold (M, F) and assume that the flag curvature along c is a constant δ, meaning that $K(c'(t); V) = \delta$ for every t and every $V \in T_{c(t)}M$ forming a flag with $c'(t)$. One can choose an orthonormal basis $(e_1, e_2, \dots, e_n)$ of the tangent space $T_{c(0)}M$ with respect to the metric $\langle\,\cdot\,,\cdot\,\rangle_{c'}$ with $e_n = c'(0)$. Using parallel transport defined by the covariant derivative ∇/dt along c we obtain a frame $(e_1(t), e_2(t), \dots, e_n(t))$ along $c = c(t)$ orthonormal with respect to $\langle\,\cdot\,,\cdot\,\rangle_{c'}$ and satisfying $e_1(t) = c'(t)$ for all $t \in \mathbb{R}$. Then with $Y(t) = \sum_{i=2}^n y_i(t) e_i(t)$ the Jacobi equation for a Jacobi field orthogonal to c' (with respect to $\langle\,\cdot\,,\cdot\,\rangle_{c'}$) decouples because of the identities

$$\frac{\nabla^2}{dt^2} Y(t) = \sum_{i=2}^n y_i''(t) e_i(t)$$

and

$$R^{c'}(Y, c')c'(t) = \sum_{i,j=2}^n y_i \bigl\langle R^{c'}(e_i), e_j \bigr\rangle_{c'} e_j(t) = \sum_{i=2}^n y_i K(e_1; e_i) e_i(t) = \delta \sum_{i=2}^n y_i e_i(t)$$

into $n-1$ ordinary differential equations

$$y_i''(t) + \delta y_i(t) = 0, \quad i = 2, \dots, n.$$

The solutions $y'' + \delta y = 0$, $y(0) = 0$, $y'(0) = 1$ are

$$y_\delta(t) = \begin{cases} 1/\sqrt{\delta}\,\sin(\sqrt{\delta}\,t) & \text{if } \delta > 0, \\ t & \text{if } \delta = 0, \\ 1/\sqrt{-\delta}\,\sinh(\sqrt{-\delta}\,t) & \text{if } \delta < 0. \end{cases}$$

Hence in the case of constant flag curvature $K(c(t)) = \delta$ along a geodesic c, we obtain for $\operatorname{conj}_{c'(0)}$ the value $\pi/\sqrt{\delta}$ if $\delta > 0$, and ∞ if $\delta \le 0$.

6. Second Variation of the Energy Functional

The first variational formula shows that geodesics can be seen as critical points of the energy functional. Therefore it is natural to study the second-order behavior of the energy functional at a geodesic. This leads to the second variational formula.

For a piecewise smooth curve $c : [0,1] \to M$ denote by W_c the set of piecewise smooth vector fields along c. The *index form* of a geodesic $c : [0,1] \to M$ is the symmetric bilinear form $I_c : W_c \times W_c \to \mathbb{R}$ defined by

$$I_c(X, Y) := \int_0^1 \Bigl(\Bigl\langle \frac{\nabla}{dt} X, \frac{\nabla}{dt} Y \Bigr\rangle_{c'}(t) - \langle R^{c'}(X), c' \rangle_{c'}(t) \Bigr)\, dt.$$

Lemma 6.1 [Shen 2001a, § 10.2; Shen 2001b, § 8.5]. *Let $c_s : [0,1] \to M$, for $s \in (-\varepsilon, \varepsilon)$, be a variation of the geodesic $c = c_0$ with fixed end points $c_s(0) = c(0)$, $c_s(1) = c(1)$; or let $c_s : S^1 \to M$ be a variation of the closed geodesic $c = c_0$ by closed curves. Let the variation vector field be $V(t) = (\partial c_s/\partial s)|_{s=0} c_s(t)$. Then*

$$\frac{d^2}{ds^2}\Big|_{s=0} E(c_s) = I_c(V, V).$$

For a geodesic $c : [0,1] \to M$ on a Finsler manifold (M, F) with an osculating Riemannian metric $\bar{g} = g_V$ defined in a neighborhood of $c([0,1])$, the index forms $I_c, \bar{I}_c$ with respect to the Finsler metric and with respect to the osculating Riemannian metric coincide.

We introduce some subspaces of the space W_c of vector fields along c defined at the beginning of this section. W_c^0 denotes the subspace consisting of vector fields that vanish at the end points. If c is a closed geodesic, also known as a *periodic geodesic* (that is, $c(0) = c(1)$ and $c'(0) = c'(1)$), we denote by W_c^1 the subspace of W_c consisting of periodic vector fields ($X(0) = X(1)$). Note that a closed geodesic can be though of as having domain $S^1 = [0,1]/\{0,1\}$.

Using Lemma 6.1 we obtain:

COROLLARY 6.2. (a) *For points $p, q \in M$ let $c : [0,1] \to M$ be a geodesic joining $p = c(0)$ and $q = c(1)$. The restriction of the index form of c to W_c^0 is denoted by I_c^0. If $c_s : [0,1] \to M$, $s \in (-\varepsilon, \varepsilon)$, is a piecewise smooth variation of c with variation vector field $Y \in W_c^0$ and with fixed end points $p = c_s(0)$, $q = c_s(1)$, we have*

$$\left.\frac{d^2}{ds^2}\right|_{s=0} E(c_s) = I_c^0(Y, Y).$$

(b) *Let $c : S^1 \to M$ be a closed geodesic and denote the restriction of the index form I_c to W_c^1 by I_c^1. If $c_s : S^1 \to M$, $s \in (-\varepsilon, \varepsilon)$ is a variation of c by closed curves with variation vector field Y, we have*

$$\left.\frac{d^2}{ds^2}\right|_{s=0} E(c_s) = I_c^1(Y, Y).$$

We now define important invariants of geodesics and closed geodesics.

DEFINITION 6.3. (a) The *index* $\operatorname{ind} c$ of a geodesic $c : [0,1] \to M$ joining points p and q of a Finsler manifold is by definition the same as the index $\operatorname{ind} W_c^0$ of the quadratic form

$$I_c^0 : W_c^0 \times W_c^0 \to \mathbb{R},$$

that is, the maximal dimension of a subspace on which I_c^0 is negative definite. The *nullity* $\operatorname{nul} c$ is the maximal dimension of a subspace $W' \subset W_c^0$ such that $I_c^0(X, Y) = 0$ for all $X \in W'$ and $Y \in W_c^0$.

(b) The Λ-*index* $\operatorname{ind}_\Lambda c$ of a closed geodesic $c : S^1 \to M$ on a Finsler manifold is the maximal dimension of a subspace on which the index form

$$I_c^1 : W_c^1 \times W_c^1 \to \mathbb{R}$$

is negative definite. The Λ-*nullity* $\operatorname{nul} c$ is the maximal dimension of a subspace $W'' \subset W_c^1$ such that $I_c^1(X, Y) = 0$ for all $X \in W''$ and $Y \in W_c^1$.

As in the Riemannian case, one can show that these numbers are finite:

LEMMA 6.4. *The index* $\operatorname{ind} c$ *and the nullity* $\operatorname{nul} c$ *of a geodesic c are finite. So are the Λ-index* $\operatorname{ind}_\Lambda c$ *and the Λ-nullity* $\operatorname{nul} c$ *of a closed geodesic c.*

Proof. Let $c : [0,1] \to M$ be a geodesic. Choose a partition $0 = t_0 < t_1 < \cdots < t_k = 1$ of the unit interval such that there is no pair of conjugate points in $[t_i, t_{i+1}]$. Define the following subspaces of the vector space W_c^0 of piecewise smooth vector fields along c vanishing at the endpoints:

$$J := \{X \in W_c^0 \mid X|[t_i, t_{i+1}] \text{ is a Jacobi field for } i = 0, \dots, k-1\},$$
$$H := \{X \in W_c^0 \mid X(t_i) = 0 \text{ for } i = 1, \dots, k-1\},$$
$$W_c^\perp = \{X \in W_c^0 \mid \langle X, c'\rangle_{c'} = 0 \text{ and } X|[t_i, t_{i+1}] \text{ is smooth for } i = 0, \dots, k-1\}.$$

Then, for $X, Y \in W_c^\perp$,

$$\begin{aligned} I_c(X,Y) &= \int_0^1 \Big(\Big\langle \frac{\nabla}{dt} X, \frac{\nabla}{dt} Y \Big\rangle_{c'} - \big\langle R^{c'}(X, c')c', Y \big\rangle_{c'} \Big)\, dt \\ &= -\int_0^1 \Big\langle \frac{\nabla^2}{dt^2} X + R^{c'}(X, c')c', Y \Big\rangle_{c'} dt \\ &\qquad + \sum_{i=1}^{k-1} \Big\langle \frac{\nabla}{dt} X(t_i^-) - \frac{\nabla}{dt} X(t_i^+), Y(t_i) \Big\rangle_{c'}. \end{aligned} \tag{6–1}$$

It follows that J and $H \cap W_c^\perp$ are orthogonal with respect to I_c, that is, $I_c(X,Y) = 0$ for all $X \in J$ and $Y \in H \cap W_c^\perp$. Therefore we can conclude that W_c^0 is the direct orthogonal sum $J \oplus (H \cap W_c^\perp)$ with the following argument: For every $X_i \in T_{c(t_i)}M$ and $X_{i+1} \in T_{c(t_{i+1})}M$ there is a unique Jacobi field Y along $c|[t_i, t_{i+1}]$ with $Y(t_i) = X_i$ and $Y(t_{i+1}) = X_{i+1}$, since $c|[t_i, t_{i+1}]$ is by assumption free of pairs of conjugate points. On the other hand, the index form I_c^0 is positive definite on H since there is no conjugate point $c(t^*)$, $t^* \in (t_i, t_{i+1})$, to the point $c(t_i)$ along $c|[t_i, t^*]$. This shows that the indices and nullities match:

$$\operatorname{ind} c = \operatorname{ind} I_c^0 = \operatorname{ind}(I_c^0|J), \quad \operatorname{nul} c = \operatorname{nul} I_c^0 = \operatorname{nul}(I_c^0|J). \tag{6–2}$$

Since J is finite-dimensional these invariants are finite. An analogous proof shows that the Λ-index and Λ-nullity of a closed geodesic are also finite. □

We call a geodesic c *nondegenerate* if $\operatorname{nul} c = 0$. This implies that the point $q = c(1)$ is not conjugate to $p = c(0)$ along c.

We call a closed geodesic $c : S^1 \to M$ *nondegenerate* if $\operatorname{nul}_\Lambda c = 1$. Since in this case $I_c^1(c', c') = 0$, the nullity is at least 1. (That's why some other authors define the nullity of a closed geodesic as $\operatorname{nul}_\Lambda c - 1$.) A Finsler metric all of whose closed geodesics are nondegenerate is called *bumpy*.

7. Results from Topology

Using the energy functional on the space $\Omega_{pq}M$ of curves on a Finsler manifold M joining two fixed points $p, q \in M$, we obtain a CW-decomposition of the space $\Omega_{pq}M$. The Morse indices of the geodesics in Ω_{pq} are related to the dimensions of the cells of the CW-decomposition. In this chapter we review results from the topology of CW-complexes and the relation between the topology of $\Omega_{pq}M$ and that of M. As general references we cite [Milnor 1969; Bredon 1995; Spanier 1966].

DEFINITION 7.1. Let A, X be topological spaces with $A \subset X$. We say that X is *obtained from A by adjoining k-cells e_j^k*, $j \in J_k$, if the following conditions hold:

(a) For every $j \in J_k$ the set e_j^k is a closed subset of X.
(b) Let $\dot{e}_j^k := e_j^k \cap A$. Then for all $i, j \in J_k$ with $i \neq j$ the subsets $e_i^k - \dot{e}_i^k$ and $e_j^k - \dot{e}_j^k$ are disjoint.
(c) The topology of $X = A \cup \bigcup_{j \in J_k} e_j^k$ is the *weak topology* with respect to the component subsets A and e_j^k, for $j \in J_k$. (This means that $U \subset X$ is open if and only if $U \cap A$ is open in A and $U \cap e_j^k$ is open in e_j^k for each j.)
(d) For every $j \in J_k$ there is a continuous map

$$\phi_j : (D^k, S^{k-1}) \to (e_j^k, \dot{e}_j^k)$$

with $\phi_j(D^k) = e_j^k$ such that the restriction $\phi_j : D^k - S^{k-1} \to e_j^k - \dot{e}_j^k$ is a homeomorphism, and such that a subset $U \subset e_j^k$ is open if and only if $U \cap \dot{e}_j^k \subset \dot{e}_j^k$ and $\phi_j^{-1}(U) \subset D^k$ are open subsets.

For $k = 0$ the space X is the disjoint union of the space A and a discrete space.

DEFINITION 7.2. A (*relative*) *CW-complex* (X, A) consists of a topological space X, a closed subspace $A \subset X$, and a sequence of closed subspaces $(X, A)^k \subset X$, $k \geq 0$ (where $(X, A)^k$ is called the *k-skeleton*) with $X = A \cup \bigcup_{k \geq 0} (X, A)^k$ such that the following conditions hold:

(a) $(X, A)^0$ is obtained form A by adjoining 0-cells and for every $k \geq 1$ the space $(X, A)^k$ is obtained from $(X, A)^{k-1}$ by adjoining k-cells.
(b) The topology of X is the weak topology of the union $A \cup \bigcup_{k \geq 0} (X, A)^k$.

Hence it is possible to build up a CW-complex recursively. Start with the topological space $(X, A)^{-1} := A$, and recursively assume that $(X, A)^{k-1}$ has been defined. Given continuous maps $\tilde{\phi}_j : S^{k-1} \to (X, A)^{k-1}$, $j \in J_k$ (called *attaching maps*), we form the subset $(X, A)^k$ as follows: Let $\dot{e}_j^k = \phi_j(S^{k-1})$ and $e_j^k := D^k \cup_{\phi_j} \dot{e}_j^k$, meaning that e_j^k is the quotient space of the disjoint union of D^k and $\dot{e}_j^k$ with the equivalence relation identifying each $x \in S^{k-1}$ with $\phi_j(x) \in \dot{e}_j^k$. The sets $D^k \setminus S^{k-1}$ and $e_j^k \setminus \dot{e}_j^k$ are homeomorphic and can be identified. Next define the *characteristic map* $\phi_j : (D^k, S^{k-1}) \to (e_j^k, \dot{e}_j^k)$ by $\phi_j(x) = x$, for $x \in D^k - S^{k-1}$, and $\phi_j(y) = \tilde{\phi}_j(y)$, for $y \in S^{k-1}$. Then set $(X, A)^k = (X, A)^{k-1} \cup \bigcup_{j \in J_k} e_j^k$. A subset $U \subset (X, A)^k$ is open if and only if the intersection $U \cap (X, A)^{k-1}$ is open in $(X, A)^{k-1}$ and for all $j \in J_k$ the preimages $\phi_j^{-1}(U \cap e_j^k)$ are open subsets of D^k. The subsets e_j^k, $j \in J_k$, are called the *closed k-cells* of the CW-complex. Because of possible identifications on the boundary, these cells are in general not homeomorphic to D^k. The subsets $e_j^k - \dot{e}_j^k$ are called *open cells*; they are homeomorphic to $D^k - S^{k-1}$, but in general these sets are not open subsets of the k-skeleton $(X,A)^k$.

A *subcomplex* of a CW-complex is the union of closed cells of the CW-complex with the same attaching resp. characteristic maps which is itself a CW-complex. For example, the k-skeleton $(X,A)^k$ is a subcomplex. For $A = \varnothing$ we simply write $X^k = (X, \varnothing)^k$.

Remark 7.3. The advantage of using CW-complexes is that one generally needs fewer cells to write a space as a CW-complexes than to triangulate it. A simple but important example: for $n \geq 1$ the n-dimensional sphere $S^n = \{x \in \mathbb{R}^{n+1} \mid \|x\|^2 = 1\}$ has the structure of a CW-complex with one 0-cell e^0 and one n-cell; the attaching map is the constant map $\phi : S^{n-1} \to e^0$.

CW-complexes are very useful if one considers homotopy properties of topological spaces. For example, Morse theory shows that any manifold is homotopy equivalent to a CW-complexes; see the proof of Corollary 7.7.

Two continuous maps $f, g : X \to Y$ are called *homotopic* if there is a continuous map $F : X \times [0,1] \to Y$, $(x,t) \mapsto F_t(x) = F(x,t)$, with $F_0 = f$ and $F_1 = g$. There is also a relative version of this concept: Let (X,A) and (Y,B) be pairs of topological spaces, meaning that $A \subset X$ and $B \subset Y$. A continuous map $f : (X,A) \to (Y,B)$ between these pairs is a continuous map $f : X \to Y$ with the property $f(A) \subset B$. Two continuous maps $f, g : (X,A) \to (Y,B)$ agreeing on A are *homotopic relative to* A, or *homotopic rel* A, if there is a continuous map $F : (X,A) \times [0,1] \mapsto (Y,B)$, $(x,t) \mapsto F_t(x) = F(x,t)$, with $F_0 = f$, $F_1 = g$ and $F_t|A = f|A = g|B$ for all $t \in [0,1]$. Two topological spaces X, Y are called *homotopy equivalent* if there are continuous maps $f : X \to Y$ and $g : Y \to X$ such that the compositions $f \circ g$ and $g \circ f$ are homotopic to the identity maps id_Y and id_X, respectively. Then f is called a *homotopy equivalence*. We call a pathwise connected topological space X *n-connected* for some $n \geq 1$ if for every $j \in \{1, \ldots, n\}$ every continuous map $f : S^j \to X$ is homotopic to a constant map. (This implies that the *homotopy groups* $\pi_j(X,p) = 0$ vanish for all $j = 1, \ldots, n$, but we do not need this concept here.) A 1-connected space is also called *simply connected*. A topological pair (X,A) is called *n-connected* for some $n \geq 1$ if for every $j \in \{1, \ldots, n\}$ every continuous map $f : S^j \to X$ is homotopic to a continuous map whose image lies in A.

A continuous map $f : X \to Y$ between CW-complexes is called *cellular* if it respects the CW-structure, that is, if the image $f(X^k)$ of the k-skeleton lies in the k-skeleton of Y: $f(X^k) \subset Y^k$.

Proposition 7.4 (Cellular Approximation Theorem). *Every continuous map $f : X \to Y$ between CW-complexes is homotopic to a cellular map.* (See [Spanier 1966, Theorem 7.6.17] for a proof.)

Thus every continuous map $f : S^j \to S^n$ between spheres is homotopic to a cellular map $F : S^j \to S^n$, where the spheres have the CW-structure described in Remark 7.3. Now suppose that $j < n$; since the j-skeleton of S^n equals the 0-skeleton (a single point), S^n is $(n-1)$-connected. The same argument shows:

PROPOSITION 7.5. *If a pathwise connected CW-complex X has no j-dimensional cells for any $j \in \{1, 2, \ldots, k\}$, the space X is k-connected.*

We note the following consequences of Whitehead's and Hurewicz's theorems [Bredon 1995, Theorems VII-11.2 and VII-10.7]:

PROPOSITION 7.6 [Bredon 1995, Corollary VII-11.14]. (a) *Let $f : X \to Y$ be a continuous map between two simply connected CW-complexes, such that the induced homomorphism $f_* : H_i(X; \mathbb{Z}) \to H_i(Y; \mathbb{Z})$ of the singular homology groups with integer coefficients is an isomorphism for all i. Then f is a homotopy equivalence.*

(b) *If a topological space X is $(n-1)$-connected for some $n \geq 2$, then for every homology class $h \in H_n(X; \mathbb{Z})$ there is a continuous map $f_h : S^n \to X$ and a generator $g \in H_n(S^n; \mathbb{Z})$ such that $(f_h)_*(g) = h$. (Hence every n-dimensional homology class can be represented by a spherical cycle.)*

COROLLARY 7.7. *A compact and $(n-1)$-connected differentiable manifold M is homotopy equivalent to the n-sphere.*

PROOF. We use the fact that an n-dimensional manifold has the homotopy type of a finite CW-complex X (meaning there are only finitely many cells) of dimension n (meaning the maximal cell dimension is n). This can be proved using a Morse function on M; one obtains a CW-structure where every critical point of index k corresponds to a k-dimensional cell. Since the critical points are nondegenerate and the manifold is compact, there are only finitely many of them. The index of a critical point on a differentiable manifold is bounded above by the dimension of the manifold.

The manifold is simply connected, therefore orientable. Hence $H_n(M; \mathbb{Z}) \cong \mathbb{Z}$ [Bredon 1995, Corollary VI-7.2]. It follows from Proposition 7.6(b) that there is a continuous map $f : S^n \to M$ inducing an isomorphism $f_* : H_n(S^n; \mathbb{Z}) \to H_n(M; \mathbb{Z})$. Since S^n and M^n are (up to homotopy equivalence) CW-complexes, we conclude from Proposition 7.6(a) that f is a homotopy equivalence. □

REMARK 7.8. The *Poincaré conjecture* states that for $n \geq 3$, an n-dimensional simply connected and compact manifold homotopy equivalent to the n-sphere is *homeomorphic* to the n-sphere. For $n \geq 5$ the conjecture was proven by S. Smale, for $n = 4$ by M. Freedman. There is an announcement of a proof for $n = 3$, by G. Perelman [Milnor 2003].

Given a topological space X and a point $p \in X$, we define the *loop space*

$$\Omega_p(X) = \{\gamma : [0, 1] \to X \mid \gamma \text{ continuous}, \gamma(0) = \gamma(1) = p\}.$$

Proposition 7.9. *If X is a pathwise connected topological space and $\Omega_p X$ is the loop space of X with $p \in X$, the homotopy groups satisfy*

$$\pi_{k-1}(\Omega_p X) \cong \pi_k X \quad \textit{for all } k \geq 1.$$

Proof. Consider the cube $I^k = [0,1]^k$, with boundary

$$\dot{I}^k = \{(x_1, \ldots, x_k) \in I^k \mid \text{there is } j = 1, \ldots, k \text{ with } x_j \in \{0,1\}\}$$

Given a continuous map $f : (I^{k+1}, \dot{I}^{k+1}) \to (X, p)$ for some point $p \in X$, we define $F(f) : (I^k, \dot{I}^k) \to (\Omega_p X, p)$ by

$$F(f)(x_1, x_2, \ldots, x_k)(t) = f(x_1, x_2, \ldots, x_{k+1}).$$

Also, for a continuous map $g : (I^k, \dot{I}^k) \to (\Omega X, p)$ we define a continuous map $G(g) : (I^{k+1}, \dot{I}^{k+1}) \to (X, p)$ by

$$G(g)(x_1, \ldots, x_{k+1}) = g(x_1, \ldots, x_k)(x_{k+1}).$$

This lets us define the isomorphism between the homotopy groups $\pi_k(\Omega X, p)$ and $\pi_{k+1}(X, p)$. □

Corollary 7.10. *Let M be a simply connected, compact, n-dimensional manifold and $p, q \in M$ arbitrary points. Consider the space*

$$\Omega_{pq} M := \{\gamma : [0,1] \to M \mid \gamma \textit{ piecewise smooth}, \gamma(0) = p, \gamma(1) = q\}.$$

If this space is homotopy equivalent to a CW-complex with no cells of dimension $j \in \{1, \ldots, k\}$, then M is k-connected. In particular, if $k = n-1$ the manifold is homotopy equivalent to the n-sphere.

Proof. The space $\Omega_{pq} M$ of piecewise continuous curves joining p and q is homotopy equivalent to the space $\Omega^*_{pq} M$ of continuous curves joining p and q with the compact-open topology. This can be shown by using the finite-dimensional approximation $\Omega(k, a)$ for $\Omega^a M = \{\gamma \in \Omega_{pq} M \mid E(\gamma) \leq a\}$ with a sufficiently large k [Milnor 1969, Theorem 17.1]. The next step is that the homotopy type of the spaces $\Omega_{pq} M$ does not depend on the chosen points. For two curves $\gamma_1, \gamma_2 : [0,1] \to M$ with $\gamma_2(1) = \gamma_1(0)$, denote by $\gamma_1 * \gamma_2 : [0,1] \to M$ their *composition*, defined by

$$\gamma_1 * \gamma_2(t) = \begin{cases} \gamma_2(2t) & \text{for } 0 \leq t \leq \frac{1}{2}, \\ \gamma_1(2t-1) & \text{for } \frac{1}{2} \leq t \leq 1. \end{cases}$$

Now fix a curve $\gamma_1 \in \Omega_{qr} M$; the map $\gamma \in \Omega_{pq} M \mapsto \gamma_1 * \gamma \in \Omega_r M$ defines a homotopy equivalence. Hence we can conclude from Proposition 7.5 that the loop space is k-connected. This finally implies by Proposition 7.9 that the manifold is $(k-1)$-connected. If $k = n-1$ we conclude from Corollary 7.7 that the manifold is homotopy equivalent to the n-sphere. □

8. Morse Theory of the Energy Functional

For a Finsler manifold (M, F) and two points $p, q \in M$ we consider the energy functional on the space Ω_{pq} of curves joining the points p and q. A critical value κ equals the energy $E(c)$ of a geodesic c joining these points. Morse theory provides a connection between invariants of critical points of a function on a manifold and global topological invariants, in our case a connection between homology or homotopy invariants of the loop space and indices of geodesics.

We introduce the space $\Omega_{pq}M$ of absolutely continuous curves $\gamma : [0,1] \to M$ satisfying $\gamma(0) = p$, $\gamma(1) = q$, and

$$\int_0^1 F^2(\gamma'(t))dt < \infty.$$

The energy functional $E : \Omega_{pq}M \longrightarrow \mathbb{R}$, given by

$$E(\gamma) = \frac{1}{2}\int_0^1 F^2(\gamma'(t))\,dt,$$

is $C^{1,1}$-differentiable, that is, it is C^1-differentiable and its derivative is locally Lipschitz continuous [Mercuri 1971].

If $c_s : [0,1] \to M$, $s \in (-\varepsilon, \varepsilon)$, is a variation of $c = c_0$ with fixed end points $p = c(0)$, $q = c(1)$, we conclude from the first variation formula (Lemma 4.1) that the variation vector field $Y(t) = (d/ds)|_{s=0}\, c_s(t)$ satisfies

$$dE(c)Y := \frac{\partial}{\partial s}\Big|_{s=0} E(c_s) = \int_0^1 \Big\langle \frac{\nabla}{dt} c', Y \Big\rangle_{c'}(t)\,dt.$$

The curve c is a *critical point* of the energy functional if $dE(c)Y = 0$ for every vector field $Y \in W_c^0$. Then it follows from Corollary 4.2 that the critical points are the geodesics $c : [0,1] \to M$, starting at $p = c(0)$ and ending at $q = c(1)$. For nonnegative $\kappa \geq 0$ we define the sublevel sets

$$\Omega_{pq}^{\kappa} = \Omega^{\kappa} := \{\sigma \in \Omega_{pq}M \mid E(\sigma) \leq \kappa\};$$

then $\Omega_{pq}^{l}M$ with $2l = \theta(p,q)^2$ contains the minimal geodesics joining p and q. For l small enough there is a unique minimal geodesic, that is, $\Omega_{pq}^{l}M$ contains exactly one element. If the case $p = q$ the subset Ω_{pp}^{0} consists of the point curve p.

We choose an arbitrary Riemannian metric g on the manifold, which induces a Hilbert space structure on $\Omega_{pq}M$. If $c : [0,1] \to M$ is a smooth curve with $c(0) = c(1)$ and X, Y are smooth vector fields along c, a Riemannian metric on $\Omega_{pq}M$ is defined by

$$g_1(X,Y) = \int_0^1 g\big(X(t), Y(t)\big)dt + \int_0^1 g\Big(\frac{\nabla}{dt}X(t), \frac{\nabla}{dt}Y(t)\Big)\,dt,$$

where ∇/dt is the covariant derivative along c induced by the Levi-Civita connection of the Riemannian manifold. The energy functional induces a Lipschitz continuous gradient vector field $\operatorname{grad} E$ through the equation

$$g_1\big(\operatorname{grad} E(c), X\big) = dE(c)\,X$$

for all X. The energy functional satisfies the Palais–Smale condition and the *negative gradient flow*, that is, the flow of the vector field $-\operatorname{grad} E$ on λ, is defined for every $t \geq 0$ [Mercuri 1971].

For the Morse theory of the energy functional we have to consider also the second derivatives of the energy functional at the critical points. For a non-Riemannian Finsler metric the square F^2 of the Finsler metric F is not C^2-differentiable at the zero section. Hence E is C^2-differentiable only at the regular curves, those curves c with $c'(t) \neq 0$ for all t. Geodesics of positive length are regular, so we can use the statement in Corollary 6.2. As remarked just before that corollary, the index form I_c^0 of the geodesic defined with the Finsler metric coincides with the index form $\bar{I}_c$ of an osculating Riemannian metric.

We will not go into details in this construction since instead of the space Ω_{pq} one can use a *finite-dimensional approximation*. This allows us to use Morse theory for a finite-dimensional compact manifold instead of the infinite-dimensional Hilbert manifold $\Omega_{pq}M$. The finite-dimensional approximation was introduced by M. Morse and is explained in [Milnor 1969, Chapter 16]. We start with a compact Finsler manifold (M, F) with injectivity radius $\operatorname{inj} > 0$ (see Definition 9.1 below). For every pair of points $p, q \in M$ with distance $\theta(p, q) < \operatorname{inj}$, there is a unique minimal geodesic from p to q. Choose $a > 0$ and $k \in \mathbb{N}$ such that $1/k < (\operatorname{inj})^2/(2a)$, and set $t_i := i/k$ for $i \in \mathbb{N}$. Define

$$\Omega_{pq}(k, a) := \big\{ c \in \Omega_{pq}^a M \mid c|[t_i, t_{i+1}] \text{ is a geodesic } \big\}.$$

Since

$$\theta^2\big(c(t_i), c(t_{i+1})\big) \leq L^2\big(c|[t_i, t_{i+1}]\big) \leq \frac{2}{k} E\big(c|[t_i, t_{i+1}]\big) \leq \frac{2a}{k} < \operatorname{inj}^2,$$

a curve $c \in \Omega_{pq}(k, a)$ is uniquely determined by the points $c(t_1), \ldots, c(t_{k-1}) \in M \times \cdots \times M$. Therefore we can identify $\Omega_{pq}(k, a)$ with the submanifold with boundary

$$\big\{(x_1, \ldots, x_{k-1}) \in M \times \cdots \times M \mid \theta(x_i, x_{i+1}) \leq \tfrac{1}{2} \operatorname{inj} \ \text{ for } i = 0, \ldots, k-1\big\}$$

of the product manifold $M \times \cdots \times M$, where we set $p = x_0$, $q = x_k$. We conclude that $\Omega_{pq}(k, a)$ has the structure of a compact manifold with boundary of dimension $(k-1) \dim M$. Then there is a strong deformation retraction

$$r_u : \Omega_{pq}^a \to \Omega_{pq}^a, \quad u \in [0, 1],$$

with $r_0(c) = c$ for all $c \in \Omega_{pq}^a$, $r_u(c) = c$ for all $c \in \Omega_{pq}(k, a)$, $u \in [0, 1]$, and $r_1(c) \in \Omega_{pq}(k, a)$ for all Ω_{pq}^a. It is defined for $u \in [t_i, t_{i+1}]$, as follows: For $t \leq t_i$, $r_u(c)(t)$ is the broken geodesic with corners $c(0), c(t_1), c(t_2), \ldots, c(t_i)$; $r_u(t)$ for

$t \in [t_i, u]$ is the minimal geodesic between $c(t_i)$ and $c(u)$ and $r_u(c)(t) = c(t)$ for $t \geq t_{i+1}$. Then the restriction $E' : \Omega_{pq}(k,a) \to \mathbb{R}$ of E given by

$$E'(x_1, \ldots, x_k) = \tfrac{1}{2} \textstyle\sum_{i=1}^{k} \theta^2(x_i, x_{i+1})$$

is a C^1-smooth function. The first variational formula implies that the critical points are the geodesics from p to q with $E(c) \leq a$, and the function is C^∞ differentiable in the neighborhood of critical points of positive length.

For a broken geodesic $c = (x_1, \ldots, x_k)$ (smooth except at $t = t_1, \ldots, t_{k-1}$, where t_i no longer bears the meaning i/k), a tangent vector $y(t) = (\partial/\partial s)|_{s=0}\, c_s(t)$ is given by a variation $c_s = (x_1^s, \ldots, x_k^s) \in \Omega_{pq}(k,a)$, $s \in (-\varepsilon, \varepsilon)$, that is, a curve in $\Omega_{pq}(k,a)$. Since the variation vector field of a geodesic variation is a Jacobi field (Lemma 5.5), the tangent space $T_x\Omega_{pq}$ consists of *broken Jacobi fields*,

$$T_x\Omega_{pq}(k,a) = \bigl\{X \in W_c \mid X|[t_i, t_{i+1}] \text{ is a Jacobi field}\bigr\}.$$

In particular, if c is a geodesic (smooth throughout), the tangent space $T_c\Omega_{pq}$ coincides with the space J introduced in the Proof of Lemma 6.4. Therefore the Proof of Lemma 6.4 also shows that restricting the index form to the tangent space $T_c\Omega_{pq}(k,a)$ changes neither the index nor the nullity.

The energy functional $E' : \Omega_{pq}M(k,a) \to \mathbb{R}$ is a differentiable function on the compact manifold $\Omega_{pq}M(k,a)$. It is a Morse function if all critical points are nondegenerate, that is, if all geodesics c joining the points p and q with energy $\leq a$ are nondegenerate — in symbols, $\operatorname{nul} c = 0$. Assume $c \in \Omega_{pq}(k,a)$ is degenerate, so there is $X \in W_c^0$ with $X \neq 0$ and $I_c^0(X,Y) = 0$ for all $Y \in W_c^0$. Since X is a piecewise smooth vector field, choose $0 = t_0 < t_1 < \cdots < t_k = 1$ such that X is smooth when restricted to each subinterval $[t_i, t_{i+1}]$, and also such that no subinterval contains a pair of conjugate points. Then we obtain from Equation 6–1:

$$0 = I_c(X,Y) = -\int_0^1 \Bigl\langle \frac{\nabla^2}{dt^2} X + R^{c'}(X, c')c', Y \Bigr\rangle_{c'} dt \\ + \sum_{i=1}^{k-1} \Bigl\langle \frac{\nabla}{dt} X(t_i^-) - \frac{\nabla}{dt} X(t_i^+), Y(t_i) \Bigr\rangle_{c'}. \quad (8\text{–}1)$$

Let $Y \in W_c^0$ be a broken Jacobi field with $Y(t_i) := (\nabla/dt)X(t_i^-) - (\nabla/dt)X(t_i^+)$, so the restrictions $Y|[t_i, t_{i+1}]$ are Jacobi fields along $c|[t_i, t_{i+1}]$. Then Equation 8–1 implies

$$I_c(X,Y) = \sum_{i=1}^{k-1} \Bigl\| \frac{\nabla}{dt} X(t_i^-) - \frac{\nabla}{dt} X(t_i^+) \Bigr\|_{c'}^2 = 0,$$

hence X is a smooth vector field. We have shown:

LEMMA 8.1. *The energy functional $E' : \Omega_{pq}(k,a) \to \mathbb{R}$ is a Morse function if and only if the point q is not conjugate to p along any geodesic c with $E(c) \leq a$ joining p and q.*

The next proposition relates the conjugate points of $p \in M$ to the critical points of the exponential map at p. (Recall that the *exponential map* $\exp_p : T_pM \to M$ is defined by $\exp_p(X) = c_X(1)$, where for $X \in T_pM$ we denote by $c_X : \mathbb{R} \to M$ the geodesic with $c'_X(0) = X$. This assumes that the metric is complete — more precisely, forward geodesically complete — an assumption that is satisfied in particular if the manifold is compact, thanks to the Finsler version of the Hopf–Rinow Theorem [Bao et al. 2000, § 6.6].)

PROPOSITION 8.2. *Let (M,F) be a complete Finsler manifold and let $p \in M$. A point $q = \exp_p(X)$ is a critical point of the exponential map $\exp_p : T_pM \to M$ if and only if q is a conjugate point of p along the geodesic $t \in [0,1] \mapsto \exp_p(tX) \in M$ from p to q.*

The proof of the Riemannian case [Milnor 1969, Theorem 18.1] carries over. As an application of Sard's Theorem one obtains:

COROLLARY 8.3 [Milnor 1969, Corollary 18.2]. *Let (M,F) be a compact Finsler manifold and let $p \in M$. For almost all points $q \in M$ (that is, up to a set of measure zero) the point q is not a conjugate point to p along any geodesic from p to q. For almost all $q \in M$ the energy functional $E : \Omega_{pq}M \to \mathbb{R}$ is a Morse function.*

It is the chief observation of Morse theory that the topology of the sublevel sets $\Omega^\kappa_{pq}M := \{\sigma \in \Omega_{pq} \mid E(\sigma) \leq \kappa\}$ and $\Omega^\kappa_{pq}(k,b)$ can only change if κ is a critical value. The change in topology can be described by the indices of the corresponding critical points. Applied to the energy functional, this line of argumentation yields (compare [Milnor 1969, Theorem 17.3]):

THEOREM 8.4 (FUNDAMENTAL THEOREM OF MORSE THEORY). *Let (M,F) be a compact Finsler manifold and $p \in M$ an arbitrary point. For almost all $q \in M$ and for all $a > 0$ the function $E' : \Omega_{pq}(k,a) \to \mathbb{R}$ is a Morse function and there are only finitely many geodesics c joining p and q with $E(c) \leq a$.*

The spaces $\Omega^\kappa_{pq}M$ and $\Omega^\kappa_{pq}(k,a)$ have the homotopy type of a CW-complex having as many m-cells as there are geodesics c joining p and q with $E(c) \leq a$ and $\operatorname{ind} c = m$.

SKETCH OF PROOF. As remarked in Proposition 8.2, for almost all $q \in M$ and all $a > 0$ the energy functional $E' : \Omega_{pq}(k,a) \to \mathbb{R}$ is a Morse function. If there is no critical value in $[\alpha,\beta]$, one can use the flow of the negative gradient field $-\operatorname{grad} E'$ on $\Omega_{pq}(k,a)$ and retract $\Omega^\beta_{pq}(k,a)$ onto $\Omega^\alpha_{pq}(k,a)$.

The behavior of a Morse function near a critical point is described by the Morse Lemma [Milnor 1969, Lemma 2.2]. Applied to E' it states that near a geodesic c one can introduce local coordinates $y = (y_1,\dots,y_r)$, with c corresponding to $0 = (0,\dots,0)$, such that

$$E'(y_1,\dots,y_r) = E(c) - \sum_{j=1}^{\operatorname{ind} c} y_j^2 + \sum_{j=\operatorname{ind} c+1}^{r} y_j^2.$$

Here $r = (k-1)\dim M = \dim \Omega_{pq}(k,a)$. Hence the index describes the dimension of a subspace on which the energy of nearby curves decreases quadratically, whereas on a complementary subspace the energy grows quadratically. This implies that the geodesics are isolated. Since $\Omega_{pq}(k,a)$ is compact, there are only finitely many geodesics joining p and q with energy $\le a$.

Assume for simplicity that there is only one geodesic of energy a joining p and q. Let $\operatorname{ind} c = m$. Then one can show that for sufficiently small $\varepsilon > 0$ the set $\Omega_{pq}^{a+\varepsilon}(k,a)$ has the homotopy type of $\Omega_{pq}^{a-\varepsilon}(k,a)$ with an m-dimensional cell attached [Milnor 1969, Theorem 3.2]. This m-cell corresponds to the set $\{(y_1,\dots,y_k,0,\dots,0) \mid y_1^2+\dots+y_m^2 < \varepsilon\}$ in the coordinates used in the Morse Lemma. □

REMARK 8.5. In Remark 5.7 we discussed Jacobi fields along a geodesic where the flag curvature is constant and positive. Now we consider the index form $I_c(\delta,l)$ of a geodesic $c = c_l(\delta) : [0,1] \to M$ of length l with constant flag curvature $K(c'(t);\sigma)$. We can use bounds for the flag curvature to estimate the index and the conjugate radius, as in the Riemannian case.

We choose $e_1, e_2, \dots, e_n$ in $T_{c(0)}M$, orthonormal with respect to $\langle\cdot,\cdot\rangle_{c'}$ and such that $c'(0) = F(c')e_1$. We extend this frame by parallel transport with respect to (∇/dt) along c. We can write vector fields $X = X(t)$ along c as $X(t) = \sum_{i=1}^n x_i(t)e_i(t)$, for smooth functions $x_i : [0,1] \to \mathbb{R}$. Then

$$I_c(\delta,l)(X,X) = \int_0^1 \bigl(x_i'(t)^2 - \delta x_i^2(t)\bigr)\,dt$$

and one shows that

$$\operatorname{ind} c_l(\delta) = \operatorname{ind} I_c(\delta,l) = k(n-1) \tag{8–2}$$

for $l \in \bigl(k\pi/\sqrt{\delta},\,(k+1)\pi/\sqrt{\delta}\bigr)$. See [Klingenberg 1995, Example 2.5.7].

Now let $\gamma : [0,1] \to M$ be a geodesic of a Finsler metric with a lower bound for the flag curvature: $K \ge \delta$. We again choose along γ an orthonormal frame $(e_1, e_2, \dots, e_n)(t)$ parallel with respect to (∇/dt). We can estimate the indexes $\operatorname{ind}\gamma$ and $\operatorname{ind} I_\gamma$ by comparing them with the index $\operatorname{ind} c_l(\delta)$ of a geodesic $c_l(\delta)$ of the same length on a space form with constant sectional curvature:

$$\begin{aligned} I_c(X,X) &= \int_0^1 \bigl(x_i'(t)^2 - K(e_1(t);e_2(t))x_i^2(t)\bigr)\,dt \\ &\le \int_0^1 \bigl(x_i'(t)^2 - \delta x_i^2(t)\bigr)\,dt = I_c(\delta,l)(X,X). \end{aligned}$$

This computation and a similar one in the case of an upper bound for the flag curvature lead to the following estimates for the distance of conjugate points and indices of geodesics. Here we use the fact that the index form I_γ of $\gamma = \gamma_X$ is positive definite for $L(\gamma) < \operatorname{conj}_X$ and degenerate for $L(\gamma) = \operatorname{conj}_X$.

LEMMA 8.6. *Let $\gamma = \gamma_X : [0,a] \to M$ be a geodesic parametrized by arc length on a Finsler manifold (M,F), with $\gamma_X'(0) = X$.*

(a) *If the flag curvature $K = K(\gamma';\sigma)$ satisfies $K \le \Delta$ (resp. $K < \Delta$) then* $\operatorname{conj}_X \ge \pi/\sqrt{\Delta}$ (*resp.* $\operatorname{conj}_X > \pi/\sqrt{\Delta}$).
(b) *If the flag curvature $K = K(\gamma';\sigma)$ satisfies $\delta \le K$ (resp. $\delta < K$) then* $\operatorname{conj}_X \le \pi/\sqrt{\Delta}$ (*resp.* $\operatorname{conj}_X < \pi/\sqrt{\Delta}$).
(c) *If the Ricci curvature* $\operatorname{Ric} = \operatorname{Ric}(\gamma')F^2$ *satisfies* $\operatorname{Ric} \ge \delta(n-1)F^2$ (*resp.* $\operatorname{Ric} > \delta(n-1)$) *then* $\operatorname{conj}_X \le \pi/\sqrt{\delta}$ (*resp.* $\operatorname{conj}_X < \pi/\sqrt{\delta}$).

SKETCH OF PROOF. The argument for cases (a) and (b) is given in Remark 8.5.

The argument in case (c) is the same as in the Riemannian case. As in the preceding remark we choose an orthonormal frame $(e_1,\dots,e_n)(t)$ along γ with $L(\gamma) = a = \pi/\sqrt{\delta}$, parallel with respect to ∇/dt, and such that $\gamma'(0) = X = F(X)e_1(0)$. Then we define the vector fields

$$X_i(t) = \sin(\sqrt{\delta}\,t)e_i(t), \quad i = 2,\dots,n.$$

We compute for the index form I_γ :

$$\begin{aligned}\sum_{i=2}^n I_\gamma(X_i,X_i) &= \sum_{i=2}^n \int_0^a \big(\cos^2(\sqrt{\delta}\,t) - K(e_1(t);e_i(t))\sin^2(\sqrt{\delta}\,t)\big)\,dt \\ &= (n-1)\sum_0^a \left(\delta\cos^2(\sqrt{\delta}\,t) - \frac{\sum_{i=2}^n K(e_1;e_i)}{n-1}\sin^2(\sqrt{\delta}\,t)\right)dt \\ &\le (n-1)\delta\int_0^a \big(\cos^2(\sqrt{\delta}\,t) - \sin^2(\sqrt{\delta}\,t)\big)\,dt = 0.\end{aligned}$$

We conclude that $\operatorname{conj}_X \le \pi/\sqrt{\delta}$. □

The *diameter* of a complete Finsler manifold M is the maximal distance of two points. By the Hopf–Rinow theorem [Bao et al. 2000, § 6.6] there is a minimal geodesic between two points of maximal distance. Since a geodesic is not minimal after the first conjugate point, the diameter is at most $\max\{\operatorname{conj}_X \mid X \in T^1M\}$. Therefore we obtain as a consequence of Lemma 8.6 the following generalization of the Bonnet–Myers theorem of Riemannian geometry:

COROLLARY 8.7 [Auslander 1955]. *Let (M,F) be a complete Finsler manifold of dimension n with Ricci curvature* $\operatorname{Ric}(V) \ge \delta(n-1)F^2(V)$ *for some positive δ and for all nonzero tangent vectors V. Then M is compact and its diameter is at most $\pi/\sqrt{\delta}$.*

Since this estimate also holds for the universal covering space, we conclude that the universal covering space is also compact, so the fundamental group of the manifold is finite. In the proof of the Sphere Theorem the following statement is of importance:

PROPOSITION 8.8. *Let (M, F) be a compact and simply connected Finsler manifold of dimension n, and let $p, q \in M$ be such that q is not conjugate to p along any geodesic joining p and q. Assume there is a number $m \geq 2$ such that every nonminimal geodesic c from p to q has index at least m. Then*:

(a) *The manifold is m-connected* (see page 280 for definition).
(b) *If $m = n - 1$, the manifold is homotopy equivalent to the n-sphere.*

PROOF. It follows from the Fundamental Theorem of Morse Theory (Theorem 8.4) that the space Ω_{pq} has the homotopy type of a CW-complex with no cells of dimension $k \in \{1, 2, \ldots, m-1\}$. By Proposition 7.5 this implies that the space Ω_{pq} is $(m-1)$-connected; Proposition 7.9 then implies that M itself is m-connected. Part (b) follows from Corollary 7.7. □

9. Shortest Nonminimal Geodesics and the Sphere Theorem

Now we come to the crucial geometric argument in the proof of the Sphere Theorem. We obtain a lower bound for the length of a nonminimal geodesic c joining two points p, q or a nonconstant geodesic loop. In contrast to a minimal geodesic, this geodesic will meet the cut locus, after which the geodesic is not minimal anymore.

The exponential map $\exp_p : T_pM \to M$ is C^∞-smooth on $T_pM \setminus \{0\}$ and C^1-smooth on T_pM [Shen 2001a, § 11.1]. The differential at $0 \in T_pM$ is an isomorphism; hence there is an $\varepsilon > 0$ such that the restriction

$$\exp_p : B_\varepsilon(T_pM) = \{X \in T_pM | F(X) < \varepsilon\} \to M$$

is a local diffeomorphism onto its image $B_\varepsilon(p) \subset M$. If a piecewise smooth curve $c : [0, a] \to M$ is minimal, that is, $L(c) = \theta\bigl(c(0), c(a)\bigr)$, it follows from Corollary 4.2 that c is a smooth geodesic.

DEFINITION 9.1. For a unit tangent vector $X \in T_pM$, set

$$t(X) = \sup\bigl\{s > 0 \mid \theta(\exp_p(sX), p) = s\bigr\}.$$

Then $q = \exp_p(t(X)X)$ is called a *cut point*. The *cut locus*

$$\mathrm{Cut}(p) := \bigl\{\exp_p(t(X)X) \mid F(X) = 1,\, t(X) < \infty\bigr\}$$

is the union of all cut points on geodesics starting from p. The *injectivity radius at p* is $\mathrm{inj}\, p := \inf\bigl\{\theta(p, q) \mid q \in \mathrm{Cut}(p)\bigr\}$. If the manifold is compact we define the *injectivity radius* of M as $\mathrm{inj} = \mathrm{inj}(M; F) = \inf\bigl\{\mathrm{inj}\ p \mid p \in M\bigr\}$. The *symmetrized injectivity radius at p* is $d(p) := \inf\bigl\{d(p, q) \mid q \in \mathrm{Cut}(p)\bigr\}$. If the manifold is compact, we define the *symmetrized injectivity radius* $d = d(M; F) = \inf\bigl\{d(p) \mid p \in M\bigr\}$. Finally, given two points p, q we define

$$\vartheta(p, q) := \inf\bigl\{\theta(p, r) + \theta(r, q) \mid r \in \mathrm{Cut}\, p\bigr\}.$$

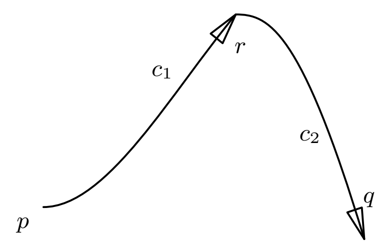
Hence $d(p)$ is the symmetrized distance between p and its cut locus, whereas $\operatorname{inj} p$ is the distance $\theta(p, \operatorname{Cut}(p))$ with respect to the distance function θ. For a reversible Finsler metric these functions coincide. In general we have the bounds

$$\tfrac{1}{2}\big(1+1/\lambda\big)\operatorname{inj} p \le d(p) \le \tfrac{1}{2}\,(1+\lambda)\operatorname{inj} p,$$
$$\frac{2}{1+\lambda}\,d(p) \le \operatorname{inj} p \le \frac{2\lambda}{1+\lambda}\,d(p),$$

which imply the corresponding estimates for global injectivity radii inj and d in case of a compact manifold. Obviously $\vartheta(p,p) = 2d(p)$ and the triangle inequality for θ implies

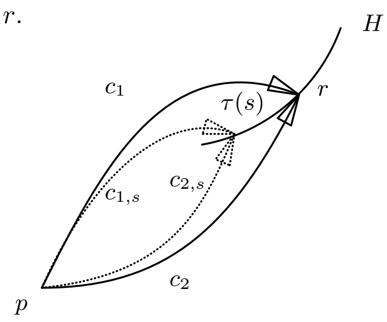
$$\vartheta(p,q) + \theta(q,p) \ge 2d(p). \tag{9–1}$$

If the manifold is compact, the cut locus of a point is also compact, so the infima in the above definitions of the injectivity radius and the symmetrized injectivity radius are actually minima.



Definition 9.2. A *broken geodesic with one corner* joining p and q is a continuous curve $c : [0,b] \to M$ such that $p = c(0)$, $q = c(b)$, and for some $a \in (0,b)$ the restrictions $c_1 = c|[0,a] \to M$ and $c_2 = c|[a,b] \to M$ are minimal geodesics. The point $r = c(a)$ is the *corner* of c. We call c *smooth at* r if $c_1'(a) = c_2'(a)$. The *length* of c is given by $L(c) = L(c_1) + L(c_2) = \theta(p,r) + \theta(r,q)$. If $p = q$, we have a closed broken geodesic, and its length is twice the symmetrized distance between p and r: $L(c) = 2d(p,r)$.

Lemma 9.3. *Let (M,F) be a compact Finsler manifold with reversibility λ and flag curvature $K \le 1$, and let $p \in M$ be a point on M. If there is a cut point $r \in \operatorname{Cut} p$ with $\theta(p,r) < \pi$, there is a local hypersurface $H \subset M$ with $r \in H$ such that for every smooth curve $\tau : (-1,1) \to H$ with $\tau(0) = r$ there are two geodesic variations $c_{1,s}, c_{2,s} : [0, \theta(p,q)] \to M$ with $c_{1,s}(\theta(p,r)) = c_{2,s}(\theta(p,r)) = \tau(s)$, $L(c_{1,s}) = L(c_{2,s})$ and such that $c_1 = c_{1,0}$, $c_2 = c_{2,0}$ are two distinct minimal geodesics joining p and r.*



Proof. We conclude from Lemma 8.6 that, since r is not conjugate to p along a minimal geodesic, there are distinct minimal geodesics $c_1, c_2 : [0, \theta(p,r)] \to M$

parametrized by arc length with $c_1(0)=c_2(0)=p$ and $c_1(\theta(p,r))=c_2(\theta(p,r))=r$. Since r is not conjugate to p along c_1 or c_2, we can choose an open neighborhood $U \subset M$ of r and open disjoint neighborhoods $U_j \subset T_pM$ of $c_j'(0)$, for $j=1,2$, such that the restrictions of the exponential map $\exp_p : T_pM \to M$ to U_1, U_2 are diffeomorphisms; see Proposition 8.2. We define functions $f_1, f_2 : U \to \mathbb{R}$ by setting

$$f_j(v) = F\big((\exp_p|U_j)^{-1}(v)\big), \quad j=1,2.$$

These functions are differentiable and of maximal rank, with $f_j(r) = \theta(p,r)$,

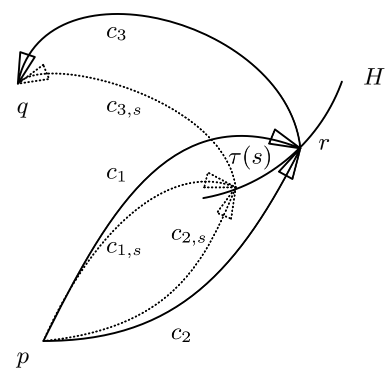
$$\operatorname{grad} f_j(r) = c_j'(\theta(p,r)) \quad \text{and} \quad df_j(r)(X) = \big\langle c_j'(\theta(p,r)), X\big\rangle_{c_j'(\theta(p,r))}$$

for $j=1,2$ and for all X, as follows from the Gauss lemma (see [Bao et al. 2000, § 6.1] and [Shen 2001a, 11.2.1]). One can view f_1, f_2 as distance functions on the Finsler manifolds (U_1', F) and (U_2', F), where U_j' is a small open neigborhood of the image of the curve $c\big([0,\theta(p,r)]\big)$. Since

$$\operatorname{grad} f_1(r) = c_1'(\theta(p,r)) \neq c_2'(\theta(p,r)) = \operatorname{grad} f_2(r),$$

the function $f_1 - f_2$ has maximal rank in an open neighborhood of r, which we again denote by U; thus $H = f^{-1}(0) = \{x \in U \mid f_1(x) = f_2(x)\}$ is a smooth hypersurface with $r \in H$. We finish by setting $c_{1,s}'(0) = f_1^{-1}(\tau(s))$ and $c_{2,s}'(0) = f_2^{-1}(\tau(s))$. □

LEMMA 9.4. *Let (M,F) be a compact Finsler manifold with reversibility λ and with flag curvature $K \le 1$. Let $p, q \in M$ be two points with $\vartheta(p,q) + \theta(q,p) < \pi(1+\lambda^{-1})$. Then there is a cut point $r \in \operatorname{Cut} p$ and a geodesic c of length $\vartheta(p,q)$ parametrized by arc length from p to q going through r.*



PROOF. Choose $r \in \operatorname{Cut} p$ such that $\vartheta(p,q) = \theta(p,r) + \theta(r,q)$. If $\theta(p,r) \ge \pi$, the definition of the reversibility (Lemma 3.9) implies that $\theta(r,p) \ge \pi/\lambda$, hence $\vartheta(p,q) + \theta(q,p) = \theta(p,r) + \theta(r,q) + \theta(q,p) \ge \theta(p,r) + \theta(r,p) = 2d(p,r) \ge \pi(1+\lambda^{-1})$, contradicting the assumption. Thus we have proved that $\theta(p,r) < \pi$.

On the other hand, if $\theta(r,q) \ge \pi$, we have $\theta(q,r) \ge \pi/\lambda$ and $\vartheta(p,q) + \theta(q,p) = \theta(p,r) + \theta(r,q) + \theta(q,p) \ge \theta(r,q) + \theta(q,r) = d(r,q) \ge \pi(1+\lambda^{-1})$. Therefore it follows also that $\theta(r,q) < \pi$.

Since $\theta(p,r) < \pi$, the point r is not conjugate to p along a minimal geodesic joining p. Therefore Lemma 9.3 gives an open hyersurface $H \subset M$ with $r \in H$ such that for every smooth curve $\tau : (-1,1) \to H$ with $\tau(0) = r$ there are variations $c_{1,s}, c_{2,s} : [0, \theta(p,r)] \to M$ such that $L(\tilde{c}_{1,s}) = L(\tilde{c}_{1,s})$, $p = c_{1,s}(0) =$

$c_{2,s}(0)$, $\tau(s) = c_{1,s}(1) = c_{2,s}(1)$ and $c_1 = c_{1,0}, c_2 = c_{2,0} : [0, \theta(p,r)] \to M$ are distinct minimal geodesics joining p and r.

Now let $c_3 : [\theta(p,r), \vartheta(p,q)] \to M$ be a minimal geodesic parametrized by arc length joining $c_3(\theta(p,r)) = r$ and $c_3(2d(p,q)) = p$. Since $\theta(r,q) < \pi$, the point q is not conjugate to r along a minimal geodesic from r to q. Therefore we can choose a geodesic variation $c_{3,s} : [\theta(p,r), \vartheta(p,q)] \to M$ with $c_{3,0} = c_3$, $c_{3,s}(\theta(p,r)) = \tau(s)$ and $c_{3,s}(\vartheta(p,q)) = q$ for all $s \in (-1,1)$.

Now we can combine the smooth geodesic variations $c_{1,s}$, $c_{2,s}$ and $c_{3,s}$ to obtain two piecewise smooth variations $\tilde{c}_{1,s}, \tilde{c}_{2,s} : [0, \vartheta(p,q)] \to M$ with

$$\tilde{c}_{j,s}(t) = \begin{cases} c_{j,s}(t) & \text{if } t \in \big[0, \theta(p,r)\big], \\ c_{3,s}(t) & \text{if } t \in \big[\theta(p,r), \vartheta(p,q)\big]. \end{cases}$$

These are variations of the broken geodesics $\tilde{c}_1 = (c_1, c_3)$ and $\tilde{c}_2 = (c_2, c_3)$ by broken geodesics with fixed end points $p = \tilde{c}_j(0)$ and $q = \tilde{c}_j(\vartheta(p,q))$ and with $\tau'(0) = (\partial/\partial s)|_{s=0}\, \tilde{c}_{j,s}(\theta(p,r))$.

We assume that $c_1'(\theta(p,r)) \neq c_3'(\theta(p,r))$ and $c_2'(\theta(p,r)) \neq c_3'(\theta(p,r))$. Since c_1 and c_2 are distinct, $c_1'(\theta(p,r))$, $c_2'(\theta(p,r))$, $c_3'(\theta(p,r))$ are pairwise disjoint. Recall from Definition 3.3 the Legendre transformation $\mathcal{L}(X)(Y) = \langle Y, X\rangle_X$. Given three pairwise distinct nonzero vectors $X_1, X_2, X_3 \in T_rM$, we have

$$\dim\big\{Y \in T_rM \mid \mathcal{L}(X_1)(Y) = \mathcal{L}(X_2)(Y) = \mathcal{L}(X_3)(Y)\big\} \leq n-2.$$

Applying this $c_1'(\theta(p,r))$, $c_2'(\theta(p,r))$, $c_3'(\theta(p,r))$, we see that there is a tangent vector $V \in T_rH \subset T_rM$ such that $\mathcal{L}\big(c_3(\theta(p,r))\big)(V)$ is not equal simultaneously to $\mathcal{L}\big(c_1(\theta(p,r))\big)(V)$ and $\mathcal{L}\big(c_2(\theta(p,r))\big)(V)$. We assume without loss of generality that

$$\mathcal{L}\big(c_1(\theta(p,r))\big)(V) - \mathcal{L}\big(c_3(\theta(p,r))\big)(V) \neq 0.$$

The first variational formula for the energy functional (Lemma 4.1), applied to the variation $\tilde{c}_{1,s}$ of the broken geodesic $\tilde{c}_1$, yields

$$\frac{d}{ds}\Big|_{s=0} E(\tilde{c}_1) = \mathcal{L}\big(c_1(\theta(p,r))\big) - \mathcal{L}\big(c_3(\theta(p,r))\big) \neq 0.$$

By using $s \mapsto \tau(-s)$ instead of $s \mapsto \tau(s)$, if necessary, we can assume that

$$\frac{d}{ds}\Big|_{s=0} E(\tilde{c}_{1,s}) < 0.$$

It follows that $\theta\big(p, \tilde{c}_{1,s}(\theta(p,r))\big) + \theta\big(\tilde{c}_{1,s}(\theta(p,r)), q\big) < \theta(p,r) + \theta(r,q)$ for small $s > 0$. Since for sufficiently small $s > 0$ the geodesics $\tilde{c}_{1,s}, \tilde{c}_{2,s} : [0, \theta(p,r)] \to M$ intersect at $t = \theta(p,r)$, the cut point $\tilde{c}_{1,s}(t_{1,s})$ of $\tilde{c}_{1,s}$ occurs no later than $\theta(p,r)$, that is, $t_{1,s} \leq \theta(p,r)$. Since

$$\begin{aligned} \theta\big(p, \tilde{c}_{1,s}(t_{1,s})\big) + \theta\big(\tilde{c}_{1,s}(t_{1,s}), q\big) &\leq \theta\big(p, \tilde{c}_{1,s}(\theta(p,r))\big) + \theta\big(\tilde{c}_{1,s}(\theta(p,r)), \tilde{c}_{1,s}(\vartheta(p,q))\big) \\ &= L\big(\tilde{c}_{1,s}\big) < \vartheta(p,q), \end{aligned}$$

we have found for sufficiently small $s > 0$ a cut point $r_{1,s} = \tilde{c}_{1,s}(t_{1,s}) \in \operatorname{Cut} p$ satisfying $\theta(p, r_{1,s}) + \theta(r_{1,s}, q) < \theta(p, r) + \theta(r, q) = \vartheta(p, q)$, which contradicts the definition of $\vartheta(p, q)$. Hence $c_1'(\theta(p, r)) = c_3'(\theta(p, r))$, that is, the broken geodesic (c_1, c_3) with break point r is actually smooth. □

LEMMA 9.5. *Let (M, F) be a compact Finsler manifold with reversibility λ and flag curvature $K \leq 1$.*

(a) *Let $p, q \in M$ with $q \notin \operatorname{Cut} p$ and assume that $\vartheta(p, q) + \theta(q, p) < \pi(1 + \lambda^{-1})$. Then $\vartheta(p, q)$ is the length of the shortest nonminimal geodesic from p to q.*
(b) *If the symmetrized injectivity radius d of M satisfies $d < \pi(1+\lambda^{-1})/2$, there is a shortest geodesic loop c with initial point p and a point $q \in \operatorname{Cut}(p)$ on this loop with $L(c) = 2d = 2d(p, q)$.*

PROOF. (a) The cut locus $\operatorname{Cut} p$ is a closed subset; hence there exists $r \in \operatorname{Cut} p$ with $\theta(p, r)+\theta(r, q) = \vartheta(p, q)$. It follows from Lemma 9.4 that there is a geodesic c from p to q through r with $L(c) = \vartheta(p, q)$. Since r is a cut point and $r \neq q$, this geodesic is not minimal.

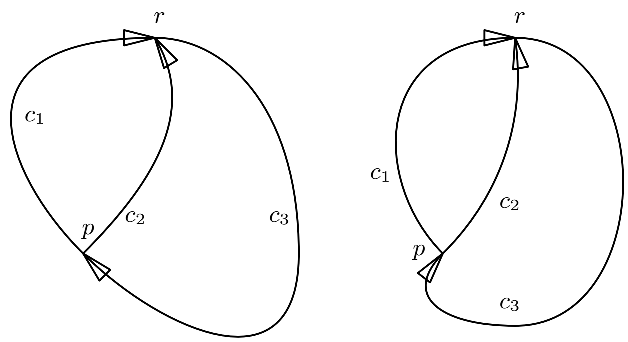
(b) Let $q \in \operatorname{Cut}(p)$ be a point with $d = d(p, q)$. We know from Lemma 9.4 that there is a geodesic loop c with $c(0) = p$ and $L(c) = 2d$. We only have to show that this curve is a shortest geodesic loop. If c_1 is a shortest geodesic loop with $c_1(0) = p$ and with cut point $q = c_1(t_0)$, the restriction $c_1|[0, t_0]$ is minimal and $L(c_1) \geq 2d(p, q)$. But we showed in Lemma 9.4 that there is a geodesic loop $c \in \Omega_p$ with $L(c) = 2d = 2d(p)$. This finally implies that $L(c_1) = 2d$. □

REMARK 9.6. If the Finsler metric is reversible, the proof of Lemma 9.4 simplifies considerably. The argument for this case was introduced by Klingenberg [1995, 2.1.11] in the Riemannian setting. The minimal geodesic c_3 coincides with one of the minimal geodesics c_1, c_2 (say c_1) up to orientation, that is, $\theta(p, r) = d(p, r)$ and $c_3(t) = c_1(2d(p, r) - t)$. By using the same argument exchanging the roles of p, r one can prove then that there is a closed geodesic c of length $2d$. If c is parametrized by arc length, $c(d)$ is the cut point, and there is no shorter geodesic loop.

If we use the same argument as in the proof of Lemma 9.4, we obtain:

LEMMA 9.7. *Let (M, F) be a compact Finsler manifold with reversibility λ and flag curvature $K \leq 1$. If the symmetrized injectivity radius satisfies $d < \pi\left(1 + \frac{1}{\lambda}\right)$, there is a point $p \in M$ and a cut point $r \in \operatorname{Cut} p$ such that either*

(a) *there is a closed geodesic $c : [0, 2d] \to M$ parametrized by arclength with $L(c) = 2d$ and $c(0) = p$, $c(\theta(p, r)) = r$, or*
(b) *there are two distinct geodesic loops $c_1, c_2 : [0, 2d] \to M$ parametrized by arc length (that is, both have the same length $2d$) with $c_1(0) = p = c_2(\theta(r, p))$ and $c_1(\theta(p, r)) = r = c_2(0)$.*



PROOF. If we use the statement of Lemma 9.4 and the argument in the proof of the same lemma, exchanging the roles of p and q, we reach the following statement: There are three minimal geodesics $c_1, c_2 : [0, \theta(p,q)] \to M$ and $c_3 : [\theta(q,p), 2d] \to M$ with $d = d(p,q)$, $p = c_1(0) = c_2(0) = c_3(2d)$, and $q = c_1(\theta(p,q)) = c_2(\theta(p,q)) = c_3(\theta(p,q))$. Without loss of generality we can assume that the broken geodesic (c_1, c_3) with corner q is smooth at q: $c_1'(\theta(p,q)) = c_3'(\theta(p,q))$. In addition there are two cases: Either the broken geodesic formed by c_3, c_1 is smooth at p, so that $c_3'(2d) = c_1'(0)$, which means that the geodesics c_1, c_3 form a closed geodesic, or the broken geodesic formed by c_3, c_2 is smooth at p. □

In the reversible case we have case (a) of the lemma. It is not clear whether case (b) occurs in the nonreversible case.

LEMMA 9.8. *Let (M, F) be a compact Finsler manifold with reversibility λ and flag curvature $K \le 1$ and let $p, q \in M$ be two points with distance $\theta(p,q) < \operatorname{inj} p$. Let $c_1 : [0,1] \to M$ be a nonminimal geodesic with $c_1(0) = p$ and $c_1(1) = q$, and let $c_0 : [0,1] \to M$ be a curve such that the reversed curve $c_0^{-1} : [0,1] \to M$, $c_0^{-1}(t) = c_0(1-t)$ is a minimal geodesic with $q = c_0(1)$, $p = c_0(0)$ and length $L(c_0) = \theta(q,p)$. If $c_s : [0,1] \to M$, $s \in [0,1]$, is a homotopy of piecewise smooth curves with fixed endpoints $p = c_s(0)$, $q = c_s(1)$ for all $s \in [0,1]$ between the curves c_0, c_1, then*

$$\theta(q,p) + \max_{s \in [0,1]} L(c_s) \ge \pi\left(1 + \frac{1}{\lambda}\right).$$

PROOF. We are assuming that $\theta(q,p) + L(c_s) < \pi(1 + \lambda^{-1})$ for all $s \in (0,1]$; hence there is a $\rho > 0$ such that $\theta(q,p) + L(c_s) \le (\pi - \rho)(1 + \lambda^{-1})$ for all $s \in [0,1]$. We show by contradiction that

$$\theta(p, c_s(t)) \le \pi - \rho \quad \text{for all } s, t \in [0,1]. \tag{9–2}$$

If there are $s, t \in [0,1]$ such that $\theta(p, c_s(t)) > \pi - \rho$, then

$$\begin{aligned}\theta(q,p) + L(c_s) &\ge \theta(p, c_s(t)) + \theta(c_s(t), q) + \theta(q,p) \\ &\ge \theta(p, c_s(t)) + \theta(c_s(t), p) \ge \theta(p, c_s(t))(1 + \lambda^{-1}) \\ &> (\pi - \rho)(1 + \lambda^{-1}),\end{aligned}$$

which contradicts our assumption.

Define the closed ball $\bar{B}_a(T_pM) := \{v \in T_pM \mid F(v) \le a\}$ of radius $a > 0$ in the tangent space T_pM at p. The subset $\bar{B}_a(p) := \{x \in M \mid \theta(p,x) \le a\} \subset M$ equals the image $\exp_p(\bar{B}_a(T_pM))$ of the exponential map $\exp_p : T_pM \to M$ for an arbitrary $a > 0$. It follows from 9–2 that $c_s(t) \in \bar{B}_{\pi-\rho}(p)$ for all $s, t \in [0,1]$. The restriction

$$F := \exp_p : \bar{B}_{\pi-\rho}(T_pM) \to \bar{B}_{\pi-\rho}(p)$$

has everywhere maximal rank since the flag curvature satisfies $K \le 1$; thus it is a local diffeomorphism. The restriction

$$\exp_p : B_{\operatorname{inj} p}(T_pM) \to B_{\operatorname{inj} p}(p)$$

is a diffeomorphism, since $\theta(q,p) < \operatorname{inj} p$ for sufficiently small $\eta > 0$ we have $c_s([0,1]) \subset B_{\operatorname{inj} p}(p)$ for all $s \in [0,\eta)$. Hence there is a uniquely defined lift

$$\tilde{c_s} : t \in [0,1) \mapsto \tilde{c}_s(t) \in \bar{B}_{\operatorname{inj} p}(T_pM)$$

for $s \in [0,\eta)$ with $c_s(t) = \exp_p(\tilde{c}_s(t))$, $0_p = \tilde{c}_s(0)$, and $X = \tilde{c}_0(1) = \tilde{c}_s(1)$ for all $s \in [0,\eta)$. Since the restriction $F = \exp_p|\bar{B}_{\pi-\delta}(T_pM)$. is a local diffeomorphism, there is a uniquely determined extension

$$\tilde{c_s} : t \in [0,1) \mapsto \tilde{c}_s(t) \in \bar{B}_{\pi-\rho}(T_pM)$$

with $c_s(t) = \exp_p(\tilde{c}_s(t))$ of the lift for all $s \in [0,1]$. It remains to show that this lift is a homotopy with fixed end points. Define $J_0 := \{s \in [0,1] | \tilde{c}_s(0) = 0_p\}$, $J_1 := \{s \in [0,1] | \tilde{c}_s(1) = X\}$; these subsets contain the nonempty interval $[0,\eta)$ and are closed in $[0,1]$. Since the restriction F of the exponential map is a local diffeomorphism the subsets $J_0, J_1 \subset [0,1]$ are also open, hence $J_0 = J_1 = [0,1]$. By assumption $c_1 : [0,1] \to M$ is a geodesic from p to q; hence $\tilde{c}_1(t) = tc_1'(0)$ for all $t \in [0,1]$, contradicting $\tilde{c}_1(1) = 0_p$. Therefore we arrive at a contradiction starting from the assumption $\theta(q,p) + L(c_s) < \pi\left(1 + \frac{1}{\lambda}\right)$ for all $s \in [0,1]$, which finally proves the claim. □

With the long homotopy lemma we are able to gain a lower bound for the length of nonminimal geodesics:

PROPOSITION 9.9. *Let (M,F) be a simply connected, compact Finsler manifold of dimension $n \ge 3$, with reversibility λ and flag curvature $\left(1 - \frac{1}{1+\lambda}\right)^2 < K \le 1$. If $p \in M$, there exists for every $\varepsilon > 0$ a point q that is a regular point of $\exp_p$ and that satisfies $\theta(q,p) < \varepsilon$ and $\vartheta(p,q) + \theta(q,p) \ge \pi\left(1 + \frac{1}{\lambda}\right)$.* (Recall that $\vartheta(p,q)$ is the length of a shortest nonminimal geodesic from p to q.)

PROOF. Since M is compact we can choose $\delta > \left(1 - \frac{1}{1+\lambda}\right)^2$ such that the flag curvature K of satisfies $\delta < K \le 1$. For a given $\varepsilon > 0$ we choose a regular point q of $\exp_p$ with $\theta(q,p) < \varepsilon$ and

$$\theta(q,p) < \pi\left(1 + \frac{1}{\lambda} - \frac{1}{\sqrt{\delta}}\right).$$

Now we assume that there is a shortest nonminimal geodesic $c_1 : [0,1] \to M$ from $p = c_1(0)$ to $q = c_1(1)$ with length $L(c_1)$ and satisfying $L(c_1) + \theta(q,p) < \pi\left(1 + \frac{1}{\lambda}\right)$.

M is simply connected, so there is a path $s \in [0,1] \mapsto c_s \in \Omega_{pq}M$ going from c_0 (the reverse of the minimal geodesic c_0^{-1} from q to p of length $\theta(q,p)$) to the geodesic c_1. This path describes a map

$$H : ([0,1], \{0,1\}) \to \left(\Omega_{pq}M, \Omega_{pq}^{\kappa-}M\right),$$

with $H(s) = c_s$, $\Omega_{pq}^{\kappa-}M := \{\gamma \in \Omega_{pq}M \mid E(\gamma) < \kappa\}$, and κ defined by $2\sqrt{\kappa} = \pi(1 + \lambda^{-1}) - \theta(q,p)$. Let $c^* : [0,1] \to M$ be a geodesic from p to q with length $L(c^*) \geq \pi\left(1 + \frac{1}{\lambda}\right) - \theta(q,p) \geq \pi/\sqrt{\delta}$. Lemma 8.6 gives the bound $\operatorname{ind} c \geq n - 1$, so $\operatorname{ind} c \geq 2$ in view of the assumption $n \geq 3$. We can conclude from the Fundamental Theorem of Morse Theory (Theorem 8.4) that the pair $(\Omega_{pq}M, \Omega_{pq}^{\kappa-}M)$ has the homotopy type of a CW-complex with no 1-dimensional cells; hence the relative homotopy group

$$\pi_1\left(\Omega_p M, \Omega_p^{\kappa-}M\right) = 0,$$

is 1-connected, as is the pair $\left(\Omega_{pq}M, \Omega_{pq}^{\kappa-}M\right)$; see Proposition 7.5. Therefore there is a map $\Phi : (u,s) \in [0,1] \times ([0,1], \{0,1\}) \mapsto \Phi_u(s) \in (\Omega_{pq}M, \Omega_{pq}^{\kappa-}M)$ with $\Phi_u(0)(t) = c_0(t)$, $\Phi_u(1)(t) = c(t)$ for all $t, u \in [0,1]$ and $\Phi_0(s) = c_s$ and $\bar{c}_s = \Phi_1(s) \in \Omega_{pq}^{\kappa-}$ for all $s \in [0,1]$. This implies that $L(\bar{c}_s) < \pi\left(1 + \frac{1}{\lambda}\right) - \theta(q,p)$ for all $s \in [0,1]$, that and $\bar{c}_0 = c_0$ is up to orientation the minimal geodesic, and that $\bar{c}_1 = c$ is a shortest nonminimal geodesic joining p and q. But we conclude from the Long Homotopy Lemma 9.8 that there is a $s^* \in (0,1)$ with $L(\bar{c}_{s^*}) \geq \left(1 + \frac{1}{\lambda}\right) - \theta(q,p)$, which is a contradiction. □

The assumption that q is a regular value of the exponential map $\exp_p$ ensures that the energy functional is a Morse function, so all geodesics joining p and q are nondegenerate. If one aims at estimating the length of geodesic loops or closed geodesics it won't be the case in general that p itself is a regular value of the exponential map $\exp_p$. For example, on the standard sphere every point p is conjugate to itself along a great circle; in particular every point p is a critical point of $\exp_p$. But the statement of Proposition 9.9 is also correct if we remove this assumption. In that case either one has to use a version of Morse theory including degenerate critical points [Rademacher 2004, Theorem 3] or one can argue as follows:

Theorem 9.10. *Let (M,F) be a simply connected, compact Finsler manifold of dimension $n \geq 3$, with reversibility λ and flag curvature $\left(1 - \frac{1}{1+\lambda}\right)^2 < K \leq 1$. Then every nonconstant geodesic loop c has length at least $\pi\left(1 + \frac{1}{\lambda}\right)$ and the injectivity radius satisfies* $\operatorname{inj} \geq \pi/\lambda$.

Proof. For a point $p \in M$ the function

$$(q,r) \in M \times \operatorname{Cut} p \mapsto \theta(p,r) + \theta(r,q) \in \mathbb{R}$$

is continuous, hence also the map

$$q \in M \mapsto \vartheta(p,q) = \inf \big\{ \theta(p,r) + \theta(r,q) \mid r \in \operatorname{Cut} p \big\}.$$

Choose a sequence $(q_i)_{i\in\mathbb{N}}$ of regular points of the exponential map $\exp_p$ with $\lim_{i\to\infty} \theta(q_i, p) = 0$. We conclude from Proposition 9.9 that

$$2d(p) = \lim_{i\to\infty} \vartheta(p, q_i) \geq \pi\Big(1 + \frac{1}{\lambda}\Big).$$

The estimate for the injectivity radius follows from Equation 9–1. □

Now we can prove the Sphere Theorem:

Theorem 9.11. *A simply connected and compact Finsler manifold of dimension* $n \geq 3$ *with reversibility* λ *and with flag curvature* K *satisfying* $\big(1 - \frac{1}{1+\lambda}\big)^2 < K \leq 1$ *is homotopy equivalent to the* n*-sphere*.

For $n = 2$, Synge's Theorem (Theorem 10.2) implies that an orientable compact surface carrying a Finsler metric of positive flag curvature $K > 0$ is diffeomorphic to the 2-sphere.

Proof of the Sphere Theorem. Since M is compact we can choose $\delta > \pi\left(1 + \lambda^{-1}\right)$ such that the flag curvature K satisfies $\delta < K \leq 1$. We choose $\varepsilon > 0$ with

$$\varepsilon < \pi\Big(1 + \frac{1}{\lambda} - \frac{1}{\sqrt{\delta}}\Big). \tag{9–3}$$

We conclude from Proposition 9.9 that there is a regular point $q \in M$ of the exponential map $\exp_p$ with $\theta(q,p) < \varepsilon$ and $\vartheta(p,q) \geq \pi\left(1 + \frac{1}{\lambda}\right) - \varepsilon$; hence $\vartheta(p,q) \geq \pi/\sqrt{\delta}$, by 9–3. We conclude from Lemma 8.6 that the index $\operatorname{ind} c$ of a nonminimal geodesic c joining p and q satisfies $\operatorname{ind} c \geq n - 1$. Then Proposition 8.8 implies that M is homotopy equivalent to the n-sphere. □

10. Length of Closed Geodesics in Even Dimensions

In even dimensions one obtains a lower bound for the length of closed geodesics for every metric of positive curvature on a simply connected manifold without assuming a lower curvature bound. The crucial point (*Synge's argument*) is that in even dimensions there is a periodic parallel vector field along a closed geodesic. By scaling the metric we can assume that the flag curvature K satisfies $0 < K \leq 1$.

Lemma 10.1. *Let* (M, F) *be a compact*, *oriented Finsler manifold of even dimension with positive flag curvature* $0 < K \leq 1$. *For every closed geodesic* c *there is a parallel and periodic vector field* W *with* $\langle W, W\rangle_{c'} = 1$ *and* $\langle W, c'\rangle_{c'} = 0$.

Proof. The covariant derivative (∇/dt) along the geodesic $c : [0,1] \to M$ with $\dot c(0) = \dot c(1)$ defines a parallel transport $P : T_pM \to T_pM$ with $P(X(0)) = X(1)$ and $X = X(t)$ is a parallel vector field along c with respect to (∇/dt).

Since $(d/dt)\big\langle X(t), X(t)\big\rangle_{c'} = 2\big\langle(\nabla/dt)X(t), X(t)\big\rangle_{c'}$, the parallel transport is an orientation-preserving isometry. Since $P(\dot c(0)) = \dot c(1)$, the parallel transport defines an isometry of the orthogonal complement

$$T_p^{\perp}M := \big\{X \in T_pM \mid \langle X, \dot c\rangle_{c'} = 0\big\}.$$

This vector space has odd dimension, so there exists a nonzero eigenvector to the eigenvalue 1, that is, a vector $v \in T_p^{\perp}M$ with $Pv = v$. Then the parallel field W along c with $W(0) = v$ is periodic: $W(1) = W(0)$. □

Now we prove a generalization of Synge's Theorem:

THEOREM 10.2. *Let (M, F) be a compact Finsler manifold of positive flag curvature.*

(a) *If M is orientable, it is simply connected.*
(b) *If M is nonorientable, its fundamental group satisfies $\pi_1(M) = \mathbb{Z}_2$.*

PROOF. Let M be orientable, and assume that $\pi_1(M) \neq 0$. Then there is a nontrivial homotopy class in M and a shortest closed curve in this nontrivial homotopy class is a closed geodesic $c : [0, 1] \to M$. By Lemma 10.1 there is a parallel and periodic vector field W along c; the index form at W satisfies

$$I_c(W; W) = \int_0^1 \big\langle R^{c'}(W, c')c', W\big\rangle_{c'} dt < 0;$$

therefore there is a variation by homotopic closed curves c_s with $L(c_s) < L(c)$ for $s > 0$, contradicting the assumption that c is a shortest closed curve in the given homotopy class. Hence $\pi_1(M) = 0$.

If M is nonorientable one passes to the orientable double cover, which by (a) is simply connected, so $\pi_1(M) = \mathbb{Z}_2$. □

THEOREM 10.3. *Let (M, F) be a simply connected compact Finsler manifold of even dimension $n \geq 2$ with reversibility λ and with flag curvature $0 < K \leq 1$. Then every nonconstant closed geodesic c has length $L(c) \geq \pi\left(1 + \frac{1}{\lambda}\right)$.*

PROOF. Let $c : S^1 \to M$ be a shortest closed geodesic with $0 < L(c) < \pi\left(1 + \frac{1}{\lambda}\right)$. By Lemma 10.1, there exists a parallel unit vector field W along c; it follows that the index form I_c on $V_c^{\perp}$ satisfies $I_c(W, W) < 0$. Let c_s, $s \in (-\varepsilon, \varepsilon)$, be a variation of $c = c_0$ with variation vector field W. It follows from the second variation formula that $E(c_s) < E(c_0)$ for all $s \in (-\varepsilon, 0) \cup (0, \varepsilon)$. Since there are no critical values of E in the interval $\big(0, E(c_0)\big)$, there is a map $h_s : S^1 \to M$, $s \in [-1, 1]$, with $c = h_0$, $L(h_1) = L(h_{-1}) = 0$ and $L(h_s) < L(c) = L(h_0)$ for all nonzero $s \in (-1, 1)$. One can generalize the Long Homotopy Lemma 9.8 to the case of homotopies $c_s : S^1 \to M$ of freely homotopic closed curves. This generalization yields a contradiction. □

11. An Example

Shen [2002] constructed Finsler metrics of constant flag curvature and Randers type. It turns out that in the Hamiltonian description these are the metrics introduced in [Katok 1973] and investigated in [Ziller 1982], as observed in [Rademacher 2004, Chapter 4]. These examples show that the estimates in Proposition 9.9 and Theorem 9.10, for the lengths of nonminimal geodesics between fixed points and of nonconstant geodesic loops, are sharp.

One can describe a Finsler metric using the Legendre transformation with a Hamiltonian function [Ziller 1982, Chapter 1]. The Katok examples on $S^2 = \{(x,y,z) \in \mathbb{R}^3 \mid x^2+y^2+z^2 = 1\}$ can be introduced as follows. We start with the standard Riemannian metric g on the 2-sphere S^2, letting g^* be the dual metric on the cotangent bundle T^*M. In the Hamiltonian description the standard metric is determined by the quadratic Hamiltonian function

$$y \in T^*S^2 \mapsto g^*(y,y) \in \mathbb{R},$$

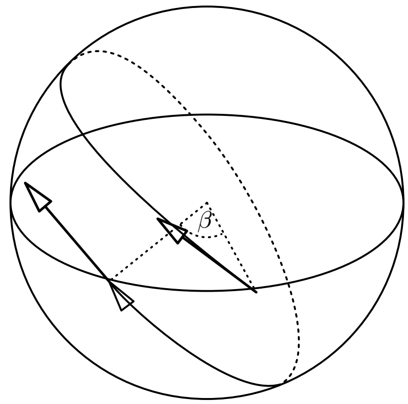
or by the 1-homogeneous Hamiltonian $H_0 : T^*S^2 \to \mathbb{R},\ H_0(y) = \sqrt{g^*(y,y)}$. Let $\psi_0^t : T^*S^2 \to T^*S^2$ be the corresponding Hamiltonian flow. Then $t \in \mathbb{R} \mapsto \tau^*(\psi_0^t(y))$ is a geodesic of the standard metric; here $\tau^* : T^*S^2 \to S^2$ is the projection of the cotangent bundle.

Let $V(x,y,z) = (y,-x,0)$ be the Killing field belonging to the 1-parameter subgroup $\phi^t : S^2 \to S^2$ generated by the rotations around the z-axis. A. Katok introduced the following perturbation of Randers type:

$$H_\varepsilon : T^*S^2 \to \mathbb{R}; H_\varepsilon(y) = \sqrt{g^*(y,y)} + \varepsilon y(V).$$

In [Bao et al. 2003] these perturbations are connected to *Zermelo navigation*. For $\varepsilon \in [0,1)$ this defines a quadratic Hamiltonian $\frac{1}{2}H_\varepsilon^2$ and using the Legendre transformation of this Hamiltonian we obtain a Finsler metric F_ε.

The description of the geodesics appears to be easier in the Hamiltonian picture: Since ϕ^t is a group of isometries leaving H_0 invariant, the Hamiltonian flow ψ_ε^t of the quadratic Hamiltonian $\frac{1}{2}H_\varepsilon^2$ is generated by two commuting flows, $\psi_\varepsilon^t = \psi_0^t \circ (\phi^{\varepsilon t})^*$. Here $(\phi^t)^* : T^*S^2 \to T^*S^2$ is the flow on the cotangent bundle induced by differentiating ϕ^t. The projection of the Hamiltonian flow onto the 2-sphere yields the geodesics of the Finsler metric. As described in [Ziller 1982, Chapter 1] one can visualize the geodesic flow of these Finsler metrics by identifying the cotangent bundle T^*S^2 with the tangent bundle T_*S^2 via the standard Riemannian metric g. Then the geodesic flow can be seen as the geodesic flow of the standard metric observed from a coordinate system rotating around the z-axis with constant speed $2\pi\varepsilon$, as shown in the figure on the next page. For irrational ε the only closed geodesics are $c_\pm(t) = (\cos 2\pi t, \pm \sin 2\pi t, 0)$, $t \in [0,1]$, i.e., the equator with both orientations. (We consider c_+ and c_- geometrically distinct; for example their lengths $L(c_\pm) = 2\pi/(1 \pm \varepsilon)$ differ.)



Using the results [Hrimiuc and Shimada 1996, Theorem 5.8], [Shen 2001a, Example 3.1.1] one obtains in geodesic polar coordinates $(r,\phi) \in (0,\pi) \times [0,2\pi]$ of the standard metric the following formula for F_ε :

$$F_\varepsilon = \frac{\sqrt{(1-\varepsilon^2 \sin^2 r)\, dr^2 + \sin^2(r)\, d\phi^2} - \varepsilon \sin^2 r\, d\phi}{1-\varepsilon^2 \sin^2 r}. \tag{11–1}$$

It is shown in [Shen 2002, Remark 3.1] that this metric has constant flag curvature 1. The reversibility of the Finsler metrics F_ε can be computed by

$$\max\left\{H_\varepsilon(-y) \mid y \in T^*S^2,\, H_\varepsilon(y) = 1\right\}.$$

Collecting the results from [Ziller 1982] and [Rademacher 2004, Chapter 5] one obtains:

Theorem 11.1 [Rademacher 2004, Theorem 5]. *There is a one-parameter family F_ε, $\varepsilon \in [0,1)$, of Finsler metrics on the 2-sphere S^2 of constant flag curvature 1. These Finsler metrics are nonreversible for $\varepsilon \in (0,1)$ and F_0 is the standard metric. The reversibility is $\lambda = (1+\varepsilon)/(1-\varepsilon)$. If ε is irrational there are exactly two geometrically distinct closed geodesics $c_\pm$ of length $L(c_\pm) = 2\pi(1\pm\varepsilon)^{-1}$. In particular the shortest closed geodesic c_+ satisfies $L(c_+) = 2\pi(1+\lambda^{-1}) = \pi/(1+\varepsilon)$. The injectivity radius and the diameter equal π.*

References

[Abresch and Meyer 1997] U. Abresch and W. Meyer, *Injectivity radius estimates and sphere theorems*, pp. 1–47 in: *Comparison geometry*, MSRI Publications **30** (1997).

[Auslander 1955] L. Auslander, "On curvature in Finsler geometry", *Trans. Amer. Math. Soc.* **79** (1955), 378–388.

[Bao et al. 2000] D. Bao, S.-S. Chern, and Z. Shen, *An introduction to Riemann–Finsler geometry*, Grad. Texts Math. **200**, Springer, New York, 2000.

[Bao et al. 2003] D. Bao, C. Robles, and Z. Shen, "Zermelo navigation on Riemannian manifolds", preprint, 2003.

[Berger 1998] M. Berger, “Riemannian geometry during the second half of the twentieth century”, *Jber. d. Dt. Math.-Verein.* **100** (1998), 45–208.

[Bredon 1995] G. E. Bredon, *Topology and geometry*, Grad. Texts Math. **139**, Springer, New York, corr. 2nd printing, 1995.

[Bryant 2002] R. Bryant, “Some remarks on Finsler manifolds with constant flag curvature”, *Houston J. Math.* **28** (2002), 221–262.

[Dazord 1968a] P. Dazord, “Variétés finslériennes de dimension paire δ-pincées”, *C. R. Acad. Sc. Paris* **266** (1968), 496–498.

[Dazord 1968b] P. Dazord, “Variétés finslériennes en forme des sphères”, *C. R. Acad. Sc. Paris* **267** (1968), 353–355.

[Hrimiuc and Shimada 1996] D. Hrimiuc and H. Shimada, “On the $\mathcal{L}$-duality between Lagrange and Hamilton manifolds”, *Nonlin. World* **3** (1996), 613–641.

[Katok 1973] A. Katok, “Ergodic properties of degenerate integrable Hamiltonian systems” (Russian), *Izs. Akad. Nauk SSSR Ser. Mat.* **37** (1973), 539–576; translated in *Math. USSR Izv.* **7** (1973), 535–572.

[Klingenberg 1961] W. Klingenberg, “Über Riemannsche Mannigfaltigkeiten mit positiver Krümmung”, *Comment. Math. Helv.* **35** (1961), 47–54.

[Klingenberg 1963] W. Klingenberg, “Manifolds with restricted conjugate locus”, *Ann. Math.* **78** (1963), 527–547.

[Klingenberg 1995] W. Klingenberg, *Riemannian geometry*, de Gruyter Studies Math. **1**, 2nd rev. ed., de Gruyter, Berlin, New York, 1995.

[Matthias 1980] H. H. Matthias, *Zwei Verallgemeinerungen eines Satzes von Gromoll und Meyer*, Bonner Math. Schr. **126** (1980).

[Mercuri 1971] F. Mercuri, “The critical point theory for the closed geodesic problem”, *Math. Zeitschr.* **156** (1971), 231–245.

[Milnor 1969] J. Milnor, *Morse theory*, Annals of Math. Studies **51** Princeton Univ. Press, Princeton 1969.

[Milnor 2003] J. Milnor, “Towards the Poincaré conjecture and the classification of 3-manifolds”, *Notices Amer. Math. Soc.* **50**:10 (2003), 1226–1233.

[Rademacher 1992] H.-B. Rademacher, *Morse-Theorie und Geschlossene Geodätische*, Bonner Math. Schr. **229** (1992).

[Rademacher 2004] H.-B. Rademacher, “A sphere theorem for non-reversible Finsler metrics”, *Math. Ann.* **328** (2004), 373–387.

[Shen 2001a] Z. Shen, “Lectures on Finsler geometry”, World Scientific, Singapore, 2001.

[Shen 2001b] Z. Shen, *Differential geometry of spray and Finsler spaces*, Kluwer, Dordrecht, 2001.

[Shen 2002] Z. Shen, “Two-dimensional Finsler metrics with constant curvature”, *Manuscr. Math.* **109** (2002), 349–366.

[Spanier 1966] E. H. Spanier, “Algebraic Topology”, McGraw-Hill, New York, 1966; reprinted by Springer, New York, 1989.

[Ziller 1982] W. Ziller, “Geometry of the Katok examples”, *Ergod. Th. Dyn. Syst.* **3** (1982), 135–157.

Hans-Bert Rademacher
Universität Leipzig
Mathematisches Institut
Augustusplatz 10/11
04109 Leipzig
Germany
rademacher@math.uni-leipzig.de