CLOSED GEODESICS ON REVERSIBLE FINSLER 2-SPHERES

GUIDO DE PHILIPPIS, MICHELE MARINI, MARCO MAZZUCCHELLI,
AND STEFAN SUHR

ABSTRACT. We extend two celebrated theorems on closed geodesics of Riemannian 2-spheres to the larger class of reversible Finsler 2-spheres: Lusternik-Schnirelmann's theorem asserting the existence of three simple closed geodesics, and Bangert-Franks-Hingston's theorem asserting the existence of infinitely many closed geodesics. In order to prove the first theorem, we employ the generalization of Grayson's curve shortening flow developed by Angenent-Oaks.

CONTENTS

1. INTRODUCTION

Since the seminal work of Hadamard [Had98], Poincaré [Poi05], Birkhoff [Bir66], and Morse [Mor96], it became evident that closed Riemannian manifolds of dimension at least 2 tend to have many closed geodesics (that is, periodic orbits of the geodesic flow). This evidence was supported by celebrated theorems of Gromoll-Meyer [GM69b] and Vigué Poirrier-Sullivan [VPS76], which together assert that

simply connected closed Riemannian manifolds with a non-monogenic rational co-homology ring always have infinitely many closed geodesics. This statement covers a large class of simply connected closed manifolds, leaving out those with the co-homology of a compact rank-one symmetric space: S^n, $\mathbb{CP}^n$, $\mathbb{HP}^n$, and $\mathrm{Ca}\mathbb{P}^2$. As of 2019, it is an open conjecture whether these spaces admit infinitely many closed geodesics for any choice of the Riemannian metric. The only known case is the one of S^2, for which the proof required a combination of spectacular work by Bangert [Ban93], Franks [Fra92], and Hingston [Hin93] (either Franks' or Hingston's work, together with Bangert's one, provide the full result). The starting point for this work is another celebrated result due to Lusternik-Schnirelmann [LS34], asserting that every Riemannian 2-sphere possesses at least three simple closed geodesics (that is, closed geodesics that are embedded circle in the Riemannian manifold). For many decades Lusternik-Schnirelmann's theorem was considered controversial due to a gap in their construction of a pseudo-gradient flow for the length function of simple closed curves, that have been subsequently addressed by many authors. Nowadays, the gap is considered filled, for instance thanks to the work of Grayson [Gra89] on the curve shortening flow.

The closed geodesic problem can be studied on closed Finsler manifolds as well. A Finsler metric on a manifold M is a continuous function $F : \mathrm{T}M \to [0, \infty)$, smooth outside the zero-section of $\mathrm{T}M$, positively homogeneous of degree 2 (i.e. $F(x, \lambda v) = \lambda F(x, v)$ for all $(x, v) \in \mathrm{T}M$ and $\lambda \geq 0$), and such that the restriction of its square F^2 to any fiber of $\mathrm{T}M$ has positive definite Hessian everywhere outside the origin. In the literature, a more general notion of Finsler metric is sometimes employed, but the one given here is the most appropriate for the study of geodesic flows. Many results, such as Gromoll-Meyer's one, remain valid essentially with the same proof in the Finsler category (see [Lu15] and references therein). However, a striking example due to Katok [Kat73], and further explored by Ziller [Zil83], shows that Lusternik-Schnirelmann's and Bangert-Franks-Hingston's theorems fail: there exists a Finsler metric on S^2 having only two closed geodesics.

A Finsler metric $F : \mathrm{T}M \to [0, \infty)$ is called reversible when $F(x, -v) = F(x, v)$ for all $(x, v) \in \mathrm{T}M$. The Katok's Finsler metric does not satisfy this property. In the current paper, we show that all the above mentioned results valid for Riemannian 2-spheres remain valid for reversible Finsler 2-spheres.

1.1. **The curve shortening semi-flow.** In [Oak94], Oaks provided a generalization of Grayson's curve shortening flow [Gra89]. As remarked by Angenent [Ang08], such generalization allows to provide a curve shortening flow on orientable Finsler surfaces: a tool to shrink embedded circles without creating self-intersections. In this section, we state a theorem that summarizes all the properties of this flow (actually, a semi-flow) that we will need for the application to the closed geodesics problem.

Let M be a closed oriented surface, equipped with a reversible Finsler metric F. We denote by $S^1 := \mathbb{R}/\mathbb{Z}$ the 1-periodic circle, and by $\mathrm{Emb}(S^1, M)$ the space of smooth embedded loops $\gamma : S^1 \hookrightarrow M$, endowed with the C^∞ topology (that is, the topology whose basis is given by the open sets $\mathcal{U} \subset \mathrm{Emb}(S^1, M)$ of the C^k topology, for all $k \in \mathbb{N}$). We consider the Finsler length functional

$$L : \mathrm{Emb}(S^1, M) \to (0, \infty), \qquad L(\gamma) = \int_0^1 F(\gamma(u), \dot{\gamma}(u))\, \mathrm{d}u. \qquad (1.1)$$

The group of diffeomorphisms $\mathrm{Diff}(S^1)$ acts freely on $\mathrm{Emb}(S^1, M)$ by reparametrizations. Notice that, since the Finsler metric F is homogeneous of degree 1 and reversible, the length functional is invariant by the $\mathrm{Diff}(S^1)$-action, i.e. $L(\gamma) = L(\gamma \circ \theta)$ for all $\gamma \in \mathrm{Emb}(S^1, M)$ and $\theta \in \mathrm{Diff}(S^1)$.

We fix an auxiliary Riemannian metric g on M, and we will simply write $\| \cdot \|$ or $\| \cdot \|_g$ for its associated norm on tangent vectors. Since (M, g) is an orientable Riemannian surface, it admits a canonical complex structure $J \in \mathrm{End}(\mathrm{T}M)$, i.e. Jv is obtaining by rotating $v \in \mathrm{T}_x M$ by a positive angle of $\pi/2$ measured with g. The positive normal to $\gamma \in \mathrm{Emb}(S^1, M)$ is the vector field

$$N_\gamma(u) := \frac{1}{\|\dot{\gamma}(u)\|} J\dot{\gamma}(u),$$

where $\| \cdot \|$ is the Riemannian norm associated to g. We set

$$V_\gamma(u) := \frac{\left(\frac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma(u), \dot{\gamma}(u)) - F_x(\gamma(u), \dot{\gamma}(u))\right) N_\gamma(u)}{\|\dot{\gamma}(u)\|}. \tag{1.2}$$

In the expression of V_γ, the terms F_x and F_v denote the partial derivatives of F with respect to some local coordinates on M (or, more precisely, local coordinates on $\mathrm{T}M$ induced by local coordinates on M). However, the covector

$$\frac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma(u), \dot{\gamma}(u)) - F_x(\gamma(u), \dot{\gamma}(u)) \in \mathrm{T}^*_{\gamma(u)} M \tag{1.3}$$

is independent of the choice of local coordinates, and vanishes identically if and only if γ is a closed geodesic of (M, F). Since $\dot{\gamma}$ is always in its kernel, we actually conclude that V_γ vanishes identically if and only if γ is a closed geodesic of (M, F). Therefore, for each $\ell \geq \mathrm{injrad}(M, F)$ and $\epsilon > 0$, the open subset

$$\mathcal{U}(\ell, \epsilon) := \left\{ \gamma \in \mathrm{Emb}(S^1, M) \ \Big| \ L(\gamma) \in (\ell - \epsilon^2, \ell + \epsilon^2), \ \max_{s \in S^1} |V_\gamma(s)| < \epsilon \right\}. \tag{1.4}$$

is a neighborhood of the set of simple closed geodesics of (M, F) with length ℓ. We will employ the notation

$$\mathrm{Emb}(S^1, M)^{<\ell} := \{\gamma \in \mathrm{Emb}(S^1, M) \mid L(\gamma) < \ell\}$$

to denote the sublevel sets of the length functional.

Theorem 1.1. *Let (M, F) be a closed, orientable, reversible Finsler manifold, and $\rho_0 > 0$. There exists a continuous map*

$$\psi : [0, \infty) \times \mathrm{Emb}(S^1, M) \to \mathrm{Emb}(S^1, M), \qquad \psi(t, \gamma) = \psi_t(\gamma),$$

with the following properties:

   (i) *It is a semi-flow, i.e. $\psi_0 = \mathrm{id}$ and $\psi_{t_2} \circ \psi_{t_1} = \psi_{t_1 + t_2}$ for all $t_1, t_2 \geq 0$.*
   (ii) *It is equivariant with respect to the action of circle diffeomorphisms, i.e. $\psi_t(\gamma \circ \theta) = \psi_t(\gamma) \circ \theta$ for all $\gamma \in \mathrm{Emb}(S^1, M)$ and $\theta \in \mathrm{Diff}(S^1)$.*
   (iii) *The length function is not increasing along the trajectories of ψ_t. More precisely $\frac{\mathrm{d}}{\mathrm{d}t} L(\phi_t(\gamma)) \leq 0$ with equality if and only if γ is a closed geodesic or $L(\gamma) \leq \rho_0$.*
   (iv) *For each $\ell > 2\rho_0$ and $\epsilon > 0$ there exists $\delta \in (0, \epsilon^2)$ such that, for all $t > 0$ large enough, we have*

$$\psi_t(\mathrm{Emb}(S^1, M)^{<\ell+\delta}) \subset \mathcal{U}(\ell, \epsilon) \cup \mathrm{Emb}(S^1, M)^{<\ell-\delta}.$$

(v) *If there are no simple closed geodesics with length in $[\ell_1, \ell_2] \subset (2\rho_0, \infty)$, then for all $t > 0$ large enough*

$$\psi_t(\mathrm{Emb}(S^1, M)^{<\ell_2}) \subset \mathrm{Emb}(S^1, M)^{<\ell_1}.$$

Most of the points in this theorem follow from Oaks [Oak94], except point (iv), which is crucial for the applications.

1.2. **Closed geodesics on Finsler 2-spheres.** We already anticipated that the semi-flow of Theorem 1.1 allows to extend the celebrated Lusternik-Schnirelmann's theorem [LS34] to the reversible Finsler setting. Actually, it will also allow to extend the characterization of simple Zoll geodesic flows on the 2-sphere, originally claimed in the Riemannian case by Lusternik [Lju66] and rigorously proved in [MS18]. We recall that a Finsler manifold is called Zoll when all its unit-speed geodesics are closed with the same minimal period, and simple Zoll if, in addition, all the geodesics are simple closed. We denote by $\sigma_s(S^2, F)$ the simple length spectrum of a Finsler 2-sphere, which is the set of lengths of its simple closed geodesics.

Theorem 1.2. *Every reversible Finsler 2-sphere (S^2, F) has at least three geometrically distinct simple closed geodesics. More precisely:*

(i) *If $\sigma_s(S^2, F)$ is a singleton, then (S^2, F) is simple Zoll.*
(ii) *If $\sigma_s(S^2, F)$ contains exactly two elements, then there exists $\ell \in \sigma_s(S^2, F)$ such that every point of S^2 lies on a simple closed geodesic of (S^2, F) of length ℓ.*
(iii) *Assume that, for any compact interval $[\ell_1, \ell_2] \subset (0, \infty)$, (S^2, F) has only finitely many simple closed geodesics with length in $[\ell_1, \ell_2]$. Then, (S^2, F) has three simple closed geodesics $\gamma_1, \gamma_2, \gamma_3$ of (S^2, F) with lengths $L(\gamma_1) < L(\gamma_2) < L(\gamma_3)$ and such that, for each $i = 1, 2, 3$, γ_i has non-trivial local homology in degree i with with $\mathbb{Z}_2$ coefficients.*

For the definition of the local homology of a closed geodesic, we refer the reader to Section 4.3. Point (iii) in Theorem 1.2 may look technical, but it is a crucial ingredient for the proof of Theorem 1.3. Even though it is claimed in [Hin93], it does not have a proper proof in the published literature.

Finally, we can state the last result, that generalizes Bangert-Franks-Hingston's theorem.

Theorem 1.3. *Every reversible Finsler 2-sphere (S^2, F) has infinitely many geometrically distinct closed geodesics.*

The main ideas for this theorem remain the same as in the Riemannian case, but nevertheless we provide a full and rather self-contained account, which insures that certain arguments of the original proof that looked Riemannian can indeed be carried over in the Finsler case. At the same time, our treatment fills some expository gaps present in the literature.

1.3. **Organization of the paper.** In Section 2 we provide a construction of the curve shortening semi-flow, and prove Theorem 1.1. In Section 3 we prove Theorem 1.2, except the technical point (iii). In Section 4, we provide the background on the classical critical point theory for the Finsler energy function, and we will prove Theorem 1.2(iii) at the end of the section. Finally, in Section 5 we prove Theorem 1.3.

2. The curve shortening semi-flow

2.1. **The evolution equation.** We consider a 1-parameter family of curves $\gamma_t \in \mathrm{Emb}(S^1, M)$ evolving according to the partial differential equation

$$\partial_t \gamma_t(u) = w_t(u) n_t(u) \tag{2.1}$$

where $w_t := V_{\gamma_t}$ and $n_t := N_{\gamma_t}$. For every $\gamma_0 \in \mathrm{Emb}(S^1, M)$, we denote by $\tau_{\gamma_0} \in [0, \infty]$ the largest extended real number such that there is a well defined solution $\gamma_t \in \mathrm{Emb}(S^1, M)$ of (2.1) for all $t \in [0, \tau_{\gamma_0})$, with $\gamma_t|_{t=0} = \gamma_0$. We set

$$\mathcal{U} := \big\{ (t, \gamma_0) \mid \gamma_0 \in \mathrm{Emb}(S^1, M), \ t \in [0, \tau_{\gamma_0}) \big\}.$$

Theorem 2.1. *There is a well defined unique map*

$$\phi : \mathcal{U} \to \mathrm{Emb}(S^1, M), \qquad \phi(t, \gamma_0) = \phi_t(\gamma_0) = \gamma_t,$$

where γ_t is the solution of (2.1) *with initial condition γ_0, satisfying the following properties:*

   (i) *The subset $\mathcal{U} \subset [0, \infty) \times \mathrm{Emb}(S^1, M)$ is an open neighborhood of $\{0\} \times \mathrm{Emb}(S^1, M)$, and ϕ is continuous.*
   (ii) *The map ϕ is equivariant under the action of $\mathrm{Diff}(S^1)$ on $\mathrm{Emb}(S^1, M)$, i.e. $\phi_t(\gamma \circ \theta) = \phi_t(\gamma) \circ \theta$ for all $\gamma \in \mathrm{Emb}(S^1, M)$ and $\theta \in \mathrm{Diff}(S^1)$.*
   (iii) *For each $\gamma \in \mathrm{Emb}(S^1, M)$ we have $\frac{\mathrm{d}}{\mathrm{d}t} L(\phi_t(\gamma)) \leq 0$, with equality if and only if γ is a closed geodesic of (M, F).*
   (iv) *For each $\gamma \in \mathrm{Emb}(S^1, M)$, if*

$$\ell_\gamma := \lim_{t \to \tau_\gamma^-} L(\phi_t(\gamma)) > 0$$

   then $\tau_\gamma = \infty$.

The proof of this theorem will be carried over in the rest of the section: point (i) will be proved in Subsection 2.3; point (ii) is a consequence of Lemma 2.3; point (iii) will be proved in Subsection 2.2. The fact that ϕ is well-defined as a map of the above form (i.e. mapping the space $\mathrm{Emb}(S^1, M)$ into itself) and point (iv) will be proved in Section 2.4. In analogy with the analogous map in the Riemannian case, we call ϕ_t the **curve shortening semi-flow** of (M, F). Notice that ϕ_t is not a flow (despite in the Riemannian literature it is often called a flow): indeed, it is only defined for $t \geq 0$, and thus satisfies $\phi_{t_1} \circ \phi_{t_2} = \phi_{t_1 + t_2}$ only for $t_1, t_2 \geq 0$.

Every closed geodesics of a closed Finsler surface (M, F) have length strictly larger than the injectivity radius $\mathrm{injrad}(M, F)$. It is sometimes convenient to have a well defined curve shortening semi-flow defined for all positive times even for those trajectories that are not converging to a closed geodesic. We can achieve this by slowing down the curve shortening semi-flow lines in the sublevel set $\{L < \mathrm{injrad}(M, F)\}$, as follows. We fix $\rho_0 > 0$ (which will be chosen smaller than

injrad(M, F) in the applications), and consider a monotone increasing smooth function $\chi : [0, \infty) \to [0, 1]$ such that supp$(\chi) = [\rho_0, \infty)$ and $\chi(\ell) = 1$ for all $\ell \in [2\rho_0, \infty)$. We define

$$\psi : [0, \infty) \times \mathrm{Emb}(S^1, M) \to \mathrm{Emb}(S^1, M), \qquad \psi(t, \gamma_0) = \psi_t(\gamma_0) = \gamma_t,$$

where γ_t is the solution of the partial differential equation

$$\partial_t \gamma_t(u) = \chi(L(\gamma_t)) V_{\gamma_t}(u) N_{\gamma_t}(u) \tag{2.2}$$

The semi-flow ψ_t is the one that we employ for Theorem 1.1. Its properties, except Theorem 1.1(iv) and (v), will be direct consequences of the above Theorem 2.1 by means of the following lemma.

Lemma 2.2. *There exists a smooth function* $T : \mathrm{Emb}(S^1, M) \times [0, \infty) \to [0, \infty)$ *monotone increasing in the second variable such that* $T(\gamma, \cdot) < \tau_\gamma$ *and*

$$\psi_t(\gamma) = \phi_{T(\gamma, t)}(\gamma), \qquad \forall \gamma \in \mathrm{Emb}(S^1, M),\ t \in [0, \infty).$$

Moreover

  (i) $T(\gamma, t_1 + t_2) = T(\phi_{T(\gamma, t_1)}(\gamma), t_2)$,
  (ii) $T(\gamma, t) = t$ *if* $L(\phi_t(\gamma)) \geq 2\rho_0$,
  (iii) $T(\gamma, t) = 0$ *if* $L(\gamma) \leq \rho_0$,
  (iv) $T(\gamma \circ \theta, t) = T(\gamma, t)$ *for all* $\theta \in \mathrm{Diff}(S^1)$.

Proof. We denote $\gamma_0 := \gamma$ and $\gamma_t := \phi_t(\gamma_0)$. The smooth map $(s, t) \mapsto \gamma_{T(\gamma, t)}(s)$ is a solution of (2.2) if and only if

$$\chi(L(\gamma_{T(\gamma, t)})) V_{\gamma_{T(\gamma, t)}} N_{\gamma_{T(\gamma, t)}} = \partial_t \gamma_{T(\gamma, t)} = (\partial_t T(\gamma, t)) V_{\gamma_{T(\gamma, t)}} N_{\gamma_{T(\gamma, t)}}.$$

Therefore, the desired function $t \mapsto T(\gamma, t)$ is a solution of the ordinary differential equation

$$\begin{aligned} \partial_t T(\gamma, t) &= \chi(L(\gamma_{T(\gamma, t)})), \\ T(\gamma, 0) &= 0. \end{aligned} \tag{2.3}$$

This readily implies that T is smooth as a function of (γ, t), and not decreasing. Point (i) readily follows from the semi-flow property $\phi_{t_1 + t_2} = \phi_{t_1} \circ \phi_{t_2}$ of the curve shortening. If $L(\gamma_{T(\gamma, t)}) \geq 2\rho_0$, then $L(\gamma_{T(\gamma, t')}) \geq 2\rho_0$ and $\chi(L(\gamma_{T(\gamma, t')})) = 1$ for all $t' \in [0, t]$, which implies point (ii). If $L(\gamma) \leq \rho_0$, then $L(\gamma_t) \leq \rho_0$ and $\chi(L(\gamma_{T(\gamma, t)})) = 0$ for all $t \in (0, \tau_\gamma)$, which implies point (iii). Finally, if we set $T_\theta(\gamma, t) := T(\gamma \circ \theta, t)$ for some $\theta \in \mathrm{Diff}(S^1)$, we readily see that T_θ is also a solution of the ordinary differential equation (2.3). Since such equation has a unique solution, we have point (iv). $\qquad \square$

The function V_γ is a generalization of the Riemannian curvature of embedded curves on oriented Riemannian surfaces. Theorem 2.1(ii) readily follows from the following statement.

Lemma 2.3. *For each* $\theta \in \mathrm{Diff}(S^1)$, *we have*

$$N_{\gamma \circ \theta} = \mathrm{sign}(\dot\theta) N_\gamma \circ \theta, \qquad V_{\gamma \circ \theta} = \mathrm{sign}(\dot\theta) V_\gamma \circ \theta.$$

Proof. The statement concerning the normal vector N_γ is clear. Since the Finsler metric F is 1-homogeneous in the fibers $\mathrm{T}_q M$, we have $F_v(q, \lambda v) = F_v(q, v)$ for all

$\lambda > 0$. Moreover, we have $F_x(q, \lambda v) = \lambda F_x(q, v)$, $F_{qv}(q, \lambda v) = F_{qv}(q, v)$, $F_{vv}(q, v) = \lambda F_{vv}(q, \lambda v)$. Therefore, if we set $r = \theta(u)$,

$$
\begin{aligned}
V_{\gamma \circ \theta}(u) &:= \frac{\left(\frac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma(\theta(u)), \dot{\gamma}(\theta(u))) - \dot{\theta}(u) F_x(\gamma(\theta(u)), \dot{\gamma}(\theta(u)))\right) N_{\gamma \circ \theta}(u)}{\|\dot{\gamma}(\theta(u))\| \, |\dot{\theta}(u)|} \\
&= \frac{\dot{\theta}(u) \left(\frac{\mathrm{d}}{\mathrm{d}r} F_v(\gamma(r), \dot{\gamma}(r)) - F_x(\gamma(r), \dot{\gamma}(r))\right) N_{\gamma \circ \theta}(u)}{\|\dot{\gamma}(r)\| \, |\dot{\theta}(u)|} \\
&= \operatorname{sign}(\dot{\theta}(u)) \frac{\left(\frac{\mathrm{d}}{\mathrm{d}r} F_v(\gamma(r), \dot{\gamma}(r)) - F_x(\gamma(r), \dot{\gamma}(r))\right) N_\gamma(r)}{\|\dot{\gamma}(r)\|} \\
&= \operatorname{sign}(\dot{\theta}(s)) V_\gamma(\theta(u)). \qquad \qquad \square
\end{aligned}
$$

2.2. **The anti-gradient of the length.** Let $X \in \mathrm{T}_\gamma \mathrm{Emb}(S^1, M)$, that is, X is a smooth 1-periodic vector field along γ. The differential of L can be computed as

$$
\mathrm{d}L(\gamma)X = \int_0^1 \left(F_x(\gamma(u), \dot{\gamma}(u)) - \tfrac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma(u), \dot{\gamma}(u))\right) X(u) \, \mathrm{d}u. \tag{2.4}
$$

Lemma 2.4. *For each $a \in C^\infty(S^1, \mathbb{R})$, we have $\mathrm{d}L(\gamma)a\dot{\gamma} = 0$.*

Proof. If we set $\gamma_\epsilon(u) := \gamma(u + \epsilon a(u))$, we have $a\dot{\gamma} = \partial_\epsilon \gamma_\epsilon|_{\epsilon=0}$. Since, for all $|\epsilon|$ small enough, γ_ϵ is an embedded curve obtained by reparametrization of γ, we have $L(\gamma_\epsilon) = L(\gamma)$ and $\mathrm{d}L(\gamma)a\dot{\gamma} = \frac{\mathrm{d}}{\mathrm{d}\epsilon}\big|_{\epsilon=0} L(\gamma_\epsilon) = 0$. $\square$

The Riemannian metric g introduces an L^2 Riemannian metric on $\mathrm{Emb}(S^1, M)$ given by

$$
\langle\!\langle X, Y \rangle\!\rangle_\gamma = \int_{S^1} g(X(u), Y(u)) \|\dot{\gamma}(u)\| \, \mathrm{d}u, \qquad \forall X, Y \in \mathrm{T}_\gamma \mathrm{Emb}(S^1, M). \tag{2.5}
$$

Thanks to the factor $\|\dot{\gamma}(u)\|$ in the integrand, the inner product is invariant under the action of $\mathrm{Diff}(S^1)$, i.e.

$$
\langle\!\langle X \circ \theta, Y \circ \theta \rangle\!\rangle_{\gamma \circ \theta} = \langle\!\langle X, Y \rangle\!\rangle_\gamma, \qquad \forall \theta \in \mathrm{Diff}(S^1). \tag{2.6}
$$

We denote by ∇L the gradient of the length functional with respect to this inner product. Namely, ∇L is the vector field on $\mathrm{Emb}(S^1, M)$ given by

$$
\mathrm{d}L(\gamma)X = \langle\!\langle \nabla L(\gamma), X \rangle\!\rangle_\gamma.
$$

Lemma 2.5. $\nabla L(\gamma) = -V_\gamma N_\gamma$.

Proof. Consider an arbitrary $X \in \mathrm{T}_\gamma \mathrm{Emb}(S^1, M)$, which we can uniquely write as $X(u) = a(u)\dot{\gamma} + b(u)N_\gamma$, where $b(u) = g(X(u), N_\gamma(u))$. By Lemma 2.4 and Equation (2.4), we compute

$$
\begin{aligned}
\mathrm{d}L(\gamma)X &= \mathrm{d}L(\gamma)a\dot{\gamma} + \mathrm{d}L(\gamma)bN_\gamma = \mathrm{d}L(\gamma)bN_\gamma \\
&= \int_0^1 \left(F_x(\gamma(u), \dot{\gamma}(u)) - \tfrac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma(u), \dot{\gamma}(u))\right) b(u) \, N(u) \, \mathrm{d}u \\
&= \int_0^1 g(-V_\gamma(u) N_\gamma(u), X(u)) \|\dot{\gamma}(u)\| \, \mathrm{d}u. \qquad \square
\end{aligned}
$$

Therefore, the curve shortening equation (2.1) can be seen as the anti-gradient flow equation of L associated to the L^2-Riemannian metric on $\mathrm{Emb}(S^1, M)$, i.e.

$$
\partial_t \gamma_t = -\nabla L(\gamma_t). \tag{2.7}
$$

The invariance (2.6), together with Lemma 2.5, provides an alternative proof of Lemma 2.3. Moreover, if a solution γ_t is well defined for $t \in [a, b]$, then

$$L(\gamma_a) - L(\gamma_b) = \int_a^b \|\nabla L(\gamma_t)\|^2 \mathrm{d}t = \int_a^b \int_{S^1} V_{\gamma_t}(u)^2 \, \|\dot{\gamma}_t(u)\| \, \mathrm{d}u \, \mathrm{d}t. \qquad (2.8)$$

It is well known that the closed geodesics of (M, F) are critical points of L, that is, those γ such that $V_\gamma \equiv 0$. Therefore, $\partial_t L(\gamma_t) \le 0$ with equality if and only if γ_t is a closed geodesic of (M, F). This settles Theorem 2.1(iii).

Remark 2.6 (Alternative curve shortening)**.** The PDE (2.1) of the curve shortening is not canonically associated to the Finsler metric F, as it also involves the auxiliary Riemannian metric g. This choice of curve shortening semi-flow turns out to be the most convenient for the later computations. Alternatively, one could also study a curve shortening semi-flow whose definition does not involve an auxiliary Riemannian metric: this is done by replacing, in (2.7), the gradient ∇ with the one induced by the following Riemannian metric on $\mathrm{Emb}(S^1, M)$

$$\langle\!\langle X, Y \rangle\!\rangle'_\gamma = \int_{S^1} F(\gamma(u), \dot{\gamma}(u))(\tfrac{1}{2}F^2)_{vv}(\gamma(u), \dot{\gamma}(u))[X(u), Y(u)] \, \mathrm{d}u,$$

$$\forall X, Y \in \mathrm{T}_\gamma \mathrm{Emb}(S^1, M).$$

For each $v \in \mathrm{T}_q M$, we define v^F to be the positive orthogonal to v with respect to the inner product $(F^2)_{vv}(q, v)[\cdot, \cdot]$ with norm $F(q, v^F) = F(q, v)$. If we set

$$Z_\gamma(u) := \frac{\left(\frac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma(u), \dot{\gamma}(u)) - F_x(\gamma(u), \dot{\gamma}(u))\right)\dot{\gamma}(u)^F}{F(\gamma(u), \dot{\gamma}(u))},$$

the alternative curve shortening semi-flow is precisely given by

$$\partial_t \gamma_t(u) = \frac{Z_{\gamma_t}(u)}{F(\gamma_t(u), \dot{\gamma}_t(u))} \dot{\gamma}_t(u)^F. \qquad \qquad \square$$

2.3. **Short-time existence.** In order to prove Theorem 2.1(i), it is convenient to work in suitable local coordinates around a fixed curve $\gamma_0 \in \mathrm{Emb}(S^1, M)$. We denote by $\exp : \mathrm{T}M \to M$ the exponential map of (M, g). There exists $\rho > 0$ and an open set $U \subset M$ of $\gamma_0(S^1)$ such that the map

$$\xi : S^1 \times (-\rho, \rho) \to U, \qquad \xi(u, r) = \exp_{\gamma_0(u)}(r \, N_{\gamma_0}(u))$$

is a diffeomorphism.

We define the smooth map

$$\Xi : C^\infty(S^1, (-\rho, \rho)) \to \mathrm{Emb}(S^1, M), \qquad \Xi(z)(u) = \xi(u, z(u)).$$

Let us show that this map is open and injective. We first define the vector field N on U by

$$N(\xi(u, r)) = \frac{\mathrm{d}}{\mathrm{d}r} \xi(u, r) = \mathrm{d}\exp_{\gamma_0(u)}(r \, N_{\gamma_0}(u)) N_{\gamma_0}(u),$$

and notice that $\|N(q)\| = 1$ for all $q \in U$. Thus, we have $\mathrm{d}\Xi(z)w = W$, where

$$W(u) = w(u) N(\Xi(z)(u)),$$

and this latter vector field along $\Xi(z)$ is non-zero provided the function w is non-zero. Hence Ξ is an immersion. Clearly, Ξ is injective, for ξ is a diffeomorphism. Finally, the equality

$$\mathrm{dist}(\Xi(z)(u), \gamma_0(u)) = |z(u)|, \qquad \forall z \in C^\infty(S^1, (-\rho, \rho)), \ u \in S^1$$

implies that Ξ is an open map onto its image.

Since $\mathrm{Diff}(S^1)$ acts freely on $\mathrm{Emb}(S^1, M)$, the map

$$\Psi : C^\infty(S^1, (-\rho, \rho)) \times \mathrm{Diff}(S^1) \to \mathrm{Emb}(S^1, M),$$

$$\Psi(z, \theta)(u) = \Xi(z)(\theta(u)) = \exp_{\gamma_0(\theta(u))}\big(z(\theta(u)) N_{\gamma_0}(\theta(u))\big)$$

is open and injective onto a neighborhood of γ_0. The differential of Ψ is given by

$$\mathrm{d}\Psi(z, \theta)(v, \tau) = V,$$

where

$$V(u) = v(\theta(u)) N(\Psi(z, \theta)(u)) + \tau(\theta(u)) \Xi(z)^{\cdot}(\theta(u)).$$

Here, we have denoted $\Xi(z)^{\cdot}(u) := \frac{\partial}{\partial u}\Xi(z)(u)$

The map Ψ pulls-back the L^2 inner product (2.5) to

$$\langle\!\langle\!\langle (v, \tau), (w, \sigma) \rangle\!\rangle\!\rangle_{(z, \theta)}$$

$$:= \langle\!\langle \mathrm{d}\Psi(z, \theta)(v, \tau), \mathrm{d}\Psi(z, \theta)(w, \sigma) \rangle\!\rangle_{\Psi(z, \theta)}$$

$$= \int_{S^1} \Big(v(u)\, w(u) + v(u)\, \sigma(u)\, a_z(u) + w(u)\, \tau(u)\, a_z(u)$$

$$+ \tau(u)\, \sigma(u)\, b_z(u)^2 \Big) b_z(u)\, \mathrm{d}u,$$

where

$$a_z(u) := g(N(\Xi(z)(u)), \Xi(z)^{\cdot}(u)),$$

$$b_z(u) := \|\Xi(z)^{\cdot}(u)\|.$$

Notice that this inner product is actually independent of $\theta \in \mathrm{Diff}(S^1)$, and therefore we will simply write

$$\langle\!\langle\!\langle (v, \tau), (w, \sigma) \rangle\!\rangle\!\rangle_{(z, \theta)} = \langle\!\langle\!\langle (v, \tau), (w, \sigma) \rangle\!\rangle\!\rangle_z \tag{2.9}$$

In order to write expressions in local coordinates, let us pull-back the Finsler metric F by ξ. We obtain the Finsler metric $G := \xi^* F$ on $S^1 \times (-\epsilon, \epsilon)$ given by

$$G(q, v) = F(\xi(q), \mathrm{d}\xi(q)v), \qquad \forall q \in S^1 \times (-\rho, \rho),\ v \in \mathbb{R}^2.$$

The composition of the length functional L with Ψ reads

$$L \circ \Psi(z, \theta) = L \circ \Xi(z) = \int_{S^1} F\big(\tfrac{\mathrm{d}}{\mathrm{d}u}\Xi(z)(u)\big)\, \mathrm{d}u = \int_{S^1} G\big(\underbrace{(u, z(u))}_{q}, \underbrace{(1, \dot{z}(u))}_{v}\big)\, \mathrm{d}u.$$

Let us compute the derivative

$$\mathrm{d}(L \circ \Xi)(z)w = \int_{S^1} \big(G_{q_2}\, w + \partial_{v_2}G\, \dot{w}\big)\, \mathrm{d}u$$
$$= \int_{S^1} \big(G_{q_2} - G_{q_1 v_2} - G_{q_2 v_2}\, \dot{z} - G_{v_2 v_2}\, \ddot{z}\big)w\, \mathrm{d}u. \tag{2.10}$$

We denote by $(v_z, \tau_z) := \nabla(L \circ \Psi)(z)$ the gradient of $L \circ \Psi$ with respect to the inner product (2.9), i.e.

$$\langle\!\langle\!\langle (v_z, \tau_z), (w, \sigma) \rangle\!\rangle\!\rangle_z = \mathrm{d}(L \circ \Psi)(z, \theta)(w, \sigma).$$

Since $L \circ \Psi(z, \theta)$ is independent of $\theta \in \mathrm{Diff}(S^1)$, we have

$$0 = \langle\!\langle\!\langle (v_z, \tau_z), (0, \sigma) \rangle\!\rangle\!\rangle_z = \int_{S^1} \Big(v_z(u)\, a_z(u)\, b_z(u) + \tau_z(u)\, b_z(u)^3 \Big) \sigma(u)\, \mathrm{d}u,$$

which implies that

$$\tau_z(u) = -v_z(u)\, \frac{a_z(u)}{b_z(u)^2}, \qquad \forall u \in S^1.$$

On the other hand, we have

$$\begin{aligned}
\langle\!\langle\!\langle (v_z, \tau_z), (w, 0) \rangle\!\rangle\!\rangle_z &= \int_{S^1} \Big(v_z(u) + \tau_z(u)\, a_z(u) \Big)\, b_z(u)\, w(u)\, \mathrm{d}u \\
&= \int_{S^1} \Big(1 - \tfrac{a_z(u)^2}{b_z(u)^2} \Big) v_z(u)\, b_z(u)\, w(u)\, \mathrm{d}u \\
&= \mathrm{d}(L \circ \Xi)(z)w.
\end{aligned} \tag{2.11}$$

Notice that the quotient $a_z(u)/b_z(u)$ is well defined. Indeed, the curve $s \mapsto \Xi(z)(u)$ is transverse to the vector field N, and therefore

$$\tfrac{a_z(u)^2}{b_z(u)^2} = g\big(N(\Xi(z)(u)), \tfrac{\Xi(z)^{\cdot}(u)}{|\Xi(z)^{\cdot}(u)|_{\Xi(z)(u)}}\big)^2 < 1.$$

Equations (2.10) and (2.11) imply that

$$v_z = \Big(1 - \tfrac{a_{\dot z}^2}{b_{\dot z}^2} \Big)^{-1} b_z(u)^{-1} \big(G_{q_2} - G_{q_1 v_2} - G_{q_2 v_2}\, \dot z - G_{v_2 v_2}\, \ddot z \big).$$

The integral curves of the anti-gradient $-\nabla(L \circ \Xi)$ are solutions

$$(z, \theta) : [0, T) \times S^1 \to (-\rho, \rho) \times S^1$$

of the partial differential equation

$$\partial_t(z, \theta) = (-v_z, v_z a/b^2). \tag{2.12}$$

In particular, z is a solution of the partial differential equation

$$\partial_t z = \tfrac{b_z}{b_z^2 - a_z^2} \Big(G_{v_2 v_2}\, \partial_u^2 z + G_{q_2 v_2}\, \partial_u z + G_{q_1 v_2} - G_{q_2} \Big). \tag{2.13}$$

Since G is a Finsler metric, the second derivative $G_{vv}(q, v)$ is positive semidefinite and its kernel is generated by v. Therefore, $G_{v_2 v_2}((u, z(u)), (1, \dot z(u))) \neq 0$, and (2.13) is a parabolic partial differential equation. The local theory for this class of equations (see, e.g., [MM12]) provides the following statement.

Proposition 2.7. *For each $z_0 \in C^\infty(S^1, (-\rho, \rho))$ there exists $\epsilon > 0$ and a unique smooth solution $z : [0, \epsilon) \times S^1 \to (-\rho, \rho)$ of (2.13) such that $z(0, \cdot) = z_0$. Moreover, z depends continuously on the initial condition z_0 in the C^∞ topology.* $\square$

Assume that $z : [0, \epsilon) \times S^1 \to (-\rho, \rho)$ is the smooth solution given by Proposition 2.7. Up to reducing $\epsilon > 0$, we can easily find a smooth $\theta : [0, \epsilon) \times S^1 \to S^1$ such that (z, θ) is a solution of the curve shortening equation (2.12) with $\theta(0, \cdot) = \mathrm{id}$. Indeed, for each $s \in S^1$, such a θ is the unique smooth solution of the ordinary differential equation

$$\partial_t \theta(t, s) = -\tau_z(\theta(t, s)).$$

The smooth map

$$\gamma : [0, \epsilon) \times S^1 \to M, \qquad \gamma(t, s) = \Psi(z(t, \cdot), \theta(t, \cdot))(s) = \xi(\theta(t, s), z(t, \theta(t, s)))$$

is thus the unique smooth solution of the curve shortening equation (2.1) with $\gamma(0, \cdot) = \Xi(z) = \gamma_0$. Summing up, we have proved the following statement, which implies Theorem 2.1(i).

Theorem 2.8 (Local existence and uniqueness). *For each $\gamma_0 \in \mathrm{Emb}(S^1, M)$ there exists $\epsilon > 0$ and a unique smooth solution $\gamma : [0, \epsilon) \times S^1 \to M$ of the curve shortening equation* (2.1) *such that $\gamma(0, \cdot) = \gamma_0$. Moreover, γ depends continuously on the initial condition γ_0 in the C^∞ topology.* $\qquad\Box$

2.4. **Long-time existence.** We denote by SM the unit tangent bundle of M with respect to the auxiliary Riemannian metric g, i.e.

$$SM = \big\{ (x, v) \in \mathrm{T}M \mid \|v\| = 1 \big\}. \tag{2.14}$$

In order to prove that ϕ is well-defined as a map onto $\mathrm{Emb}(S^1, M)$, and that that there is long-time existence of solutions of the curve shortening equation (Theorem 2.1(iv)), it suffices to show that the factor w_t in the right-hand side of (2.1) can be expressed by means of a suitable smooth function

$$V : \mathbb{R} \times SM \to \mathbb{R}, \qquad V(\kappa_t(u), \gamma_t(u), \tau_t(u)),$$

and invoke the general results of Angenent [Ang90] and Oaks [Oak94]. Here, κ_t denotes the Riemannian curvature of γ_t measured with respect to the auxiliary Riemannian metric g, and $\tau_t(u) := \dot{\gamma}_t(u)/\|\dot{\gamma}_t(u)\|$ its unit tangent vector. By expanding the definition of w_t, we have

$$
\begin{aligned}
w_t &= \frac{\big(\frac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma_t, \dot{\gamma}_t) - F_x(\gamma_t, \dot{\gamma}_t)\big) n_t}{\|\dot{\gamma}_t\|} \\
&= F_{vv}(\gamma_t, \tau_t)[\ddot{\gamma}_t/\|\dot{\gamma}_t\|^2, n_t] + F_{xv}(\gamma_t, \tau_t)[\tau_t, n_t] - F_x(\gamma_t, \tau_t)n_t.
\end{aligned}
$$

Since $F_{vv}(x, v)v = 0$, the first summand in the last line can be rewritten as

$$
\begin{aligned}
F_{vv}(\gamma_t, \tau_t)[\ddot{\gamma}_t/\|\dot{\gamma}_t\|^2, n_t] &= F_{vv}(\gamma_t, \tau_t)[n_t, n_t]\, g(\ddot{\gamma}_t/\|\dot{\gamma}_t\|^2, n_t) \\
&= F_{vv}(\gamma_t, \tau_t)[n_t, n_t]\kappa_t - F_{vv}(\gamma_t, \tau_t)[n_t, n_t]g(\Gamma_{\gamma_t}[\tau_t, \tau_t], n_t).
\end{aligned}
$$

Here,

$$\Gamma_x[v, w] = \Gamma_{ij}^k(x)v_i w_j \partial_{x_k},$$

where Γ_{ij}^k are the Christoffel symbols of the metric g with respect to the local coordinates employed in the above expression. Inserting this into the expression of w_t, we obtain

$$
\begin{aligned}
w_t = {}&F_{vv}(\gamma_t, \tau_t)[n_t, n_t]\kappa_t - F_{vv}(\gamma_t, \tau_t)[n_t, n_t]g(\Gamma_{\gamma_t}[\tau_t, \tau_t], n_t) \\
&+ F_{xv}(\gamma_t, \tau_t)[\tau_t, n_t] - F_x(\gamma_t, \tau_t)n_t.
\end{aligned}
$$

Notice that the first summand $F_{vv}(\gamma_t, \tau_t)[n_t, n_t]\kappa_t$ is well defined independently of the local coordinates, as F_{vv} is simply the fiberwise Hessian of F. Therefore, since w_t is also well defined, the remaining summands

$$-F_{vv}(\gamma_t, \tau_t)[J\tau_t, J\tau_t]g(\Gamma_{\gamma_t}[\tau_t, \tau_t], J\tau_t) + F_{xv}(\gamma_t, \tau_t)[\tau_t, J\tau_t] - F_x(\gamma_t, \tau_t)$$

are well defined independently of the local coordinates as well. The expression above shows that w_t is of the form $w_t = V(\kappa_t, \gamma_t, \tau_t)$, where $V : \mathbb{R} \times SM \to \mathbb{R}$ is the smooth function

$$
\begin{aligned}
V(\kappa, x, v) = {}&F_{vv}(x, v)[Jv, Jv]\kappa - F_{vv}(x, v)[Jv, Jv]g(\Gamma_x[v, v], Jv) \\
&+ F_{xv}(x, v)[v, Jv] - F_x(x, v)Jv \\
=: {}&A(x, v)\kappa + B(x, v).
\end{aligned} \tag{2.15}
$$

The reversibility of F readily imply that $V(\kappa, x, v) = -V(-\kappa, x, -v)$. The function V thus satisfies in particular the assumptions required in [Ang90, Oak94]. By

[Ang91, Theorem 1.3], the map ϕ takes values inside $\mathrm{Emb}(S^1, M)$. Finally, Theorem 2.1(iv) follows from [Oak94, Corollary 6.2].

2.5. L^∞ **bounds on** V_γ. For any $\gamma_0 \in \mathrm{Emb}(S^1, M)$, we will write $\gamma_t = \phi_t(\gamma_0)$ for the corresponding solution of (2.1), and $\ell_t := L(\gamma_t)$ for its length. We denote by ∇_t, ∇_u, and ∇_s the covariant derivatives associated with the Levi-Civita connection of g. It is convenient to introduce the vector field $D_s = \|\dot\gamma_t(u)\|^{-1}\partial_u$ on $\mathbb{R} \times S^1$, which acts on smooth real-valued functions $f : \mathbb{R} \times S^1 \to \mathbb{R}$, $f(t, u) = f_t(u)$ by

$$D_s f_t(u) = \frac{\partial_u f_t(u)}{\|\dot\gamma_t(u)\|}.$$

We recall the classical Frenet formulas from plane Riemannian geometry:

$$\nabla_u \tau_t = \kappa_t \|\dot\gamma_t\| n_t, \qquad \nabla_u n_t = -\kappa_t \dot\gamma_t.$$

By means of the PDE (2.1), we also have the following formulas.

Lemma 2.9. $\nabla_t \tau_t = (D_s w_t) n_t$, $\nabla_t n_t = -(D_s w_t)\tau_t$.

Proof. Let us compute the covariant derivative $\nabla_t \tau_t$. Since $\|\tau_t\| = \|n_t\| \equiv 1$, we have

$$g(\nabla_t \tau_t, \tau_t) = g(\nabla_t n_t, n_t) = g(\nabla_u n_t, n_t) = g(\nabla_u \tau_t, \tau_t) = 0.$$

Moreover

$$g(\nabla_t \dot\gamma_t, n_t) = g(\nabla_u \partial_t \gamma_t, n_t) = g(\nabla_u(w_t n_t), n_t) = \dot w_t,$$

which readily implies

$$\nabla_t \tau_t = g(\nabla_t \tau_t, n_t) n_t = \frac{\dot w_t}{\|\dot\gamma_t\|} n_t = (D_s w_t) n_t,$$

$$\nabla_t n_t = g(\nabla_t n_t, \tau_t)\tau_t = -\frac{g(n_t, \nabla_t \dot\gamma_t)}{\|\dot\gamma_t\|}\tau_t = -\frac{\dot w_t}{\|\dot\gamma_t\|}\tau_t = -(D_s w_t)\tau_t. \qquad \square$$

Lemma 2.10. $\partial_t \|\dot\gamma_t(u)\| = -\kappa_t(u)w_t(u)\|\dot\gamma_t(u)\|$.

Proof. By means of the commutativity $\nabla_t \partial_u = \nabla_u \partial_t$ and of the PDE (2.1), we compute

$$\partial_t\|\dot\gamma_t(u)\| = \frac{g(\nabla_t \dot\gamma_t(u), \dot\gamma_t(u))}{\|\dot\gamma_t(u)\|} = \frac{g(\nabla_u \partial_t \gamma_t(u), \dot\gamma_t(u))}{\|\dot\gamma_t(u)\|} = \frac{g(\nabla_u(w_t n_t), \dot\gamma_t(u))}{\|\dot\gamma_t(u)\|}$$

$$= w_t(u)\frac{g(\nabla_u n_t, \dot\gamma_t(u))}{\|\dot\gamma_t(u)\|} = -w_t(u)\frac{g(\nabla_u \dot\gamma_t, n_t(u))}{\|\dot\gamma_t(u)\|} = -\kappa_t(u)w_t(u)\|\dot\gamma_t(u)\|.$$

$$\square$$

Lemma 2.11. *The curvature* κ_t *evolves according to the PDE*

$$\partial_t \kappa_t(u) = D_s^2 w_t(u) + w_t(u)\kappa_t^2(u) + w_t(u)k_g(\gamma_t(u)),$$

where k_g *denotes the Gaussian curvature of* (M, g), *i.e.* $k_g(x) = g(R(v, Jv)v, Jv)$ *for all* $v \in S_x M$.

Proof. The lemma follows by direct computation:

$$
\begin{aligned}
\partial_t \kappa_t &= \partial_t \frac{g(\nabla_u \tau_t, n_t)}{\|\dot{\gamma}_t\|} \\
&= \left(\partial_t \frac{1}{\|\dot{\gamma}_t\|}\right) \kappa_t \|\dot{\gamma}_t\| + \frac{1}{\|\dot{\gamma}_t\|} g(\nabla_t \nabla_u \tau_t, n_t) + \frac{1}{\|\dot{\gamma}_t\|} \underbrace{g(\nabla_u \tau_t, \nabla_t n_t)}_{=0} \\
&= \frac{\kappa_t w_t \|\dot{\gamma}_t\|}{\|\dot{\gamma}_t\|^2} \kappa_t \|\dot{\gamma}_t\| + \frac{1}{\|\dot{\gamma}_t\|} g(\nabla_u \nabla_t \tau_t, n_t) + \frac{1}{\|\dot{\gamma}_t\|} g(R(\dot{\gamma}_t, \partial_t \gamma_t)\tau_t, n_t) \\
&= w_t \kappa_t^2 + D_s^2 w_t + w_t k_g \circ \gamma_t.
\end{aligned}
$$ $\qquad \square$

We set $\ell_t := L(\gamma_t)$, and denote by $\Gamma_t : \mathbb{R}/\ell_t \mathbb{Z} \to M$ the reparametrized γ_t with unit speed with respect to the auxiliary Riemannian metric g. Namely $\Gamma_t(s) = \gamma_t \circ \sigma_t^{-1}(s)$, where

$$
\sigma_t(u) = \int_0^u \|\dot{\gamma}_t(r)\| \, \mathrm{d}r,
$$

and therefore $\ell_t = \sigma_t(1)$. We also set

$$
\begin{aligned}
W_t(s) &:= w_t \circ \sigma_t^{-1}(s), \\
N_t(s) &:= n_t \circ \sigma_t^{-1}(s), \\
K_t(s) &:= \kappa_t \circ \sigma_t^{-1}(s).
\end{aligned}
\tag{2.16}
$$

Notice that

$$
K_t(s) = g(\nabla_s \dot{\Gamma}_t, N_t), \qquad W_t(s) = V(K_t(s), \dot{\Gamma}_t(s)).
$$

Moreover

$$
\dot{W}_t \circ \sigma_t = D_s w_t, \qquad \dot{K}_t \circ \sigma_t = D_s \kappa_t.
$$

If $f : SM \to \mathbb{R}$ is a smooth function, we will denote by $\nabla^\mathrm{h} f$ and $\nabla^\mathrm{v} f$ the duals of the horizontal and vertical projections respectively of its gradient with respect to the Sasaki metric of SM induced by g. These operators allow to express $\partial_t w_t$ as

$$
\begin{aligned}
\partial_t w_t &= (\partial_\kappa V)\partial_t \kappa_t + (\nabla^\mathrm{h} V)\partial_t \gamma_t + (\nabla^\mathrm{v} V)\nabla_t \tau_t \\
&= (\partial_\kappa V)\big(D_s^2 w_t + w_t \kappa_t^2 + w_t k_g(\gamma_t)\big) + (\nabla^\mathrm{h} V)n_t \, w_t + (\nabla^\mathrm{v} V)n_t \, D_s w_t.
\end{aligned}
$$

We set

$$
A(x, v) := \partial_\kappa V(x, v) = F_{vv}(x, v)[Jv, Jv].
$$

Notice that A is uniformly bounded from below by a positive constant. From now on, we will consider it evaluated at $(\gamma_t(u), \tau_t(u))$. Notice that $A D_s^2 w_t = D_s(A D_s w_t) - (D_s A)(D_s w_t)$, and

$$
D_s A = \frac{1}{\|\dot{\gamma}_t\|}\big((\nabla^\mathrm{h} A)\dot{\gamma}_t + (\nabla^\mathrm{v} A)\nabla_u \tau_t\big) = (\nabla^\mathrm{h} A)\tau_t + (\nabla^\mathrm{v} A)n_t \, \kappa_t.
$$

Therefore, $\partial_t w_t$ can be written as

$$
\begin{aligned}
\partial_t w_t &= A D_s^2 w_t + A w_t \kappa_t^2 + B D_s w_t + C w_t \\
&= D_s(A D_s w_t) + A w_t \kappa_t^2 + E D_s w_t + H \kappa_t D_s w_t + C w_t,
\end{aligned}
$$

where B, C, E, and H are smooth functions on SM evaluated at $(\gamma_t(u), \tau_t(u))$.

We are now going to employ the open sets $\mathcal{U}(\ell, \epsilon)$ defined in (1.4).

Lemma 2.12. *For all $\ell > 0$ there exist constants $c > 0$ and $\epsilon_0 > 0$ with the following properties: for all $\epsilon \in (0, \epsilon_0]$, $\gamma_0 \in \mathcal{U}(\ell, \epsilon)$, and $t \geq 0$ such that $L(\gamma_0) - L(\gamma_t) \leq \epsilon^2$, we have $\|W_t\|_{L^2} \leq c\,\epsilon$.*

Proof. We consider an arbitrary $\gamma_0 \in \mathcal{U}(\ell, \epsilon)$, and its evolution γ_t. We recall that the corresponding w_t has the form $w_t = V(\kappa_t, \gamma_t, \tau_t)$, where $\tau_t := \dot{\gamma}_t / \|\dot{\gamma}_t\|$.

We compute

$$
\partial_t \|W_t\|_{L^2}^2 = \int_0^1 \Big(2w_t\,(\partial_t w_t)\,\dot{\sigma}_t + w_t^2(\partial_t \dot{\sigma}_t)\Big) \mathrm{d}u
$$

$$
= \int_0^1 \Big(2w_t\,D_s(AD_s w_t) + 2Aw_t^2\kappa_t^2 + 2E\,w_t\,D_s w_t + 2H\,\kappa_t w_t D_s w_t
$$

$$
+ 2Cw_t^2 - \kappa_t w_t^3 \Big)\dot{\sigma}_t\,\mathrm{d}u
$$

$$
= \int_0^{\ell_t} \Big(-2A(\dot{W}_t)^2 + 2AW_t^2 K_t^2 + 2EW_t\dot{W}_t + 2HK_t W_t \dot{W}_t
$$

$$
+ 2CW_t^2 - K_t W_t^3 \Big)\mathrm{d}s.
$$

From now on, we will denote by $c \geq 1$ a positive constant (independent of γ_t), that may increase on different inequalities. The above expression for $\partial_t \|W_t\|_{L^2}^2$ readily implies

$$
\partial_t \|W_t\|_{L^2}^2 \leq -c^{-1}\|\dot{W}_t\|_{L^2}^2 + c\left(\|W_t\dot{W}_t\|_{L^1} + \|K_t W_t \dot{W}_t\|_{L^1} + \|W_t^2 K_t^2\|_{L^1}\right.
$$
$$
\left. + \|K_t W_t^3\|_{L^1} + \|W_t\|_{L^2}^2\right). \tag{2.17}
$$

By the Peter-Paul inequality, for every $\rho > 0$, the term $\|W_t\dot{W}_t\|_{L^1}$ can be bounded as

$$
\|W_t\dot{W}_t\|_{L^1} \leq \rho^2\|\dot{W}_t\|_{L^2}^2 + \tfrac{1}{4\rho^2}\|W_t\|_{L^2}^2,
$$

and the term $\|K_t W_t \dot{W}_t\|_{L^1}$ as

$$
\|K_t W_t \dot{W}_t\|_{L^1} \leq \rho^2\|\dot{W}_t\|_{L^2}^2 + \tfrac{1}{4\rho^2}\|K_t W_t\|_{L^2}^2 \leq \rho^2\|\dot{W}_t\|_{L^2}^2 + \tfrac{1}{4\rho^2}\|K_t\|_{L^\infty}^2\|W_t\|_{L^2}^2.
$$

We will fix a sufficiently small constant $\rho > 0$ so that, in the inequality (2.17), the term $-c^{-1}\|\dot{W}_t\|_{L^2}^2$ will be able to absorb the terms $\rho^2\|\dot{W}_t\|_{L^2}^2$, still producing a negative factor in front of $\|\dot{W}_t\|_{L^2}^2$.

Equation (2.15) readily implies that the curvature K_t is related to W_t by $K_t = A^{-1}W_t + P$, where, once again, P is a smooth function on SM evaluated at $(\Gamma_t(s), \dot{\Gamma}_t(s))$. Therefore $\|K_t\|_{L^\infty} \leq c\,(\|W_t\|_{L^\infty}^2 + 1)$.

Inserting these estimates in (2.17), we obtain

$$
\partial_t \|W_t\|_{L^2}^2 \leq c\|W_t\|_{L^2}^2 + c\|W_t\|_{L^\infty}^2\|W_t\|_{L^2}^2 - c^{-1}\|\dot{W}_t\|_{L^2}^2.
$$

We require $\epsilon > 0$ to be small enough so that, since $\ell_0 - \ell_t \leq \epsilon^2$,

$$
c^{-1} \leq \ell_t \leq c.
$$

If we bound from above the term $-c^{-1}\|\dot{W}_t\|_{L^2}^2$ by means of the inequality

$$
\|W_t\|_{L^\infty}^2 \leq 2\ell_t^{-1}\|W_t\|_{L^2}^2 + 2\ell_t\|\dot{W}_t\|_{L^2}^2 \leq c\|W_t\|_{L^2}^2 + c\|\dot{W}_t\|_{L^2}^2,
$$

we further obtain

$$
\partial_t \|W_t\|_{L^2}^2 \leq c\|W_t\|_{L^2}^2 + c\|W_t\|_{L^\infty}^2(\|W_t\|_{L^2}^2 - c^{-1}).
$$

We claim that, if $\epsilon > 0$ is small enough (independently of γ), then $\|W_t\|^2_{L^2} < c^{-1}$ for all $t \geq 0$ such that $\ell_0 - \ell_t < \epsilon^2$. Indeed, assume that this is not the case. If $\epsilon^2 < 1/(c\ell_0)$, since $\|W_0\|^2_{L^2} \leq \ell_0 \epsilon^2 < c^{-1}$, there must be $\tau > 0$ such that $\|W_t\|^2_{L^2} < c^{-1}$ for all $t \in [0, \tau)$, $\|W_\tau\|^2_{L^2} = c^{-1}$, and $\ell_0 - \ell_\tau < \epsilon^2$. For all $t \in [0, \tau]$ we have the inequality $\partial_t \|W_t\|^2_{L^2} \leq c\|W_t\|^2_{L^2}$, and thus

$$c^{-1} = \|W_\tau\|^2_{L^2} \leq e^{c\tau} \|W_0\|^2_{L^2} \leq e^{c\tau}\epsilon^2 \ell_0.$$

If $\epsilon^2 \leq e^{-c} c^{-1} \ell_0^{-1}$, then $\tau \geq 1$. Therefore, since $\|W_\tau\|^2_{L^2} \leq e^{c(\tau-t)}\|W_t\|_{L^2}$ for all $t \in [0, \tau]$, by (2.8) we have

$$c^{-1} = \|W_\tau\|^2_{L^2} \leq e^c \int_{\tau-1}^\tau \|W_t\|^2_{L^2}\mathrm{d}t = e^c(\ell_{\tau-1} - \ell_\tau) \leq e^c \epsilon^2,$$

which is impossible if $\epsilon^2 < e^{-c}c^{-1}$.

Summing up, we showed that $\partial_t\|W_t\|^2_{L^2} \leq c\|W_t\|^2_{L^2}$ provided $\ell_0 - \ell_t \leq \epsilon^2$, and therefore

$$\|W_t\|^2_{L^2} \leq e^{ct} \int_0^t \|W_r\|^2_{L^2}\,\mathrm{d}r \leq e^c(\ell_0 - \ell_t) \leq e^c\epsilon^2,$$

$$\forall t \geq 0 \text{ such that } \ell_0 - \ell_t \leq \epsilon^2. \qquad \square$$

Lemma 2.13. *For all $\ell > 0$ there exist constants $c > 0$ and $\epsilon_0 > 0$ with the following properties: for all $\epsilon \in (0, \epsilon_0]$, $\gamma_0 \in \mathcal{U}(\ell, \epsilon)$, and $t \geq 0$ such that $L(\gamma_0) - L(\gamma_t) \leq \epsilon^2$, we have $\|\dot{W}_t\|_{L^2} \leq c\epsilon$.*

Proof. We consider an arbitrary $\gamma_0 \in \mathcal{U}(\ell, \epsilon)$ and its evolution γ_t. Once again, we will denote by $c \geq 1$ a large enough constant independent of γ_t and ϵ, possibly growing throughout the computations.

$$\begin{aligned}
\partial_t\|\dot{W}_t\|^2_{L^2} &= \int_0^1 \Big(2(D_s w_t)\partial_t(D_s w_t)\,\dot\sigma_t - (D_s w_t)^2 \kappa_t w_t \dot\sigma_t\Big)\mathrm{d}u \\
&= \int_0^1 \Big(2(D_s w_t)\partial_t \dot w_t + (D_s w_t)^2 \kappa_t w_t \dot\sigma_t\Big)\mathrm{d}u \\
&= \int_0^1 \Big(2(D_s w_t)D_s\big(A\,D_s^2 w_t + A w_t \kappa_t^2 + B\,D_s w_t + C\,w_t\big) \\
&\qquad + (D_s w_t)^2 \kappa_t w_t\Big)\dot\sigma_t \mathrm{d}u \\
&= \int_0^{\ell_t} \Big(-2A(\ddot W_t)^2 - 2A\ddot W_t W_t K_t^2 - 2B\ddot W_t \dot W_t - C\ddot W_t W_t \\
&\qquad + (\dot W_t)^2 K_t W_t\Big)\mathrm{d}s \\
&\leq -c^{-1}\|\ddot W_t\|^2_{L^2} + c\big(\|\ddot W_t W_t K_t^2\|_{L^1} + \|\ddot W_t \dot W_t\|_{L^1} + \|\ddot W_t W_t\|_{L^1} \\
&\qquad + \|(\dot W_t)^2 K_t W_t\|_{L^1}\big).
\end{aligned}$$

As before, by employing the Peter-Paul inequality and expressing K_t as an affine function of W_t, we have

$$
\begin{aligned}
\partial_t \|\dot{W}_t\|_{L^2}^2 \leq & -c^{-1}\|\ddot{W}_t\|_{L^2}^2 + c\big(\|W_t^3\|_{L^2}^2 + \|W_t^2\|_{L^2}^2 + \|W_t\|_{L^2}^2 + \|\dot{W}_t\|_{L^2}^2 \\
& + \|(\dot{W}_t)^2 W_t^2\|_{L^1} + \|(\dot{W}_t)^2 W_t\|_{L^1}\big) \\
= & -c^{-1}\|\ddot{W}_t\|_{L^2}^2 + c\big(\|W_t^6\|_{L^1} + \|W_t^4\|_{L^1} + \|W_t\|_{L^2}^2 + \|\dot{W}_t\|_{L^2}^2 \\
& + \|(\dot{W}_t)^2 W_t^2\|_{L^1} + \|(\dot{W}_t)^2 W_t\|_{L^1}\big).
\end{aligned}
\tag{2.18}
$$

Now, let $\epsilon > 0$ be small enough so that Lemma 2.12 holds, and $c^{-1} \leq \ell_t \leq c$. Therefore, if $\tau > 0$ is such that $\ell_0 - \ell_\tau < \epsilon^2$, we have $\|W_t\|_{L^2}^2 < c\,\epsilon^2$ for all $t \in [0, \tau]$. We introduce a large constant $d \geq 1$ that we will fix later. We set

$$
I := \big\{ t \in [0, \tau] \mid \|\dot{W}_t\|_{L^2}^2 \geq d\|W_t\|_{L^2}^2 \big\}.
$$

Notice that it is enough to prove the assertion of the lemma for all $t \in I$, and thus from now on we will fix one such t. By means of an integration by parts and Cauchy-Schwarz's inequality, we have

$$
\|\dot{W}_t\|_{L^2}^2 \leq -\int_0^{\ell_t} W_t \ddot{W}_t \, ds \leq \|W_t\|_{L^2}\|\ddot{W}_t\|_{L^2} \leq d^{-1}\|\dot{W}_t\|_{L^2}\|\ddot{W}_t\|_{L^2},
$$

and thus

$$
\|\dot{W}_t\|_{L^2} \leq d^{-1}\|\ddot{W}_t\|_{L^2}.
$$

We employ this inequality to bound from above the positive terms in (2.18) as follows.

$$
\begin{aligned}
\|W_t^4\|_{L^1} \leq & \|W_t\|_{L^2}^2 \|W_t\|_{L^\infty}^2 \leq c\,\epsilon^2\big(\|W_t\|_{L^1}^2 + \|\dot{W}_t\|_{L^1}^2\big) \leq c\,\epsilon^2\|\dot{W}_t\|_{L^2}^2, \\
\|W_t^6\|_{L^1} \leq & \|W_t\|_{L^2}^2 \|W_t\|_{L^\infty}^4 \leq c\,\epsilon^2\big(\|W_t\|_{L^1}^4 + \|\dot{W}_t\|_{L^1}^4\big) \\
& \leq c\,\epsilon^2\big(\|W_t\|_{L^1}^4 + \|W_t \ddot{W}_t\|_{L^1}^2\big) \leq c\,\epsilon^2\big(\|W_t\|_{L^1}^4 + \|W_t\|_{L^2}^2\|\ddot{W}_t\|_{L^2}^2\big) \\
& \leq c\,\epsilon^4\big(\|\dot{W}_t\|_{L^2}^2 + \|\ddot{W}_t\|_{L^2}^2\big), \\
\|(\dot{W}_t)^2 W_t\|_{L^1} \leq & \|W_t\|_{L^1}\|\dot{W}_t\|_{L^\infty}^2 \leq c\|W_t\|_{L^2}\|\ddot{W}_t\|_{L^2}^2 \leq c\,\epsilon\|\ddot{W}_t\|_{L^2}^2, \\
\|(\dot{W}_t)^2 W_t^2\|_{L^1} \leq & \|W_t\|_{L^2}^2\|\dot{W}_t\|_{L^\infty}^2 \leq c\,\epsilon^2\|\ddot{W}_t\|_{L^2}^2.
\end{aligned}
$$

We require $\epsilon > 0$ to be small enough so that the negative term $-c^{-1}\|\ddot{W}_t\|_{L^2}^2$ can absorb the terms $c\,\epsilon\|\ddot{W}_t\|_{L^2}^2$, $c\,\epsilon^2\|\ddot{W}_t\|_{L^2}^2$, $c\,\epsilon^4\|\ddot{W}_t\|_{L^2}^2$, thus obtaining

$$
\partial_t \|\dot{W}_t\|_{L^2}^2 \leq -c^{-1}\|\ddot{W}_t\|_{L^2}^2 + c\|\dot{W}_t\|_{L^2}^2 \leq (cd^{-1} - c^{-1})\|\ddot{W}_t\|_{L^2}^2, \qquad \forall t \in I.
$$

We now fix $d > c^2$, so that $(cd^{-1} - c^{-1}) < 0$ and

$$
\partial_t \|\dot{W}_t\|_{L^2}^2 \leq 0, \qquad \forall t \in I.
$$

Namely, for all $t \in [0, \tau]$, either $\|\dot{W}_t\|_{L^2}^2 \leq d\|W_t\|_{L^2}^2$ or $\partial_t \|\dot{W}_t\|_{L^2}^2 \leq 0$, and we conclude

$$
\max_{t \in [0, \tau]} \|\dot{W}_t\|_{L^2}^2 \leq d \max_{t \in [0, \tau]} \|W_t\|_{L^2}^2 \leq d\,c\,\epsilon^2. \qquad \square
$$

Lemmas 2.12 and 2.13 readily imply the following.

Lemma 2.14. *For all $\ell \geq \mathrm{injrad}(M, F)$ there exist constants $c \geq 1$ and $\epsilon_0 > 0$ with the following properties: for all $\epsilon \in (0, \epsilon_0]$, $\gamma \in \mathcal{U}(\ell, \epsilon)$, and $t \geq 0$ such that $L(\gamma) - L(\phi_t(\gamma)) \leq \epsilon^2$, we have $\|V_{\phi_t(\gamma)}\|_{L^\infty} < c\,\epsilon$.* $\qquad \square$

The analog of Theorem 1.1(iv) is well known in the classical setting of Lusternik-Schnirelmann theory: it is based on the fact that a smooth function drops of at least a fixed amount along a gradient flow line that crosses a given shell around a critical set at a given level. In the context of the Riemannian curve shortening semi-flow, the analogous property is claimed[1] by Grayson [Gra89]. We now employ the L^∞ bound of Lemma 2.14 to provide a complete proof of Theorem 1.1(iv) in our general Finsler setting.

Proof of Theorem 1.1(iv). Since the surface M is compact, we have a positive constant

$$a := \min_{(x,v)\in SM} \frac{\|v\|_g}{F(x,v)} > 0$$

where $\|\cdot\|_g$ is the norm associated to the auxiliary Riemannian metric g. For a given $\ell > 2\rho_0$, we consider the constants $c \geq 1$ and $\epsilon_0 \in (0,(\ell-2\rho_0)^{1/2})$ given by Lemma 2.14. For a given $\epsilon \in (0,\epsilon_0)$, we set

$$\delta := \frac{\epsilon^2}{4c^2}, \qquad \tau := \frac{1}{(\ell-\delta)a},$$

so that in particular $\mathcal{U}(\ell,(2\delta)^{1/2}) \subset \mathcal{U}(\ell,\epsilon)$. We fix an arbitrary $\gamma_0 \in \mathrm{Emb}(S^1,M)$ with $L(\gamma_0) < \ell + \delta$, and denote by $\gamma_t := \psi_t(\gamma_0)$, for $t \geq 0$, its evolution by means of the semi-flow ψ_t. Assume that

$$L(\gamma_\tau) \geq \ell - \delta. \tag{2.19}$$

We claim that there exists $\tau' \in [0,\tau]$ such that $\gamma_{\tau'} \in \mathcal{U}(\ell,(2\delta)^{1/2})$. Otherwise, we would have $\|V_{\gamma_t}\|_{L^\infty} \geq (2\delta)^{1/2}$ for all $t \in [0,\tau]$ and, by (2.8),

$$L(\gamma_\tau) = L(\gamma_0) - \int_0^\tau \int_{S^1} V_{\gamma_t}(u)^2 \|\dot\gamma(u)\|\,du\,dt \leq L(\gamma_0) - \tau 2\delta a L(\gamma_\tau) < \ell - \delta,$$

contradicting (2.19). By Lemma 2.14, for every $t \geq \tau'$ such that $L(\gamma_t) \geq \ell - \delta$ we have $\gamma_t \in \mathcal{U}(\ell,\epsilon)$. $\qquad\square$

2.6. **Compactness.** Finally Theorem 1.1(v) will be a consequence of the following compactness result. We denote by $\mathbb{P}\mathrm{T}M \to M$ the projectivized tangent bundle of M, whose fiber over any $x \in M$ is the real projective space

$$\mathbb{P}(\mathrm{T}_x M) = \frac{\mathrm{T}_x M \setminus \{0\}}{\sim}$$

where $[v] = [\lambda v]$ for all $v \in \mathrm{T}_x M \setminus \{0\}$ and $\lambda \in \mathbb{R}$.

Lemma 2.15. *Let $K \subseteq \mathbb{P}\mathrm{T}M$ be a compact subset. If no element of K is tangent to a simple closed geodesic of (M,F) of length ℓ, then for all $\epsilon > 0$ small enough no vector in K is tangent to some curve $\gamma \in \mathcal{U}(\ell,\epsilon)$.*

Proof. If the Lemma does not hold, then there exists a sequence $\gamma_n \in \mathcal{U}(\ell,1/n)$ such that $F(\gamma_n,\dot\gamma_n) \equiv L(\gamma_n)$ and $[\dot\gamma_n(0)] \in K$. The lifted curves $(\gamma_n,\dot\gamma_n/L(\gamma_n))$ are contained in the Finsler unit tangent bundle $\{(x,v) \in \mathrm{T}M \mid F(x,v) = 1\}$, which is a compact subset of $\mathrm{T}M$. We consider the function V_γ defined in (1.2). Since $\|V_{\gamma_n}\|_{L^\infty} < 1/n \to 0$ as $n \to \infty$, the sequence γ_n is bounded in the C^2-topology.

---

[1]Quoted from the last sentence in [Gra89, page 109]: *"Any curve leaving a small neighborhood of a geodesic shortens a fixed amount before moving very far."*

Therefore, up to a subsequence, γ_n converges in the C^1-topology to some γ such that $F(\gamma, \dot\gamma) \equiv L(\gamma) = \ell$ and $\dot\gamma(0) \in K$. Now, consider the Finsler energy

$$E : W^{1,2}(S^1, M) \to [0, \infty), \qquad E(\zeta) = \int_{S^1} F(\zeta(u), \dot\zeta(u))^2 \, \mathrm{d}u,$$

Since each γ_n has constant speed, we have

$$\mathrm{d}E(\gamma_n)X = 2 \int_{S^1} F(\gamma_n(u), \dot\gamma_n(u)) \big(F_x(\gamma_n(u), \dot\gamma_n(u)) - \tfrac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma_n(u), \dot\gamma_n(u)) \big) \, X(u) \, \mathrm{d}u$$

$$= 2L(\gamma_n) \, \mathrm{d}L(\gamma_n)X = 2L(\gamma_n) \int_{S^1} V_{\gamma_n}(u) \, g(N_{\gamma_n}(u), X(u)) \, \|\dot\gamma_n(u)\|_g \, \mathrm{d}u.$$

This, together with $\|V_{\gamma_n}\|_{L^\infty} \to 0$ and the fact that E is a $C^{1,1}$ function, readily implies that the limit curve γ is a critical point of E, and therefore a closed geodesic. In order to reach a contradiction, we simply have to show that γ is simple closed.

On an orientable reversible Finsler surface, a closed geodesic that is the C^1-limit of embedded circles is itself simple. Indeed, γ cannot have a transverse self-intersection, for the same would be true for γ_n for n large enough. Moreover, γ cannot have a self-tangency with opposite orientation, i.e. of the form $\gamma(u_1) = \gamma(u_2)$ and $\dot\gamma(u_1) = -\dot\gamma(u_2)$ for some $u_1 < u_2$; indeed, since F is reversible, we would have $\gamma(u_1 + r) = \gamma(u_2 - r)$ for all $r > 0$, and then $\dot\gamma(\frac{u_1+u_2}{2}) = 0$, contradicting the fact that γ is a geodesic. Finally, γ cannot be an iterated curve, i.e. of the form $\gamma(u) = \zeta(mu)$ for some simple closed geodesic $\zeta : S^1 \to M$ and $m \geq 2$; otherwise, since M is an orientable surface, a tubular neighborhood of ζ would be diffeomorphic to the annulus $S^1 \times (-1, 1)$, ζ being its zero section $S^1 \times \{0\}$; any closed curve sufficiently C^1-close to γ would wind $m \geq 2$ times around the annulus $S^1 \times (-1, 1)$, and therefore would have self-intersections. $\square$

Proof of Theorem 1.1(v). As in the proof of Theorem 1.1(iv)], we set

$$a := \min_{(x,v) \in SM} \frac{\|v\|_g}{F(x,v)} > 0.$$

By choosing $K = \mathbb{P}\mathrm{T}M$ in Lemma 2.15, we infer that there exists $\epsilon > 0$ such that $\mathcal{U}(\ell, \epsilon) = \varnothing$ for all $\ell \in [\ell_1, \ell_2]$. Namely, for all $\gamma_0 \in \mathrm{Emb}(S^1, M)$ with length $L(\gamma_0) \in [\ell_1, \ell_2]$ we have $\|V_{\gamma_0}\|_{L^\infty} \geq \epsilon$. We fix such a γ_0, and denote by $\gamma_t := \psi_t(\gamma_0)$ its evolution with the semi-flow ψ_t. Equation (2.8) implies that, for all $\tau \geq 0$ such that $L(\gamma_\tau) \geq \ell_1$, we have

$$L(\gamma_\tau) \leq L(\gamma_0) - \int_0^\tau \int_{S^1} V_{\gamma_t}(u)^2 \, \|\dot\gamma_t(u)\| \, \mathrm{d}u \, \mathrm{d}t \leq \ell_2 - \tau\epsilon^2 a L(\gamma_\tau),$$

and thus

$$\tau \leq \tau_0 \frac{\ell_2 - \ell_1}{\epsilon^2 a \ell_1}.$$

Therefore, for all $t > \tau_0$, we have $\psi_t(\mathrm{Emb}(S^1, M)^{<\ell_2}) \subset \mathrm{Emb}(S^1, M)^{<\ell_1}$. $\square$

3. Existence of simple closed geodesics

3.1. **Lusternik-Schnirelmann theory.** Let (M, F) be a closed, orientable, reversible, Finsler surface. We consider the space of unparametrized embedded loops

$$\Pi = \frac{\mathrm{Emb}(S^1, M)}{\mathrm{Diff}(S^1)}.$$

Here, $\mathrm{Diff}(S^1)$ acts by reparametrization on $\mathrm{Emb}(S^1, M)$. From now on, the length functional (1.1) will be considered as a continuous function on Π, i.e.

$$L : \Pi \to [0, \infty).$$

For any subset $\mathcal{W} \subset \Pi$ and $\ell \in (0, \infty]$, we set

$$\mathcal{W}^{<\ell} := \{\gamma \in \mathcal{W} \mid L(\gamma) < \ell\}. \tag{3.1}$$

Throughout this paper, we shall denote by $H_*(\,\cdot\,; \mathbb{F})$ the singular homology with coefficients in a field $\mathbb{F}$; we shall remove $\mathbb{F}$ from the notation whenever the arguments will not require a specific field. If σ is a singular chain in Π, we denote by $\mathrm{supp}(\sigma)$ its support, which is a compact subset of Π. Each non-zero homology class $h \in H_*(\Pi^{<b}, \Pi^{<a})$, where $0 < a < b \le \infty$, defines a min-max value

$$\ell(h) := \inf_{[\sigma]=h} \ \max L|_{\mathrm{supp}(\sigma)} \in [a, b).$$

Such value turns out to be the (positive) length of a simple closed geodesic of (M, F). This will be a rather direct consequence of the existence of the semi-flow of Theorem 1.1 and of the following statement. We will employ the open subsets $\mathcal{U}(\ell, \epsilon) \subset \mathrm{Emb}(S^1, M)$ defined in (1.4), which depend on an auxiliary Riemannian metric g on M. Since such open subsets are invariant under the action of $\mathrm{Diff}(S^1)$, we can consider their quotients

$$\mathcal{W}(\ell, \epsilon) := \frac{\mathcal{U}(\ell, \epsilon)}{\mathrm{Diff}(S^1)},$$

which are open subset in Π.

Lemma 3.1. *For each non-zero $h \in H_*(\Pi^{<b}, \Pi^{<a})$, the associated min-max $\ell = \ell(h)$ is the length of a simple closed geodesic of (M, F). For each $\epsilon > 0$ there exists $\delta \in (0, \epsilon^2)$ such that h can be represented by a relative cycle σ with*

$$\mathrm{supp}(\sigma) \subset \Pi^{<\ell-\delta} \cup \mathcal{W}(\ell, \epsilon). \tag{3.2}$$

Moreover, if there are only finitely many simple closed geodesics with length in $(\ell - \epsilon^2, \ell + \epsilon^2)$, there exists a simple closed geodesic γ of length ℓ such that, if we denote by $\mathcal{V}(\gamma, \epsilon)$ the connected component of $\mathcal{W}(\ell, \epsilon)$ containing γ, the inclusion induces a non-zero homomorphism $H_d(\mathcal{V}(\gamma, \epsilon), \mathcal{V}(\gamma, \epsilon)^{<\ell-\delta}) \to H_d(\Pi, \Pi^{<\ell})$.

Proof. We set $\rho_0 := a/3$, and consider the semi-flow ψ_t of Theorem 1.1. Since, by Theorem 1.1(ii), ψ_t is equivariant with respect to the action of $\mathrm{Diff}(S^1)$, it induces a well-defined continuous semi-flow on the quotient of its domain, which we still denote by $\psi_t : \Pi \to \Pi$. Given $\epsilon > 0$, we consider the associated $\delta \in (0, \epsilon^2)$ provided by Theorem 1.1(iv). By the definition of the min-max value $\ell := \ell(h)$, we can find a relative cycle σ representing h and such that $\max L|_{\mathrm{supp}(\sigma)} < \ell + \delta$. For each $t > 0$, the relative cycle $(\psi_t)_*\sigma$ still represents h. By Theorem 1.1(iv), if we choose $t > 0$ large enough, we have $\mathrm{supp}((\psi_t)_*\sigma) \subset \Pi^{<\ell-\delta} \cup \mathcal{W}(\ell, \epsilon)$. This proves (3.2).

Now, assuming by contradiction that ℓ is not the length of a simple closed geodesic of (M, F), by choosing $K = \mathbb{P}\mathrm{T}M$ in Lemma 2.15 we would have that $\mathcal{W}(\ell, \epsilon) = \varnothing$ for all $\epsilon > 0$ small enough. However, by the result of the previous paragraph, this would allow us to find a relative cycle σ representing h and such that $\mathrm{supp}(\sigma) \subset \Pi^{<\ell-\delta}$, contradicting the definition of $\ell = \ell(h)$.

We are left to prove the moreover part of the statement. For that, notice that we can assume that $\epsilon > 0$ is arbitrarily small (if the theorem holds for some ϵ, it

also holds for larger values of ϵ). In particular, we assume that ϵ is small enough so that, by our assumptions, there are only finitely many simple closed geodesics $\gamma_1, ..., \gamma_k$ with length in the interval $[\ell - \epsilon^2, \ell + \epsilon^2]$, and they all have length ℓ. We denote by $\mathcal{V}_{\epsilon,i}$ the connected component of $\mathcal{W}_\epsilon := \mathcal{W}(\ell, \epsilon)$ containing γ_i. If needed, we further lower $\epsilon > 0$, so that $\mathcal{V}_{\epsilon,i} \cap \mathcal{V}_{\epsilon,j} = \varnothing$ if $i \neq j$. We set

$$\mathcal{V}_\epsilon := \mathcal{V}_{\epsilon,1} \cup ... \cup \mathcal{V}_{\epsilon,k}.$$

We can also lower $\delta > 0$ so that $a \leq \ell - \delta$. The inclusions induce the commutative diagram

$$
\begin{array}{ccc}
H_*(\Pi^{<\ell-\delta} \cup \mathcal{W}_\epsilon, \Pi^{<a}) & \xrightarrow{\ i_* \ } & H_*(\Pi, \Pi^{<a}) \\
\downarrow & & \downarrow{\scriptstyle j_*} \\
H_*(\Pi^{<\ell-\delta} \cup \mathcal{W}_\epsilon, \Pi^{<\ell-\delta}) & \longrightarrow & H_*(\Pi, \Pi^{<\ell}) \\
{\scriptstyle \cong}\uparrow & \nearrow{\scriptstyle k_*} & \\
H_*(\mathcal{W}_\epsilon, \mathcal{W}_\epsilon^{<\ell-\delta}) & &
\end{array}
$$

The homology class h is contained in the image of i_* according to (3.2), and the lower vertical arrow is an isomorphism by excision. Moreover, by the very definition of the min-max value $\ell = \ell(h)$, we have that $j_*(h) \neq 0$. Overall, this shows that the homomorphism k_* is non-zero.

We set $\mathcal{W}'_\epsilon := \mathcal{W}_\epsilon \setminus (\mathcal{V}_{\epsilon,1} \cup ... \cup \mathcal{V}_{\epsilon,k})$, and claim that

$$\epsilon' := \inf \left\{ \|V_\gamma\|_{L^\infty} \ \big| \ \gamma \in \mathcal{W}'_\epsilon \right\} > 0,$$

where V_γ is the function defined in (1.2). Otherwise, we could find a sequence $\gamma_n \in \mathcal{W}'_\epsilon$ with $\|V_{\gamma_n}\|_{L^\infty} < 1/n$. As in the proof of Lemma 2.15, one can easily show that, up to extracting a subsequence, γ_n converges to a simple closed geodesic γ with length $L(\gamma) \in [\ell - \epsilon^2, \ell + \epsilon^2]$. But this would imply that $V_\gamma \equiv 0$ and $L(\gamma) = \ell$, and thus that $\gamma \in \mathcal{V}_{\epsilon,1} \cup ... \cup \mathcal{V}_{\epsilon,k}$, which is impossible since $\gamma_n \in \mathcal{W}'_\epsilon$ for all $n \in \mathbb{N}$.

Notice that $\mathcal{W}_{\epsilon'} \subset \mathcal{V}_\epsilon$, and once again the inclusion induces a non-zero homomorphism $H_*(\mathcal{W}_{\epsilon'}, \mathcal{W}_{\epsilon'}^{<\ell-\delta}) \to H_*(\Pi, \Pi^{<\ell})$, and therefore a non-zero homomorphism

$$\bigoplus_{i=1}^k H_d(\mathcal{V}_{\epsilon,i}, \mathcal{V}_{\epsilon,i}^{<\ell-\delta}) \to H_d(\Pi, \Pi^{<\ell}).$$

We denote by $I_\epsilon \subseteq \{1, ..., k\}$ the subset of those i such that the homomorphism $H_d(\mathcal{V}_{\epsilon,i}, \mathcal{V}_{\epsilon,i}^{<\ell-\delta}) \to H_d(\Pi, \Pi^{<\ell})$ is non-zero. Notice that $I_{\epsilon_1} \subseteq I_{\epsilon_2}$ if $0 < \epsilon_1 < \epsilon_2$. Therefore, there exists

$$i \in \bigcap_{\epsilon \in (0, \epsilon_0]} I_\epsilon,$$

and the simple closed geodesic γ_i satisfies the desired properties. $\qquad\square$

Assume now to have a homology class $w \in H_{d+i}(\Pi, \Pi^{<\rho})$ and a cohomology class $w \in H^i(\Pi)$ whose cap product $w \frown h \in H_d(\Pi, \Pi^{<\rho})$ is non-zero. Given any relative cycle σ representing h we can produce a relative cycle σ' representing $w \frown h$ and such that $\mathrm{supp}(\sigma') \subset \mathrm{supp}(\sigma)$. This readily implies that

$$\ell(w \frown h) \leq \ell(h).$$

We can now state a version of the classical Lusternik-Schnirelmann theorem for the length functional.

Theorem 3.2. *If $w \frown h \neq 0$ and $\ell(w \frown h) = \ell(h)$, then for every $\epsilon > 0$ we have $w|_{\mathcal{W}(\ell(h),\epsilon)} \neq 0$ in $H^*(\mathcal{W}(\ell(h),\epsilon))$.*

Proof. Let $\epsilon > 0$ be given, and set $\ell := \ell(h)$ and $\mathcal{W} := \mathcal{W}(\ell,\epsilon)$. By Lemma 3.1, h can be represented by a relative cycle σ such that $\mathrm{supp}(\sigma) \subset \mathcal{W} \cup \Pi^{<\ell}$. By applying sufficiently many barycentric subdivisions to the singular simplexes in σ, we can assume that the relative cycle decomposes as $\sigma = \sigma' + \sigma''$, where σ' and σ'' are chains with $\mathrm{supp}(\sigma') \subset \Pi^{<\ell}$ and $\mathrm{supp}(\sigma'') \subset \mathcal{W}$. Let $w \in H^*(\Pi)$ be a cohomology class such that $w|_{\mathcal{W}} = 0$ in $H^*(\mathcal{W})$ and $w \frown h \neq 0$. The cohomology long exact sequence of the pair $\mathcal{W} \subset \Pi$ provides a relative cocycle μ representing w that vanishes on all singular simplexes contained in $\mathcal{W}$. This implies

$$w \frown h = [\mu \frown (\sigma' + \sigma'')] = [\mu \frown \sigma'].$$

Namely, $w \frown h$ is represented by the relative cycle $\mu \frown \sigma'$ whose support is contained in the sublevel set $\Pi^{<\ell}$, which implies that $\ell(w \frown h) < \ell$. $\qquad\square$

3.2. **Topology of the space of embedded circles on the 2-sphere.** Once the results of the previous subsection are established, the proofs of points (i) and (ii) in Theorem 1.2 are analogous to ones of the Riemannian case in [MS18]. In this subsection, we provide the arguments for the reader's convenience. We will adopt the notation of the previous section, with M equal to the unit sphere $S^2 \subset \mathbb{R}^3$. It will be crucial to consider the singular homology H_* with coefficients in $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$, and therefore we will include it in the notation.

We first recall, from [Bal78, MS18], some basic information concerning the topology of the space of its unparametrized embedded loops Π. It is convenient to slightly enlarge this space as follows: we denote by Π_0 the space of constant loops on S^2, and set $\overline{\Pi} := \Pi \cup \Pi_0$. We endow $\overline{\Pi}$ with the quotient C^∞-topology as a subspace of $C^\infty(S^1, M)/\mathrm{Diff}(S^1)$. The relevant topology of $\overline{\Pi}$, at least for what concerns the application to Theorem 1.2, is provided by the subspace of round circles. More precisely, let

$$E = \big\{ ([x], \lambda x) \in \mathbb{RP}^2 \times \mathbb{R}^3 \mid x \in S^2, \ \lambda \in [-1,1] \big\}.$$

Namely, E is the total space of the canonical line bundle $\pi : E \to \mathbb{RP}^2$ with fiber $[-1,1]$. We consider the embedding

$$\iota : E \to \overline{\Pi}, \qquad \iota(e) = \gamma_e,$$

where, if $e = ([x], y)$, $\gamma_e \in \overline{\Pi}$ is the (possibly constant) loop in the intersection of S^2 with the affine plane orthogonal to x and passing through y. The fundamental group of this space is $\pi_1(E) \cong \pi_1(\mathbb{RP}^2) \cong \mathbb{Z}_2$. Its cohomology ring with $\mathbb{Z}_2$ coefficient is given by $H^*(E; \mathbb{Z}_2) = \mathbb{Z}_2[u]/(u^3)$, where u is the generator of $H^2(E; \mathbb{Z}_2) \cong H^2(\mathbb{RP}^2; \mathbb{Z}_2) \cong \mathbb{Z}_2$. Moreover, by the Thom isomorphism theorem,

$$H^*(E, \partial E; \mathbb{Z}_2) \cong H^{*-1}(E; \mathbb{Z}_2) = \langle v, v \smile u, v \smile u^2 \rangle,$$

where $v \in H^1(E, \partial E)$ denotes the Thom class of the bundle $\pi : E \to \mathbb{RP}^2$. Since we work with $\mathbb{Z}_2$ coefficients, the homology becomes simply the dual of the cohomology, and in particular there exists $k_3 \in H_3(E, \partial E; \mathbb{Z}_2)$ such that $(v \smile u^2)k_3 = 1$. Therefore, we also have the classes $k_2 := u \frown k_3$ and $k_i = u \frown k_2$, and overall we have $H_*(E, \partial E; \mathbb{Z}_2) = \langle k_1, k_2, k_3 \rangle$.

It is not hard to show that the map ι induces an injective homomorphism of fundamental groups $\iota_* : \pi_1(E) \hookrightarrow \pi_1(\overline{\Pi})$, see [MS18, Section 2.2]. Since $\pi_1(E) \cong H_1(E; \mathbb{Z}_2) \cong \mathbb{Z}_2$ and since $H_1(\overline{\Pi}; \mathbb{Z}_2)$ is the tensor product of $\mathbb{Z}_2$ with the abelianization of $\pi_1(\overline{\Pi})$, then ι also induces an injective homomorphism of homology groups $\iota_* : H_1(E; \mathbb{Z}_2) \hookrightarrow H_1(\overline{\Pi}; \mathbb{Z}_2)$. Dually, ι induces a surjective cohomology homomorphism $\iota^* : H^1(\overline{\Pi}; \mathbb{Z}_2) \twoheadrightarrow H^1(E; \mathbb{Z}_2)$. We fix, once for all, a cohomology class $w \in H^1(\overline{\Pi}; \mathbb{Z}_2)$ such that $\iota^* w = u$. Since $\partial E = \Pi_0 \cong S^2$ is simply connected, the long exact sequence of homology groups readily implies that $H_1(E, \partial E; \mathbb{Z}_2) \cong H_1(E; \mathbb{Z}_2)$ and $H_1(\overline{\Pi}, \Pi_0; \mathbb{Z}_2) \cong H_1(\overline{\Pi}; \mathbb{Z}_2)$, and therefore ι induces an injective homomorphism of relative homology groups $\iota_* : H_1(E, \partial E; \mathbb{Z}_2) \hookrightarrow H_1(\overline{\Pi}, \Pi_0; \mathbb{Z}_2)$. This readily implies that $\iota_* k_2$ and $\iota_* k_3$ are both non-zero in $H_*(\overline{\Pi}, \Pi_0; \mathbb{Z}_2)$, since

$$ w^2 \frown \iota_* k_3 = w \frown \iota_*(u \frown k_3) = w \frown \iota_* k_2 = \iota_*(u \frown k_2) = \iota_* k_1 \neq 0. $$

Now, let us get rid of the space of constant curves Π_0. We recall that the systole $\mathrm{sys}(S^2, F)$ is the length of the shortest closed geodesic of (S^2, F).

Lemma 3.3. *For all $\rho \in (0, \mathrm{sys}(S^2, F))$, the inclusion $\Pi_0 \subset \overline{\Pi}^{<\rho}$ is a homotopy equivalence. Therefore, the inclusions induce the homology isomorphisms*

$$ H_*(\overline{\Pi}, \Pi_0; \mathbb{Z}_2) \xrightarrow[\cong]{j_*} H_*(\overline{\Pi}, \overline{\Pi}^{<\rho}; \mathbb{Z}_2) \xleftarrow[\cong]{l_*} H_*(\Pi, \Pi^{<\rho}; \mathbb{Z}_2). $$

Proof. We fix $\rho_2 \in (0, \mathrm{sys}(S^2, F))$, and we consider a small enough $\rho_1 \in (0, \rho_2)$ that we will fix later on. Since there are no simple closed geodesics of (S^2, F) with length in $[\rho_1, \rho_2]$, Theorem 1.1(v) (with choice of parameter $\rho_0 \in (0, \rho_1/2)$) implies that $\psi_t(\overline{\Pi}^{<\rho_2}) \subset \overline{\Pi}^{<\rho_1}$.

We recall that we consider S^2 as the unit sphere in $\mathbb{R}^3$. If $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^3$, we fix a sufficiently large constant $a \geq 1$ so that

$$ a^{-1}\|v\| \leq F(x, v) \leq a\|v\|, \qquad \forall (x, v) \in \mathrm{T}S^2. $$

For each $\gamma \in \overline{\Pi}$, we denote by $\widehat{\gamma} \in \mathbb{R}^3$ its average, i.e. if we parametrize γ with constant speed $F(\gamma, \dot\gamma) \equiv L(\gamma)$, we have

$$ \widehat{\gamma} := \int_{S^1} \gamma(u)\, \mathrm{d}u. $$

Notice that

$$ \max_{u \in S^1} \|\gamma(u) - \widehat{\gamma}\| \leq aL(\gamma)/2. $$

We require $\rho_1 < a^{-1}$, so that $\widehat{\gamma} \in A := B^2(3/2) \setminus B^2(1/2)$ for all $\gamma \in \overline{\Pi}^{<\rho_1}$, where $B^2(R) \subset \mathbb{R}^3$ denotes the open ball of radius R centered at the origin. We consider the radial projection $\pi : A \to S^2$, $\pi(x) = x/\|x\|$, and define the continuous homotopy

$$ r_t : \overline{\Pi}^{<\rho_1} \to \overline{\Pi}, \qquad r_t(\gamma)(u) = \pi((1-t)\gamma(u) + t\widehat{\gamma}). $$

Notice that the time-1 map is a retraction $r_1 : \overline{\Pi}^{<\rho_1} \to \Pi_0$. Moreover

$$ L(r_t(\gamma)) \leq a\|\mathrm{d}\pi\|_{L^\infty} \int_{S^1} \|\dot\gamma(u)\|\, \mathrm{d}u \leq a^2 \|\mathrm{d}\pi\|_{L^\infty} L(\gamma). $$

We require $\rho_1 < \rho_2 a^{-2}\|\mathrm{d}\pi\|_{L^\infty}^{-1}$, so that every r_t is a map of the form

$$ r_t : \overline{\Pi}^{<\rho_1} \to \overline{\Pi}^{<\rho_2}. $$

Therefore the inclusion $\Pi_0 \subset \overline{\Pi}^{<\rho}$ is a homotopy equivalence. The homology long exact sequence of the triple $\Pi_0 \subset \overline{\Pi}^{<\rho} \subset \overline{\Pi}$ readily implies that j_* is an isomorphism. Finally, the excision property implies that l_* is an isomorphism as well. $\square$

We consider the isomorphisms j_* and l_* provided by Lemma 3.3, and define

$$h_i := l_*^{-1} j_* \iota_* k_i \in H_*(\Pi, \Pi^{<\rho}; \mathbb{Z}_2), \qquad i = 1, 2, 3. \tag{3.3}$$

Notice that

$$w|_\Pi \frown h_{i+1} = h_i.$$

We denote by $E_0 \subset E$ the zero-section of the line bundle $\pi : E \to \mathbb{RP}^2$. Notice that ι restricts as a map of the form $\iota_0 := \iota|_{E_0} : E_0 \to \Pi$.

Lemma 3.4. *For each* $z \in H^2(\Pi; \mathbb{Z}_2)$ *such that* $\iota_0^* z \neq 0$ *in* $H^2(E_0; \mathbb{Z})$, *we have*

$$z \frown h_3 = h_1.$$

Proof. For each $r \in [0, 1]$, we consider the subset

$$E_r = \big\{ ([x], \lambda x) \in E \mid \lambda \in [-r, r] \big\}. \subset E,$$

Notice that this notation agrees with the definition of E_0 as the zero-section of the line bundle $\pi : E \to \mathbb{RP}^2$. We fix $r \in (0, 1)$ sufficiently close to 1 so that $\iota(\partial E_r) \subset \Pi^{<\rho}$. By deformation and excision, we have that the inclusions induce isomorphisms

$$H_*(E, \partial E; \mathbb{Z}_2) \xrightarrow{\cong} H_*(E, E \setminus E_0; \mathbb{Z}_2) \xleftarrow{\cong} H_*(E_r, E_r \setminus E_0; \mathbb{Z}_2) \xleftarrow{\cong} H_*(E_r, \partial E_r; \mathbb{Z}_2).$$

We denote by $k_i' \in H_*(E_r, \partial E_r; \mathbb{Z}_2)$ the image of $k_i \in H_*(E, \partial E; \mathbb{Z}_2)$ under the composition of the above isomorphisms. Notice that $k_1' = u|_{E_r} \frown k_3'$. The restriction $\iota_r = \iota|_{E_r} : E_r \to \Pi$ induces a homomorphism

$$(\iota_r)_* : H_*(E_r, \partial E_r; \mathbb{Z}_2) \to H_*(\Pi, \Pi^{<\rho}; \mathbb{Z}_2),$$

which allows to express the homology classes (3.3) as $h_i = (\iota_r)_* k_i'$. Since the inclusion $E_0 \subset E_r$ is a homotopy equivalence, a cohomology class $z \in H^2(\Pi; \mathbb{Z}_2)$ satisfies $\iota_0^* z \neq 0$ if and only if $\iota_r^* z \neq 0$, and thus if and only if $\iota_r^* z \frown k_3' = k_1'$. Therefore, if this is the case, we have

$$z \frown h_3 = z \frown (\iota_r)_* k_3' = (\iota_r)_* (\iota_r^* z \frown k_3') = (\iota_r)_* k_1' = h_1. \qquad \square$$

We set

$$\ell_i := \ell(h_i), \qquad i = 1, 2, 3. \tag{3.4}$$

Since every ℓ_i is the length of a simple closed geodesic of (S^2, F), if the simple length spectrum $\sigma_{\mathrm{s}}(S^2, F)$ is a singleton we have $\ell_1 = \ell_2 = \ell_3$. In this case Theorem 1.2(i) is a consequence of the following statement.

Theorem 3.5. *If* $\ell_1 = \ell_2 = \ell_3$, *then* (S^2, F) *is simple Zoll.*

Proof. We consider a circle bundle $\mathrm{pr} : P \to \Pi$, whose total space is given by $P = \{(\gamma, x) \in \Pi \times S^2 \mid x \in \gamma\}$ and whose projection is $\mathrm{pr}(\gamma, x) = \gamma$. We consider the projectivized tangent bundle

$$\mathbb{P}\mathrm{T}S^2 = \big\{ V_x \mid x \in S^2, \ V_x \text{ 1-dimensional vector subspace of } \mathrm{T}_x S^2 \big\},$$

and define the continuous evaluation map $\mathrm{ev}: P \to \mathbb{P}\mathrm{T}S^2$, $\mathrm{ev}(\gamma, x) = \mathrm{T}_x\gamma$. Since $\mathbb{P}\mathrm{T}S^2$ is a closed 3-manifold, we have $H^3(\mathbb{P}\mathrm{T}S^2; \mathbb{Z}_2) \cong \mathbb{Z}_2$, and we denote by m a generator of $H^3(\mathbb{P}\mathrm{T}S^2; \mathbb{Z}_2)$. We consider the pull-back bundle

$$P_0 = \iota_0^* P = \{(e, p) \in E_0 \times P \mid \iota_0(e) = \mathrm{pr}(p)\},$$

and the commutative diagram

$$
\begin{array}{ccc}
P_0 & \stackrel{\tilde{\iota}_0}{\hookrightarrow} P \stackrel{\mathrm{ev}}{\longrightarrow} & \mathbb{P}\mathrm{T}S^2 \\
\downarrow{\scriptstyle\mathrm{pr}} & \downarrow{\scriptstyle\mathrm{pr}} & \\
E_0 & \stackrel{\iota_0}{\hookrightarrow} \Pi &
\end{array}
$$

Here, $\tilde{\iota}_0(e, p) = p$ is the projection onto the second factor. Notice that $\mathrm{ev} \circ \tilde{\iota}_0$ is a homeomorphism. Moreover, since $H^3(E_0; \mathbb{Z}_2)$ and $H^4(E_0; \mathbb{Z}_2)$ are trivial, the Gysin sequence of the pull-back bundle $\mathrm{pr}: P_0 \to E_0$ readily implies that

$$\mathrm{pr}_* : H^3(P_0; \mathbb{Z}_2) \to H^2(E_0; \mathbb{Z}_2)$$

is an isomorphism. This implies that $\mathrm{pr}_* \tilde{\iota}_0^* \mathrm{ev}^* m \neq 0$ in $H^2(E_0; \mathbb{Z}_2)$. We set

$$z := \mathrm{pr}_* \mathrm{ev}^* m \in H^2(\Pi; \mathbb{Z}_2).$$

Since $\iota_0^* z = \mathrm{pr}_* \tilde{\iota}_0^* \mathrm{ev}^* m \neq 0$, Lemma 3.4 implies that $h_1 = z \frown h_3$.

Now, assume by contradiction that $\ell_1 = \ell_2 = \ell_3 =: \ell$, but there exists $(x, v) \in SS^2$ such that the geodesic $\gamma_{x,v}(t) := \exp_x(tv)$ is not a simple closed geodesic of minimal period ℓ (namely, $\gamma_{x,v}$ is not a closed geodesic, or it is closed but not simple closed, or it is simple closed but its length is not ℓ). By Lemma 2.15 there exists $\epsilon > 0$ small enough so that v is not tangent to any curve $\gamma \in \mathcal{W} := \mathcal{W}(\ell, \epsilon)$ passing through x. Namely, if we set $P' := \mathrm{pr}^{-1}(\mathcal{W})$, the restriction $\mathrm{ev}|_{P'} : P' \to \mathbb{P}\mathrm{T}S^2$ is not surjective. Since $\ell_1 = \ell_3$ and $h_1 = z \frown h_3$, Theorem 3.2 implies that $z|_{\mathcal{W}} \neq 0$ in $H^2(\mathcal{W}; \mathbb{Z}_2)$. However, since $z|_{\mathcal{W}} = (\mathrm{pr}|_{P'})_* \mathrm{ev}|_{P'}^* m$, this implies that the homomorphism

$$\mathrm{ev}|_{P'}^* : H^3(\mathbb{P}\mathrm{T}S^2; \mathbb{Z}_2) \to H^3(P'; \mathbb{Z}_2)$$

is non-zero, which is impossible since $\mathrm{ev}|_{P'}$ is not surjective. $\qquad\square$

If the simple length spectrum $\sigma_{\mathrm{s}}(S^2, F)$ contains exactly two elements, we must have $\ell_1 = \ell_2$ or $\ell_2 = \ell_3$. In this case Theorem 1.2(i) is a consequence of the following statement.

Theorem 3.6. *If $\ell_i = \ell_{i+1}$ for some $i \in \{1, 2\}$, then every point of S^2 lies on a simple closed geodesic of (S^2, F) of length ℓ_i.*

Proof. Assume by contradiction that $\ell := \ell_i = \ell_{i+1}$, but that some point $x \in S^2$ does not lie on a simple closed geodesic of length ℓ. We consider the subset $\mathcal{U} = \{\gamma \in \overline{\Pi} \mid x \notin \gamma\}$. It is easy to see that $\mathcal{U}$ is contractible: if we denote by $B^2 \subset \mathbb{R}^2$ the unit open ball, and we consider a homeomorphism $\theta : S^2 \setminus \{x\} \to B^2$, the homotopy $r_t : \mathcal{U} \to \mathcal{U}$, $t \in [0, 1]$, given by

$$r_t(\gamma) = \theta^{-1}((1 - t)\theta(\gamma))$$

defines a contraction of $\mathcal{U}$ onto a point curve in $\Pi_0 \cap \mathcal{U}$. In particular $H^1(\mathcal{U}; \mathbb{Z}_2)$ is trivial.

By applying Lemma 2.15 with $K = \mathbb{P}(\mathrm{T}_x S^2)$, we infer that there exists $\mathcal{W} = \mathcal{W}(\ell, \epsilon)$, for $\epsilon > 0$ small enough, such that none of the curves $\gamma \in \mathcal{W}$ passes through

x. Since $h_i = w|_\Pi \frown h_{i+1}$ and $\ell_i = \ell_{i+1}$, Theorem 3.2 implies that $w|_\mathcal{W} \neq 0$ in $H^1(\mathcal{W}; \mathbb{Z}_2)$. However, since $\mathcal{W} \subset \mathcal{U}$, we have $w|_\mathcal{U} \neq 0$ in $H^1(\mathcal{U}; \mathbb{Z}_2)$ as well, contradicting the conclusion of the previous paragraph. $\square$

4. Critical point theory of the energy functional

In this section we shall recall the background on the variational theory of Finsler closed geodesics. The reader can find more details and proofs in [Rad92, BL10, CJM11, AM18] and references therein. Throughout the section, we shall consider a Finsler manifold (M, F) of arbitrary dimension, except in certain statements where we will assume M to be a surface. The Finsler metric F is not required to be reversible, unless specifically stated.

4.1. **The energy functional.** We denote by $\Lambda = W^{1,2}(S^1, M)$ the free loop space of M of regularity $W^{1,2}$, and consider the energy functional

$$E : \Lambda \to [0, \infty), \qquad E(\gamma) = \int_{S^1} F(\gamma(u), \dot{\gamma}(u))^2 \, \mathrm{d}u.$$

Unlike in the Riemannian case, in the Finsler setting E is $C^{1,1}$, but possibly not C^2. Its critical points with positive critical value are precisely those $\gamma \in \Lambda$ that are closed geodesics of (M, F) parametrized with constant speed $F(\gamma, \dot{\gamma}) \equiv E(\gamma)^{1/2}$. For each $\gamma \in \Lambda$, we denote by

$$\gamma^m \in \Lambda, \qquad \gamma^m(t) = \gamma(mt)$$

its m-th iterate, whose energy is $E(\gamma^m) = m^2 E(\gamma)$. Clearly, iterates of critical points of E are again critical points. Identifying different iterates of the same closed geodesic detected with global variational methods is the crux of the matter in the closed geodesics problem.

The C^1 regularity of E is actually enough to define a smooth pseudo-gradient flow of E on Λ. It is well known that E satisfies the Palais-Smale condition with respect to a suitable complete Riemannian metric on Λ, and therefore we can perform the usual deformations of critical point theory. Since E is even $C^{1,1}$, it has a well define Gateaux Hessian $\mathrm{d}^2 E(\gamma)$ at every critical point. However, the C^2 regularity would be needed in order to apply the classical Morse-Gromoll-Meyer lemma [GM69a]. A simple way to circumvent the potential lack of C^2 regularity and, at the same time, work in a finite dimensional setting consists in employing Morse's finite dimensional approximations of Λ. We consider the (non-symmetric) Finsler distance

$$d : M \times M \to [0, \infty), \qquad d(x, y) = \min_\gamma \int_0^1 F(\gamma(u), \dot{\gamma}(u)) \, \mathrm{d}u, \qquad (4.1)$$

where the minimum ranges over all absolutely continuous curves $\gamma : [0, 1] \to M$ joining x and y. For each integer $k \geq 2$, we consider the space

$$\Lambda_k = \left\{ \boldsymbol{x} = (x_0, ..., x_{k-1}) \in M^{\times k} \ \middle| \ \sum_{i \in \mathbb{Z}_k} d(x_i, x_{i+1})^2 < \mathrm{injrad}(M, F)^2 \quad \forall i \in \mathbb{Z}_k \right\}.$$

We identify Λ_k with a subspace of Λ as follows: every $\boldsymbol{x} \in \Lambda_k$ corresponds to the curve $\gamma_{\boldsymbol{x}} \in \Lambda$ such that each restriction $\gamma_{\boldsymbol{x}}|_{[i/k, (i+1)/k]}$, for $i \in \mathbb{Z}_k$, is the shortest geodesic parametrized with constant speed joining x_i and x_{i+1}. In the following,

we will indistinctively write $\boldsymbol{x}$ or $\gamma_{\boldsymbol{x}}$ for the same object. The restriction of the energy to Λ_k has the form

$$E_k = E|_{\Lambda_k} : \Lambda_k \to \big[0, k\,\mathrm{injrad}(M, F)^2\big), \qquad E_k(\boldsymbol{x}) = k \sum_{i \in \mathbb{Z}_k} d(x_i, x_{i+1})^2.$$

Since the distance d is smooth away from the diagonal, E_k is smooth on the subspace of those $\boldsymbol{x}$ with $x_i \neq x_{i+1}$ for all $i \in \mathbb{Z}_k$. The critical points of E_k are precisely those $\boldsymbol{x}$ such that $\gamma_{\boldsymbol{x}}$ is a closed geodesic of (M, F) parametrized with constant speed and having energy $E_k(\boldsymbol{x}) = E(\gamma_{\boldsymbol{x}}) < k^2\mathrm{injrad}(M, F)^2$. In particular E_k is smooth on a sufficiently small neighborhood of its critical points with positive energy. Finally, for each compact interval $[a, b] \subset (-\infty, k\,\mathrm{injrad}(M, F)^2)$, the preimage $E_k^{-1}[a, b]$ is compact, which allows us to apply the gradient flow deformations from critical point theory.

4.2. **The Morse index and nullity.** Let $h : V \times V \to \mathbb{R}$ be a symmetric bilinear form on a vector space V. Its index $\mathrm{ind}(h)$ is defined as the supremum of the dimension of the subspaces $W \subset V$ such that $h|_W$ is negative definite. Its nullity $\mathrm{nul}(h)$ is defined as the dimension of $\ker(h) = \{v \in V \mid h(v, \cdot) = 0\}$. Notice that the sum $\mathrm{ind}(h) + \mathrm{nul}(h)$ is the supremum of the dimension of the subspaces $Z \subset V$ such that $h|_Z$ is negative semi-definite.

Let us consider a closed geodesic $\gamma \in \mathrm{crit}(E) \cap E^{-1}(0, \infty)$, and the associated Gateaux Hessian $h := \mathrm{d}^2 E(\gamma)$. The Morse index and nullity of γ are defined by

$$\mathrm{ind}(\gamma) := \mathrm{ind}(h), \qquad \mathrm{nul}(\gamma) := \mathrm{nul}(h) - 1.$$

It is well known that the indices are always finite. The reason for the -1 appearing in the definition of the nullity is that $\mathrm{nul}(h)$ is always larger than or equal to 1, as the vector field $\dot\gamma$ belongs to $\ker(h)$. If $x_0 := \gamma(0)$, we denote by

$$\Omega := \{\zeta \in \Lambda \mid \zeta(0) = x_0\}$$

the space of loops based at x_0. The critical points of $E|_\Omega$ are the geodesic loops, that is, those $\zeta \in \Lambda$ whose restriction $\zeta|_{(0,1)}$ is a geodesic parametrized with constant speed. The Morse index and nullity of γ in the based loop space are defined as

$$\mathrm{ind}_\Omega(\gamma) := \mathrm{ind}(h|_{\mathrm{T}_\gamma \Omega}), \qquad \mathrm{nul}_\Omega(\gamma) := \mathrm{nul}(h|_{\mathrm{T}_\gamma \Omega}).$$

The behavior of the Morse indices under iteration of the closed geodesic has been thoroughly studied since the seminal work of Bott [Bot56]. Without invoking Bott's theory, one has the following properties, which are rather immediate or can be proved as an exercise.

Lemma 4.1. *Let* (M, F) *be a Finsler manifold with a closed geodesic* $\gamma \in \mathrm{crit}(E)$. *The Morse indices of* γ *satisfy the following properties.*

(i) $\mathrm{ind}(\gamma) \geq \mathrm{ind}_\Omega(\gamma)$.
(ii) $\mathrm{ind}(\gamma) + \mathrm{nul}(\gamma) \geq \mathrm{ind}_\Omega(\gamma) + \mathrm{nul}_\Omega(\gamma)$.
(iii) *If* $\mathrm{ind}(\gamma) > 0$, *then* $\mathrm{ind}(\gamma^m) \to \infty$ *as* $m \to \infty$.
(iv) $\mathrm{ind}(\gamma^m) \geq \mathrm{ind}(\gamma)$ *and* $\mathrm{nul}(\gamma^m) \geq \mathrm{nul}(\gamma)$ *for all* $m \in \mathbb{N}$.
(v) $\mathrm{ind}_\Omega(\gamma^m) \geq m\,\mathrm{ind}_\Omega(\gamma)$ *and* $\mathrm{ind}_\Omega(\gamma^m) + \mathrm{nul}_\Omega(\gamma^m) \geq m\,(\mathrm{ind}_\Omega(\gamma) + \mathrm{nul}_\Omega(\gamma))$ *for all* $m \in \mathbb{N}$. $\hfill\square$

The following proposition summarizes those subtler results concerning the Morse indices of closed geodesics that we will need in the proof of Theorem 1.3. In the literature, most of these results are proved in the Riemannian setting: points (i–iv)

can be found in [BTZ82], point (vi) in [Kli95], and point (vii) in [Ban93]. In the Finsler setting, the differences in the proofs are essentially cosmetic, but we include them for the reader's convenience.

Proposition 4.2. *Let (M, F) be an orientable Finsler manifold, and $\gamma \in \mathrm{crit}(E) \cap E^{-1}(0, \infty)$ a closed geodesic. The indices of γ satisfy the following properties.*

(i) $\mathrm{nul}(\gamma) \leq 2 \dim(M) - 2$.
(ii) $\mathrm{nul}_\Omega(\gamma) \leq \dim(M) - 1$.
(iii) $\mathrm{ind}(\gamma) \leq \mathrm{ind}_\Omega(\gamma) + \dim(M) - 1$.
(iv) $\mathrm{ind}(\gamma) + \mathrm{nul}(\gamma) \leq \mathrm{ind}_\Omega(\gamma) + \mathrm{nul}_\Omega(\gamma) + \dim(M) - 1$.

Moreover, if M is an orientable surface, they further satisfy the following properties.

(v) *If $\mathrm{nul}_\Omega(\gamma) = 1$ then $\mathrm{ind}_\Omega(\gamma^m) = m \, \mathrm{ind}_\Omega(\gamma) + m - 1$ and $\mathrm{nul}_\Omega(\gamma^m) = \mathrm{nul}_\Omega(\gamma)$ for all $m \in \mathbb{N}$,*
(vi) *If $\mathrm{nul}(\gamma) = 2$ then $\mathrm{nul}(\gamma^m) = 2$ and $\mathrm{ind}(\gamma^m)$ is odd for all $m \in \mathbb{N}$.*
(vii) *If $\mathrm{ind}_\Omega(\gamma^m) > 0$ for some $m \geq 1$, then $\mathrm{ind}(\gamma) > 0$ and indeed there exists a nowhere vanishing smooth vector field ζ along γ that is 1-periodic, everywhere transverse to $\dot\gamma$, and such that $\mathrm{d}^2 E(\gamma)(\zeta, \zeta) < 0$.*

Proof. We can assume without loss of generality that $E(\gamma) = 1$, so that $F(\gamma, \dot\gamma) \equiv 1$. We set
$$G : \mathrm{T}M \to [0, \infty), \qquad G(x, v) = \tfrac{1}{2} F(x, v)^2,$$
which is a $C^{1,1}$ function, smooth outside the zero-section, and fiberwise positively homogeneous of degree 2. The function G defines a 1-form λ on $\mathrm{T}M$ by
$$\lambda_{(x,v)}(w) = G_v(x, v) \circ \mathrm{d}\pi(x, v)w, \qquad \forall w \in \mathrm{T}_{(x,v)}(TM),$$
which is the dual of the classical Liouville one form on T^*M. The 2-form $-\mathrm{d}\lambda$ is a symplectic form on the complement of the zero-section in $\mathrm{T}M$. We treat G as a Hamiltonian, and consider its associated Hamiltonian vector field X defined by $-\mathrm{d}\lambda(X, \cdot) = \mathrm{d}G$. We denote by $\phi_t : \mathrm{T}M \to \mathrm{T}M$ the associated Hamiltonian flow of X. Its flow lines are the speed vectors of the geodesics of (M, F) parametrized with constant speed. In particular, the curve $\tilde\gamma(t) := (\gamma(t), \dot\gamma(t))$ is the periodic orbit of ϕ_t corresponding to the closed geodesic γ. Since G is autonomous, the Hamiltonian flow ϕ_t preserves each level set $G^{-1}(\ell^2)$. The energy level of $\tilde\gamma$ is
$$G(\tilde\gamma(t)) = \tfrac{1}{2} F(\tilde\gamma(t))^2 = 1/2,$$
and we denote by $SM := G^{-1}(1/2) = F^{-1}(1)$ the corresponding energy hypersurface, which is the unit tangent bundle of (M, F). The 1-form λ restricts to a contact form $\alpha := \lambda|_{SM}$, and X restricts to the Reeb vector field of (SM, α). Namely $\alpha(X) = 1$ and $\mathrm{d}\alpha(X, \cdot) = 0$. In particular $\phi_t^* \alpha = \alpha$. Notice that
$$\mathrm{d}\pi(\tilde\gamma(t))\xi_{\tilde\gamma(t)} = \ker(G_v(\tilde\gamma(t))).$$
Let L be the vector field on TM defined by
$$L(x, v) = \tfrac{\mathrm{d}}{\mathrm{d}t}\big|_{t=1}(x, tv).$$
This is a Liouville vector field, meaning that $\mathrm{d}\lambda(L, \cdot) = \lambda$, and is transverse to SM. The vector bundle $\mathrm{TT}M$ splits as a direct sum
$$\mathrm{TT}M = \mathrm{span}\{X, L\} \oplus \xi, \tag{4.2}$$

and this decomposition is symplectically orthogonal, i.e.

$$d\lambda(V, W) = 0, \qquad \forall V \in \mathrm{span}\{X, L\}, \ W \in \xi.$$

We denote by $\Phi_t := d\phi_t(\tilde\gamma(0)) : T_{\tilde\gamma(0)}TM \to T_{\tilde\gamma(t)}TM$ the linearized Hamiltonian flow along $\tilde\gamma$. Its flow lines have the form

$$\dot{\tilde\zeta}(t) = \Phi_t(\tilde\zeta(0)) = (\zeta(t), \dot\zeta(t)).$$

where ζ is a Jacobi vector field along γ, that is, a solution of the Jacobi equation

$$\partial_t(G_{vv}\dot\zeta + G_{xv}\zeta) - G_{xx}\zeta - G_{vx}\dot\zeta = 0,$$

which is the linearization of the Hamiltonian equation $\dot{\tilde\gamma} = X \circ \tilde\gamma$ at γ. Here and in the following, the second derivatives G_{xx}, G_{xv}, G_{vx}, G_{vv} are meant to be evaluated at $\tilde\gamma(t)$. The linearized flow Φ_t preserves the splitting (4.2). Indeed, $\phi_t^*\alpha = \alpha$ implies that $\Phi_t(\xi) = \xi$. Moreover, $\Phi_t(X) = X$ and $\Phi_t(L) = tX + L$, that is, $\Phi_t|_{\mathrm{span}\{X,L\}}$ can be written in the frame X, L as the symplectic matrix

$$\Phi_t|_{\mathrm{span}\{X,L\}} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \in \mathrm{Sp}(2). \tag{4.3}$$

Let $W^{1,2}(S^1, \gamma^*TM)$ be the Hilbert space of 1-periodic $W^{1,2}$-vector fields along γ. The Hessian $h := d^2E(\gamma)$ is the symmetric bilinear form on $W^{1,2}(S^1, \gamma^*TM)$ given by

$$h(\zeta, \eta) = 2\int_{S^1}\Big(\langle G_{xx}\zeta, \eta\rangle + \langle G_{vx}\dot\zeta, \eta\rangle + \langle G_{xv}\zeta, \dot\eta\rangle + \langle G_{vv}\dot\zeta, \dot\eta\rangle\Big)dt. \tag{4.4}$$

In this expression, we adopt a common abuse of notation: we write the integrand in local coordinates (this can be made precise by splitting the domain of integration S^1 as a finite union of intervals over which the local coordinates are available). A bootstrap argument, together with an integration by parts, implies that the kernel of h is precisely given by the 1-periodic Jacobi vector fields. In particular $\mathrm{nul}(h) = \dim\ker(\Phi_1 - I)$, and therefore $\mathrm{nul}(\gamma) = \mathrm{nul}(h) - 1 = \dim\ker(\Phi_1|_{\xi_{\tilde\gamma(0)}} - I)$, which implies point (i).

We consider the subspace

$$Z = \big\{\zeta \in W^{1,2}(S^1, \gamma^*TM) \ \big| \ G_v(\gamma, \dot\gamma)\zeta \equiv 0\big\}.$$

We claim that

$$\mathrm{ind}(h|_Z) = \mathrm{ind}(h), \qquad \mathrm{nul}(h|_Z) = \mathrm{nul}(h) - 1.$$

Indeed, a straightforward computation shows that the h-orthogonal

$$Z^h := \big\{\zeta \in W^{1,2}(S^1, \gamma^*TM) \ \big| \ h(\zeta, \cdot)|_Z = 0\big\}$$

is precisely the space of those $\zeta \in W^{1,2}(S^1, \gamma^*TM)$ of the form $\zeta(t) = f(t)\dot\gamma(t)$ for some $f : S^1 \to \mathbb{R}$, and we have $W^{1,2}(S^1, \gamma^*TM) = Z \oplus Z^h$, $\mathrm{ind}(h|_{Z^h}) = 0$, $\ker(h|_{Z^h}) = \mathrm{span}_{\mathbb{R}}\{\dot\gamma\}$, and $\mathrm{nul}(h|_{Z^h}) = 1$. From now on, we will simply write h for the restriction $h|_Z$, so that in particular

$$\mathrm{nul}(\gamma) = \mathrm{nul}(h) = \dim\ker(\Phi_1|_{\xi_{\tilde\gamma(0)}} - I).$$

Analogously, if we set

$$Z_0 = \big\{\zeta \in Z \ \big| \ \zeta(0) = \zeta(1) = 0\big\},$$

we have
$$\mathrm{ind}_\Omega(\gamma) = \mathrm{ind}(h|_{Z_0}), \qquad \mathrm{nul}_\Omega(\gamma) = \mathrm{nul}(h|_{Z_0}).$$

The kernel of $h|_{Z_0}$ is the space of Jacobi fields ζ such that $\zeta(0) = \zeta(1) = 0$ and $G_v(\gamma, \dot\gamma)\zeta \equiv 0$, and thus
$$\mathrm{nul}(h|_{Z_0}) = \dim \ker G_v(\gamma(t), \dot\gamma(t)) = \dim(M) - 1,$$

which proves point (ii).

Let us reduce the setting to finite dimension. Let $k \geq 2$ be a large enough integer such that no restriction $\gamma|_{[a,b]}$ with $b - a < k^{-1}$ contains conjugate points; namely, there are no Jacobi vector fields along γ vanishing on more than one point of $[a, b]$. We consider the finite dimensional vector space $V \subset Z$ of those vector fields $\zeta \in Z$ such that, for all $i = 0, ..., k - 1$, each restriction $\zeta|_{[i/k,(i+1)/k]}$ is a Jacobi vector field. The Morse indices of h and $h|_V$ are the same
$$\mathrm{ind}(h|_V) = \mathrm{ind}(h), \qquad \mathrm{nul}(h|_V) = \mathrm{nul}(h).$$

Indeed, an integration by parts in (4.4) shows that the h-orthogonal to V is the subspace
$$V^h = \left\{ \zeta \in Z \mid h(\zeta, \cdot)|_V = 0 \right\} = \left\{ \zeta \in Z \mid \zeta(\tfrac{i}{k}) = 0 \quad \forall i = 0, ..., k - 1 \right\},$$

and we have $Z = V \oplus V^h$ and $\mathrm{ind}(h|_{V^h}) + \mathrm{nul}(h|_{V^h}) = 0$. Analogously, if we set $V_0 := V \cap Z_0$, we have
$$\begin{aligned} \mathrm{ind}(h|_{V_0}) &= \mathrm{ind}(h|_{Z_0}) \\ \mathrm{nul}(h|_{V_0}) &= \mathrm{nul}(h|_{Z_0}) \end{aligned} \tag{4.5}$$

From now on, we will simply write h for the restriction $h|_V$.

The h-orthogonal $V_0^h = \left\{ \zeta \in V \mid h(\zeta, \cdot)|_{V_0} = 0 \right\}$ is precisely the space of vector fields $\zeta \in Z$ such that $\zeta|_{(0,1)}$ is a Jacobi vector field. We denote by $\mathrm{Ver} := \ker \mathrm{d}\pi \subset \mathrm{T}\mathrm{T}M$ the vertical sub-bundle of $\mathrm{T}\mathrm{T}M$. Each intersection
$$(\xi \cap \mathrm{Ver})_{\tilde\gamma(t)} = (SM \cap \mathrm{Ver})_{\tilde\gamma(t)}$$

has dimension $\dim(M) - 1$. For each $\zeta \in V_0^h$, we set
$$\tilde\zeta(t) := (\zeta(t), \dot\zeta(t)) = \Phi_t(\tilde\zeta(0^+)), \qquad \forall t \in (0, 1).$$

Notice that there is an isomorphism
$$V_0^h \to (\Phi_1 - I)|_{\xi_{\tilde\gamma(0)}}^{-1}(\xi \cap \mathrm{Ver})_{\tilde\gamma(0)}, \quad \zeta \mapsto \tilde\zeta(0^+).$$

In particular
$$\begin{aligned} \dim(V_0^h) &\leq \dim \ker(\Phi_1 - I)|_{\xi_{\tilde\gamma(0)}} + \dim(\xi \cap \mathrm{Ver})_{\tilde\gamma(0)} \\ &= \mathrm{nul}(h) + \dim(M) - 1. \end{aligned} \tag{4.6}$$

Moreover, the evaluation map $V_0^h \to \ker(G_v(\tilde\gamma(0)))$, $\zeta \mapsto \zeta(0)$ is surjective, and its kernel is precisely $\ker(h|_{V_0})$. Therefore
$$\begin{aligned} \dim(V_0^h) &= \dim \ker(h|_{V_0}) + \dim \ker(G_v(\gamma(0), \dot\gamma(0))) \\ &= \mathrm{nul}(h|_{V_0}) + \dim(M) - 1. \end{aligned} \tag{4.7}$$

The general formula relating the Morse indices of a quadratic form to the ones of its restriction to a subspace (see, e.g., [Maz16, Section A.2]) gives
$$\mathrm{ind}(h) = \mathrm{ind}(h|_{V_0}) + \mathrm{ind}(h|_{V_0^h}) + \mathrm{nul}(h|_{V_0^h}) - \mathrm{nul}(h). \tag{4.8}$$

In particular, by (4.6), we have

$$\mathrm{ind}(h) \leq \mathrm{ind}(h|_{V_0}) + \dim(V_0^h) - \mathrm{nul}(h) \leq \mathrm{ind}(h|_{V_0}) + \dim(M) - 1,$$

which proves point (iii). By (4.7) and (4.8), we have

$$\mathrm{ind}(h) + \mathrm{nul}(h) \leq \mathrm{ind}(h|_{V_0}) + \dim(V_0^h) \leq \mathrm{ind}(h|_{V_0}) + \mathrm{nul}(h|_{V_0}) + \dim(M) - 1,$$

which proves point (iv).

From now on, let us now assume that M is an orientable surface. The classical index theorem of Morse [Mor96] allows us to express $\mathrm{ind}(h|_{Z_0})$ and $\mathrm{nul}(h|_{Z_0})$ as

$$\mathrm{ind}(h|_{V_0}) = \sum_{t \in (0,1)} \dim\big(\Phi_t(\mathrm{Ver}_{\tilde\gamma(0)}) \cap \mathrm{Ver}_{\tilde\gamma(t)}\big),$$

$$\mathrm{nul}(h|_{V_0}) = \dim\big(\Phi_1(\mathrm{Ver}_{\tilde\gamma(0)}) \cap \mathrm{Ver}_{\tilde\gamma(1)}\big).$$

Notice that the Liouville vector field L takes values in the vertical sub-bundle Ver, and Equation (4.3) implies that

$$\Phi_t(\mathrm{Ver}_{\tilde\gamma(0)}) \cap \mathrm{Ver}_{\tilde\gamma(t)} = \Phi_t((\xi \cap \mathrm{Ver})_{\tilde\gamma(0)}) \cap (\xi \cap \mathrm{Ver})_{\tilde\gamma(t)}, \qquad \forall t \neq 0.$$

Since the fibers of the bundle $\xi \cap \mathrm{Ver}$ have dimension 1, we can express these index formulas by means of a single vector field η, as follows. Let us fix an arbitrary non-zero vector $\tilde\eta_0 \in (\xi \cap \mathrm{Ver})_{\tilde\gamma(0)}$, and define

$$\tilde\eta(t) = (\eta(t), \dot\eta(t)) := \Phi_t(\tilde\eta_0),$$

so that η is a Jacobi field along γ such that $\eta(0) = 0$ and $G_v(\gamma, \dot\gamma)\eta \equiv 0$. Since M is an orientable surface, the normal bundle of γ is trivial, and we can find a nowhere-vanishing 1-periodic smooth vector field μ along γ such that $G_v(\gamma, \dot\gamma)\mu \equiv 0$, so that we can express η as

$$\eta(t) = f(t)\mu(t)$$

for some smooth function $f : \mathbb{R} \to \mathbb{R}$. Notice that, since η is a Jacobi vector field that does not vanish identically, it has isolated zeroes, and in particular $\dot f(t) \neq 0$ whenever $f(t) = 0$. The index theory of Morse reduces to

$$\mathrm{ind}(h|_{Z_0}) = \#\{t \in (0,1) \mid f(t) = 0\}, \qquad \mathrm{nul}(h|_{Z_0}) = \begin{cases} 1 & \text{if } f(1) = 0, \\ 0 & \text{if } f(1) \neq 0. \end{cases}$$

If $\mathrm{nul}(h|_{Z_0}) = 1$, then $\eta(1) = 0$, and therefore

$$\tilde\eta(t+1) = \frac{\dot f(1)}{\dot f(0)}\tilde\eta(t), \qquad \forall t \in \mathbb{R},$$

This readily implies that $\mathrm{nul}_\Omega(\gamma^m) = 1$ and

$$\mathrm{ind}_\Omega(\gamma^m) = \#\{t \in (0,m) \mid f(t) = 0\} = m\,\mathrm{ind}_\Omega(\gamma) + m - 1.$$

This proves point (v).

With an integration by parts in (4.4), we readily see that the quadratic form h on the space V_0^h can be expressed in local coordinates as

$$\begin{aligned} h(\zeta, \zeta) &= \langle G_{vv}\dot\zeta(1^-) + G_{xv}\zeta(1) - G_{vv}\dot\zeta(0^+) - G_{xv}\zeta(0), \zeta(0)\rangle \\ &= \mathrm{d}\lambda((\Phi_1 - I)\tilde\zeta(0^+), \tilde\zeta(0^+)). \end{aligned} \tag{4.9}$$

Let us assume that $\mathrm{nul}(\gamma) = 2$, so that $\Phi_1|_{\xi_{\tilde\gamma(0)}} = I$, $(\Phi_1 - I)\tilde\zeta(0^+) = 0$ for all $\zeta \in V_0^h$, and $\mathrm{nul}(h|_{V_0}) = 1$. Equation (4.9) implies that $h|_{V_0^h} = 0$. Since $\dim(V_0^h) = 1 + \mathrm{nul}(h|_{V_0}) = 2$, this implies that $\mathrm{ind}(h|_{V_0^h}) = 0$ and $\mathrm{nul}(h|_{V_0^h}) = 2$. Therefore,

Equation (4.8) becomes $\mathrm{ind}(h) = \mathrm{ind}(h|_{V_0})$. Since $\Phi_1|_{\xi_{\tilde{\gamma}(0)}} = I$, in particular the above Jacobi field η is (smoothly) 1-periodic, and so is the function f. Therefore, since f has non-zero derivative at his zeroes, it must vanish an odd number of times in the open interval $(0, 1)$. Equation (4.5) allows to conclude that $\mathrm{ind}(\gamma) = \mathrm{ind}(h) = \mathrm{ind}(h|_{V_0})$ is odd. We can now repeat the same argument for all the iterates γ^m, since

$$\mathrm{nul}(\gamma^m) = \dim \ker(\Phi_1|_{\xi_{\tilde{\gamma}(0)}}^m - I) = 2,$$

and conclude that $\mathrm{ind}(\gamma)^m$ is odd as well for all $m \in \mathbb{N}$. This proves point (vi).

Finally, let us assume that $\mathrm{ind}_\Omega(\gamma^m) > 0$ for some $m \geq 1$, which is equivalent to the fact that the Jacobi field η introduced above vanishes at some positive time. Let $\tau > 0$ be the minimum $t > 0$ such that $\eta(t) = 0$. Up to replacing μ with $-\mu$, we can assume that $f|_{(0,\tau)} > 0$, so that $\dot{f}(0) > 0$ and $\dot{f}(\tau) < 0$. If $\tau \leq 1$, we consider the 1-periodic vector field along γ

$$\theta(t) = \begin{cases} \eta(t), & \text{if } t \in [0, \tau], \\ 0, & \text{if } t \in [\tau, 1], \end{cases}$$

which satisfies $h(\theta, \theta) = 0$ and

$$\begin{aligned} h(\theta, \mu) &= \langle G_{vv}\,\dot{\eta}(\tau) - G_{vv}\,\dot{\eta}(0), \mu(0) \rangle \\ &= \dot{f}(\tau)\,\langle G_{vv}\,\mu(\tau), \mu(\tau) \rangle - \dot{f}(0)\,\langle G_{vv}\,\mu(0), \mu(0) \rangle \\ &< 0. \end{aligned}$$

For each $\epsilon > 0$ the piecewise smooth vector field $\theta + \epsilon\mu$ is 1-periodic and everywhere transverse to $\dot{\gamma}$. Moreover,

$$h(\theta + \epsilon\mu, \theta + \epsilon\mu) = 2\epsilon h(\mu, \theta) + \epsilon^2 h(\mu\mu)$$

which is negative if $\epsilon > 0$ is sufficiently small. Assume now that $\tau > 1$. In this case, there exists $t > 0$ such that $f(t) = f(t+1) > 0$, $\dot{f}(t) > 0$ and $\dot{f}(t+1) < 0$. We define θ to be the 1-periodic vector field along γ such that $\theta|_{[t,t+1]} = \eta|_{[t,t+1]}$. Notice that θ is everywhere transverse to $\dot{\gamma}$, and satisfies

$$\begin{aligned} h(\theta, \theta) &= \langle G_{vv}\,\dot{\eta}(t+1), \eta(t+1) \rangle - \langle G_{vv}\,\dot{\eta}(t), \eta(t) \rangle \\ &= \langle G_{vv}\,(\dot{\eta}(t+1) - \dot{\eta}(t)), \eta(t) \rangle \\ &= (\dot{f}(t+1) - \dot{f}(t))\,f(t)\,\langle G_{vv}\,\mu(t), \mu(t) \rangle \\ &< 0. \end{aligned}$$

In both cases, we can approximate θ with a C^0-close 1-periodic smooth vector field ζ. Such a ζ will still be everywhere transverse to $\dot{\gamma}$ and will still satisfy $h(\zeta, \zeta) < 0$. This completes the proof of point (vii). $\qquad\square$

4.3. **Local homology.** The last index that is usually employed in critical point theory is the local homology, whose construction we now recall for closed geodesics of Finsler manifolds (M, F). For any $\mathcal{U} \subset \Lambda$, $U \subset \Lambda_k$, and $\ell > 0$, we set

$$\mathcal{U}^{<\ell} := \big\{ \gamma \in \mathcal{U} \mid E(\gamma) < \ell^2 \big\}, \qquad U^{<\ell} := \big\{ \boldsymbol{x} \in U \mid E_k(\boldsymbol{x}) < \ell^2 \big\}.$$

Notice that $\mathcal{U}^{<\ell}$ and $U^{<\ell}$ are sublevel sets of the energy functional E, whereas in (3.1) we denoted by $\mathcal{W}^{<\ell}$ a sublevel set of the length functional L. Nevertheless, the notation is consistent: $\mathcal{W}$ was indeed a subset of the space of unparametrized loops Π, and if we parametrize any $\gamma \in \mathcal{W}$ with constant speed and period 1 we have $L(\gamma)^2 = E(\gamma)$.

The energy functional E is invariant under the circle action

$$u \cdot \gamma = \gamma(u + \cdot) \in \Lambda, \qquad \forall u \in S^1, \ \gamma \in \Lambda.$$

Therefore, every closed geodesic $\gamma \in \mathrm{crit}(E) \cap E^{-1}(\ell^2)$ (with $\ell > 0$) belongs to a circle of critical points of E

$$S^1 \cdot \gamma := \left\{ \gamma(u + \cdot) \in \Lambda \ \middle| \ u \in S^1 \right\}$$

A closed geodesic γ is said to be isolated when the critical circles of its iterates $S^1 \cdot \gamma^m$ are isolated in $\mathrm{crit}(E)$. Under this assumption, the local homology of γ and of $S^1 \cdot \gamma$ are the relative homology groups

$$C_*(\gamma) := H_*(\Lambda^{<\ell} \cup \{\gamma\}, \Lambda^{<\ell}), \qquad C_*(\gamma) := H_*(\Lambda^{<\ell} \cup S^1 \cdot \gamma, \Lambda^{<\ell}).$$

As we already mentioned, we will denote the coefficient field in the notation only when we will need to employ a specific one.

The local homology groups of the critical circles of closed geodesics are the "building blocks" for the homology of the free loop space Λ. Indeed, if $\gamma \in \mathrm{crit}(E) \cap E^{-1}(\ell^2)$ is an isolated closed geodesic and the interval $(\ell, \ell + \epsilon)$ does not contain critical values of E, the inclusion induces an injective homomorphism

$$C_*(S^1 \cdot \gamma) \hookrightarrow H_*(\Lambda^{<\ell+\epsilon}, \Lambda^{<\ell})$$

The local homology of an isolated closed geodesic often embeds into the local homology of its critical circle. For instance, at least when we employ coefficients in $\mathbb{Q}$, if γ is prime (that is, $\gamma = \zeta^m$ if and only if $\zeta = \gamma$ and $m = 1$), then for all odd numbers $m \in \mathbb{N}$ such that $\mathrm{nul}(\gamma) = \mathrm{nul}(\gamma^m)$ the inclusion induces an injective homomorphism

$$C_*(\gamma^m; \mathbb{Q}) \hookrightarrow C_*(S^1 \cdot \gamma^m; \mathbb{Q}).$$

The local homology of an isolated closed geodesic γ does not vary (up to a shift in degree) under iterations that preserve the nullity. In particular, if $\mathrm{ind}(\gamma) = \mathrm{ind}(\gamma^m)$ and $\mathrm{nul}(\gamma) = \mathrm{nul}(\gamma^m)$, the iteration map $\psi^m : \Lambda \hookrightarrow \Lambda$, $\psi^m(\zeta) = \zeta^m$ induces the local homology isomorphisms

$$\psi^m_* : C_*(\gamma) \xrightarrow{\cong} C_*(\gamma^m), \qquad \psi^m_* : C_*(S^1 \cdot \gamma) \xrightarrow{\cong} C_*(S^1 \cdot \gamma^m).$$

Even though the energy function E may not be C^2, the local homology $C_*(\gamma)$ is isomorphic to the local homology of a smooth function at an isolated critical point of index $\mathrm{ind}(\gamma)$ and nullity $\mathrm{nul}(\gamma)$. Indeed, if $k \in \mathbb{N}$ is large enough so that the closed geodesic $\gamma \in \mathrm{crit}(E) \cap E^{-1}(\ell^2)$ belongs to Λ_k, the inclusion induces the homology isomorphism

$$H_*(\Lambda_k^{<\ell} \cup \{\gamma\}, \Lambda_k^{<\ell}) \xrightarrow{\cong} C_*(\gamma).$$

The energy $E_k = E|_{\Lambda_k}$ is smooth in a neighborhood of the critical point γ (indeed, E_k is smooth at all those $\zeta \in \Lambda_k$ such that $\zeta(\frac{i}{k}) \neq \zeta(\frac{i+1}{k})$ for all $i \in \mathbb{Z}_k$). Let $\Sigma \subset M$ be an embedded hypersurface intersecting γ transversely at $x_0 := \gamma(0)$. We define the smooth hypersurface

$$\Sigma_k := \left\{ \zeta \in \Lambda_k \ \middle| \ \zeta(0) \in \Sigma \right\} \subset \Lambda_k.$$

Now, γ becomes an isolated critical point of $E_k|_{\Sigma_k}$ of index $\mathrm{ind}(\gamma)$ and nullity $\mathrm{nul}(\gamma)$, and the inclusions induce the homology isomorphisms

$$H_*(\Sigma_k^{<\ell} \cup \{\gamma\}, \Sigma_k^{<\ell}) \xrightarrow{\cong} C_*(\gamma).$$

Since $\boldsymbol{x}$ is an isolated critical point of $E_k|_{\Sigma_k}$, its local homology can also be expressed by means of the so-call Gromoll-Meyer neighborhoods [GM69a]: these are suitable arbitrarily small compact neighborhoods $W \subset \Sigma_k$ of $\boldsymbol{x}$ such that, for all $\delta \geq 0$ small enough, the inclusion induces the homology isomorphism

$$H_*(\Sigma_k^{<\ell} \cup \{\boldsymbol{x}\}, \Sigma_k^{<\ell}) \overset{\cong}{\longleftarrow} H_*(W^{<\ell} \cup \{\boldsymbol{x}\}, W^{<\ell-\delta}) \overset{\cong}{\longrightarrow} H_*(W, W^{<\ell-\delta}).$$

These neighborhoods are particularly useful to prove certain technical statements concerning the local homology, for instance the following one that we will employ in the proof of Corollary 5.6: if a closed geodesic $\gamma \in \mathrm{crit}(E) \cap E^{-1}(\ell^2)$ whose critical circle $S^1 \cdot \gamma$ is isolated in $\mathrm{crit}(E)$ admits an arbitrarily small open neighborhood $\mathcal{U} \subset \Lambda$ or $\mathcal{U} \subset \Lambda_k$ such that $\mathcal{U}^{<\ell}$ is not connected, then the local homology $C_1(\gamma)$ is non-zero.

We close this section by proving the following proposition relating the local homology in the setting Λ to the one in the setting Π of Section 3. By applying the proposition to the three min-max values $\ell(h_i)$, for $i = 1, 2, 3$, defined in (3.4), we will infer Theorem 1.2(iii).

Proposition 4.3. *Let (M, F) be a closed, orientable, reversible Finsler surface, $\rho > 0$ a constant, and $h \in H_d(\Pi, \Pi^{<\rho})$ a non-trivial homology class. Assume that there are only finitely many simple closed geodesics of (M, F) having length in a neighborhood of $\ell(h)$. Then, there exists a simple closed geodesic $\gamma \in \mathrm{crit}(E) \cap E^{-1}(\ell(h)^{1/2})$ with non-zero local homology $C_d(\gamma) \neq 0$.*

Proof. We first apply Lemma 3.1 and obtain a simple closed geodesic γ of length $\ell := \ell(h)$ and, for every $\epsilon > 0$, an open neighborhood $\mathcal{V}(\gamma, \epsilon) \subset \Pi$ and a constant $\delta \in (0, \epsilon^2)$ such that the homomorphism

$$H_*(\mathcal{V}(\gamma, \epsilon), \mathcal{V}(\gamma, \epsilon)^{<\ell-\delta}) \to H_*(\Pi, \Pi^{<\ell})$$

induced by the inclusion is non-zero.

Let $\Sigma \subset M$ be an embedded open hypersurface (i.e. an open segment) intersecting γ transversely. We choose Σ and $\epsilon_0 > 0$ small enough so that every $\zeta \in \mathcal{V}(\gamma, \epsilon_0)$ intersect Σ in a single point and, by the implicit function theorem, the map $\mathcal{V}(\gamma, \epsilon_0) \to \Sigma$, $\zeta \mapsto \zeta \cap \Sigma$ is continuous. Throughout this section, we uniquely parametrize every $\zeta \in \mathcal{V}(\gamma, \epsilon_0)$ so that

$$\zeta : S^1 \to M, \qquad F(\zeta, \dot{\zeta}) \equiv L(\zeta), \qquad \zeta(0) \in \Sigma.$$

With this choice of parametrizations, $\mathcal{V}(\gamma, \epsilon_0)$ is compact in the C^1 topology. Moreover, every C^1-open neighborhood of γ contains $\mathcal{V}(\gamma, \epsilon)$ for a sufficiently small $\epsilon \in (0, \epsilon_0]$.

Let us consider an embedding $M \hookrightarrow \mathbb{R}^3$, which exists since M is an orientable closed surface. Let $U \subset \mathbb{R}^3$ be a tubular neighborhood of M with associated smooth retraction $\pi : U \to M$. We consider a family of mollifiers $\theta_s(u) = \theta(u/s)/s$, where $s \in (0, 1)$ and $\theta : S^1 \to [0, \infty)$ is a smooth function supported in $(-1/2, 1/2)$ and with integral 1. We denote by $*$ the convolution operation. Since $\mathcal{V}(\gamma, \epsilon_0)$ is C^1-compact and θ_s tends to the Dirac delta as $s \to 0$, there exists $s_0 > 0$ and $\epsilon_1 \in (0, \epsilon_0]$ such that we have a well defined continuous map

$$c : [0, s_0] \times \mathcal{V}(\gamma, \epsilon_1) \to \mathrm{Emb}(S^1, M), \qquad c(s, \zeta)(u) = c_s(\zeta)(u) = \pi(\zeta * \theta_s(u)).$$

Notice that $c_0(\zeta) = \zeta$ for all $\zeta \in \mathcal{V}(\gamma, \epsilon_1)$. Since the length function is continuous on the C^1-compact subset $\mathcal{V}(\gamma, \epsilon_1)$, there exists $s_1 \in (0, s_0]$ such that

$$L(c_s(\zeta)) < L(\zeta) + \delta/2, \qquad \forall s \in [0, s_1], \ \zeta \in \mathcal{V}(\gamma, \epsilon). \tag{4.10}$$

By the continuity of the convolution, there exists an open subset $\mathcal{U} \subset W^{1,2}(S^1, M)$ containing $\mathcal{V}(\gamma, \epsilon)$ such that c_{s_1} extends as a continuous map

$$c_{s_1} : \mathcal{U} \to \mathrm{Emb}(S^1, M), \qquad c_{s_1}(\zeta)(u) = \pi(\zeta * \theta_{s_1}(u)),$$

and

$$L(c_{s_1}(\zeta)) < L(\zeta) + \delta, \qquad \forall \zeta \in \mathcal{U}. \tag{4.11}$$

We consider an integer

$$k > \frac{\ell + \epsilon_1^2}{\mathrm{injrad}(M, F)}$$

that we will soon fix, and the space Σ_k of broken closed geodesics intersecting Σ at time 0. We define a continuous homotopy

$$r_t : \mathcal{V}(\gamma, \epsilon_1) \to W^{1,2}(S^1, M), \ t \in [0, 1],$$

as follows: we uniquely parametrize every $\zeta \in \mathcal{V}(\gamma, \epsilon_0)$ so that

$$F(\zeta, \dot\zeta) \equiv L(\zeta), \qquad \zeta(0) \in \Sigma;$$

for all $i = 0, ..., k - 1$, we define

$$r_t(\zeta)|_{[i/k, (i+1-t)/k]} := \zeta|_{[i/k, (i+1-t)/k]},$$

and $r_t(\zeta)|_{[(i+1-t)/k, (i+1)/k]}$ as the shortest geodesic of (M, F) parametrized with constant speed and joining its endpoints. We require k to be large enough so that every r_t has image inside the open subset $\mathcal{U} \subset W^{1,2}(S^1, M)$. Clearly,

$$E(r_t(\zeta)) \leq E(\zeta) = L(\zeta)^2, \qquad \forall t \in [0, 1].$$

We consider a Gromoll-Meyer neighborhood $W \subset \Sigma_k \cap \mathcal{U}$ of $\gamma = r_1(\gamma)$. Notice that, by (4.11), we have

$$L(c_{s_1}(\zeta)) < L(\zeta) + \delta \leq E(\zeta)^{1/2} + \delta, \qquad \forall \zeta \in W,$$

and in particular $c_{s_1}(W^{<\ell-\delta}) \subset \Pi^{<\ell}$. We fix a constant $\epsilon_2 \in (0, \epsilon_1]$ small enough so that $r_1(\mathcal{V}(\gamma, \epsilon_1)) \subset W$. Overall, we have the homomorphisms

$$
\begin{array}{ccc}
H_*(\mathcal{V}(\gamma, \epsilon_2), \mathcal{V}(\gamma, \epsilon_2)^{<\ell-\delta}) & \xrightarrow{\ \ \ \ \ i_* \ \ \ \ \ } & H_*(\Pi, \Pi^{<\ell}) \\
& (r_1)_* \searrow \qquad \nearrow (c_{s_1})_* & \\
& H_*(W, W^{<\ell-\delta}) &
\end{array}
\tag{4.12}
$$

where i_* is the non-zero homomorphism induced by the inclusion (see the first paragraph of the proof). All we need to do in order to complete the proof is to show that the diagram (4.12) commutes. This is a consequence of the fact that the inclusion i is homotopic to the composition $c_{s_1} \circ r_1$ via the continuous homotopy

$$h_t : \mathcal{V}(\gamma, \epsilon_2) \to \Pi, \qquad h_t = \begin{cases} c_{2ts_1}, & \text{if } t \in [0, 1/2], \\ c_{s_1} \circ r_{2t-1}, & \text{if } t \in [1/2, 1], \end{cases}$$

which satisfies $h_0 = i$, $h_1 = c_{s_1} \circ r_1$, and $h_t(\mathcal{V}(\gamma, \epsilon_2)^{<\ell-\delta}) \subset \Pi^{<\ell}$ for all $t \in [0, 1]$ according to (4.10) and (4.11). $\qquad \square$

5. Infinitely many closed geodesics

5.1. **The Birkhoff map.** Let (S^2, F) be a reversible Finsler sphere, and $SS^2 = \{(x, v) \mid F(x, v) = 1\}$ its Finsler unit tangent bundle with base projection $\pi : SM \to M$, $\pi(x, v) = x$. As we already recalled in the proof of Proposition 4.2, SS^2 admits the contact form $\alpha = G_v \, d\pi$, where $G(x, v) = \frac{1}{2}F(x, v)^2$, and the associated Reeb vector field X on SM defined by $\alpha(X) \equiv 1$ and $d\alpha(X, \cdot) \equiv 0$. The flow $\phi_t : SM \to SM$ of X is precisely the geodesic flow of (S^2, F).

Let $\gamma : S^1 \hookrightarrow S^2$ a simple closed geodesic of (S^2, F). Without loss of generality, let us assume that $F(\gamma, \dot\gamma) \equiv 1$. The complement $S^2 \setminus \gamma$ is the disjoint union of two open balls $B_0, B_1 \subset S^2$. We consider the open annuli

$$A_i := \big\{(x, v) \in SS^2 \ \big| \ x \in \gamma(S^1), \ v \pitchfork \dot\gamma(t) \text{ and points inside } B_i\big\}, \quad i = 0, 1.$$

Since the Reeb vector field X is transverse to A_i, we readily see that

$$d\alpha|_{A_i} = (X \lrcorner (\alpha \wedge d\alpha))|_{A_i}$$

is a symplectic form on A_i. We assume that the first return time

$$\tau_i : A_i \to (0, \infty], \qquad \tau_i(x, v) := \inf\big\{t > 0 \ \big| \ \phi_t(x, v) \in A_{1-i}\big\} \in (0, +\infty]$$

is finite for all $(x, v) \in A_i$ (here, we adopt the usual convention $\inf \varnothing = +\infty$). Under this assumption, there is a well defined first return map

$$\psi_i : A_i \to A_{1-i}, \qquad \psi_i(x, v) = \phi_{\tau_i(x,v)}(x, v),$$

which is a diffeomorphism. Since

$$\psi_i^* \alpha - \alpha = \phi_t^* \alpha|_{t=\tau_i} + \alpha(\partial_t \phi_t(z))|_{t=\tau_i} d\tau_i - \alpha = \alpha + \alpha(X) d\tau_i - \alpha = d\tau_i,$$

the first return map is an exact symplectomorphism $\psi_i : (A_i, d\alpha) \to (A_{1-i}, d\alpha)$. Notice that

$$\partial A_0 = \partial A_1 = \{(\gamma(t), \dot\gamma(t)), (\gamma(t), -\dot\gamma(t)) \mid t \in S^1\},$$

and we readily see that $d\alpha|_{\overline{A}_i}$ vanishes on ∂A_i.

For each $t \in [0, 1)$, we choose a non-zero $w_t \in \ker d\pi(\tilde\gamma(0))$ depending smoothly on t, and we extend it to a vector field

$$\tilde\eta_t(s) = (\eta_t(s), \dot\eta_t(s)) := d\phi_{s-t}(\tilde\gamma(t)) w_t. \tag{5.1}$$

Namely, η_t is a non-trivial Jacobi vector field along γ such that $G_v(\gamma, \dot\gamma)\eta_t \equiv 0$ and $\eta_t(t) = 0$. We recall that the points $\gamma(t), \gamma(s)$, with $t \neq s$, are conjugate when $\eta_t(s) = 0$. For each $t \in [0, 1)$, we set

$$t_{-1} := \sup\{s < t \mid \eta_t(s) = 0\}, \qquad t_1 := \inf\{s > t \mid \eta_t(s) = 0\},$$

and $t_{\pm 2} := (t_{\pm 1})_{\pm 1}$ (here, once again, we set $\sup \varnothing = -\infty$ and $\inf \varnothing = +\infty$). Namely, t_i is the time of the $|i|$-th conjugate point to $\gamma(t)$ after t if $i > 0$, or before t if $i < 0$.

Lemma 5.1. *Assume that, for some $t \in [0, 1)$, t_1 is finite. Then, for all $t \in [0, 1)$ both t_1 and t_{-1} are finite, and the first return maps ψ_i can be extended as homeomorphisms*

$$\psi_i : \overline{A}_i \to \overline{A}_{1-i}, \qquad \psi_i(\gamma(t), \pm\dot\gamma(t)) = (\gamma(t_{\pm 1}), \pm\dot\gamma(t_{\pm 1})). \tag{5.2}$$
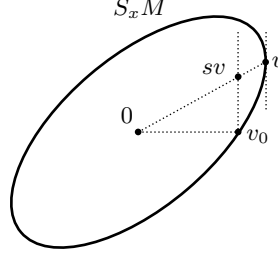
FIGURE 1. The value of $s \in [0, 1]$ such that $v_0 - s\,v \in \ker G_v(x, v)$ can be found geometrically: the parallel to the tangent line $\mathrm{T}_v(S_x M)$ passing through v_0 intersects the segment joining the origin and v at sv.

Proof. Let μ be a nowhere vanishing 1-periodic vector field along γ such that $G_v(\gamma, \dot{\gamma})\mu \equiv 0$. The Jacobi fields η_t can be written as $\eta_t(s) = f(t, s)\mu(s)$ for some smooth function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Up to replacing μ with $-\mu$, we can assume that $f(t, t + \epsilon) > 0$ for all $t \in S^1$ and $\epsilon > 0$ small enough. Since the Jacobi fields η_t are non-trivial, we have $\partial_s f(t, s) \neq 0$ whenever $f(t, s) = 0$. This readily implies that, if t_1 is finite for some $t \in \mathbb{R}$, the same is true for all $t \in \mathbb{R}$, and the function $t \mapsto t_1$ is continuous and monotone increasing. Since $(t_1)_{-1} = t$, we infer that the function $t \mapsto t_{-1}$ is well defined, continuous and monotone increasing as well.

We fix an arbitrary $t \in [0, 1)$ and $(x, v) := (\gamma(t), \dot{\gamma}(t))$. In order to complete the proof, we are left to show that, for each sequence $v_n \in S_x S^2$ of vectors pointing inside A_i and such that $v_n \to \pm v$, we have $\tau_i(x, v_n) \to \pm(t_{\pm 1} - t)$. Indeed, this implies that $\phi_{\tau_i(x, v_n)}(x, v_n) \to (\gamma(t_{\pm 1}), \pm\dot{\gamma}(t_{\pm 1}))$, and therefore the extension (5.2) of ψ_i is continuous and bijective. Since the annuli $\overline{A}_i$ and $\overline{A}_{i-1}$ are compact and Hausdorff, such an extension is a homeomorphism.

Let us focus on the case $v_n \to v$, the other one being analogous. We set $\gamma_n(s) := \exp_x((s - t)\,v_n)$ and $\sigma_n := \tau_i(x, v_n)$. We claim that

$$\liminf_{n \to \infty} \sigma_n \geq t_1 - t.$$

Otherwise, we could extract a subsequence such that $\sigma_n \to \sigma \in (0, t_1 - t)$; however, since the geodesic $\gamma|_{[t, t+\sigma]}$ has no conjugate points, this would contradict the fact that the exponential map $\exp_x$ is a local diffeomorphism at σv. The fact that $f(t_1 + \epsilon) < 0$ if $\epsilon > 0$ is small enough readily implies that $\gamma_n(t_1 + \epsilon) \in A_{1-i}$ for all n large enough, and therefore $\sigma_n < t_1 - t + \epsilon$. This implies that $\sigma_n \to t_1 - t$. $\qquad\square$

We set $A := A_0$. The previous lemma implies that the annulus A is a surface of sections for the geodesic flow: a surface that is transverse to the vector field X on its interior, and whose boundary is the union of periodic orbits of the flow. The composition $\psi := \psi_1 \circ \psi_0 : A \to A$ is the first return map of the surface of section A, and extends to a homeomorphism of $\overline{A}$ as

$$\psi(\gamma(t), \pm\dot{\gamma}(t)) = (\gamma(t_{\pm 2}), \pm\dot{\gamma}(t_{\pm 2})).$$

As customary in the Riemannian literature, we will call ψ the Birkhoff map of γ. With a suitable change of coordinates, A becomes the standard symplectic annulus.

Lemma 5.2. *There exists a smooth homeomorphism*

$$\sigma : S^1 \times [-1, 1] \to \overline{A}$$

of the form $\sigma(t,s) = (\gamma(t), \nu(t,s))$, where $\nu(t,s) \in S_{\gamma(t)}M$ is the unique tangent vector (pointing inside B_0 or tangent to γ) such that

$$\dot{\gamma}(t) - s\,\nu(t,s) \in \ker G_v(\gamma(t), \nu(t,s)).$$

The map σ restrict to a diffeomorphism $\sigma : S^1 \times (-1,1) \to A$, and $\sigma^*\alpha = s\,\mathrm{d}t$.

Proof. We fix $x = \gamma(t)$ and $v_0 = \dot{\gamma}(t)$. For each $v \in S_xM$ there is a unique $s(v) \in [-1,1]$ such that $v_0 - s(v)v \in \ker G_v(x,v)$, see Figure 1. Clearly, $s(v)$ depends smoothly on v. We choose an arbitrary parametrization of the fiber

$$v : [0,1] \xrightarrow{\cong} \overline{A} \cap S_xM$$

such that $v(0) = v_0$ and $v(1) = -v_0$. Notice that $\dot{v}(r) \in \ker G_v(x,v(r))$, and there exists $\lambda(r) \in \mathbb{R}$ such that $v_0 = s(v(r))v(r) + \lambda(r)\dot{v}(r)$. By the strict convexity of S_xM, we have $\lambda(r) = 0$ if and only if $r \in \{0,1\}$. Since

$$G_v(x,v)v_0 = G_v(x,v)s(v)v = s(v),$$

we have

$$
\begin{aligned}
\mathrm{d}s(v(r))\dot{v}(r) &= \tfrac{\mathrm{d}}{\mathrm{d}r}s(v(r)) = \tfrac{\mathrm{d}}{\mathrm{d}r}G_v(x,v(r))v_0 = G_{vv}(x,v(r))\dot{v}(r)\,v_0 \\
&= s(v(r))G_{vv}(x,v(r))\dot{v}(r)\,v(r) + \lambda(r)G_{vv}(x,v(r))\dot{v}(r)\,\dot{v}(r) \\
&= s(v(r))G_v(x,v(r))\dot{v}(r) + \lambda(r)G_{vv}(x,v(r))\dot{v}(r)\,\dot{v}(r) \\
&= \lambda(r)G_{vv}(x,v(r))\dot{v}(r)\,\dot{v}(r).
\end{aligned}
$$

The last term is non-zero for all $r \in (0,1)$. Therefore, $s : \overline{A} \cap S_xM \to [-1,1]$ is a diffeomorphism that restricts to a diffeomorphism $s : A \cap S_xM \to (-1,1)$. We set $\sigma(x,\cdot)$ to be the inverse homeomorphism. The obtained map $\sigma : S^1 \times [-1,1] \to \overline{A}$ is thus a homeomorphism that restricts to a diffeomorphism $\sigma : S^1 \times (-1,1) \to A$. The pull-back of the contact form α by σ is

$$(\sigma^*\alpha)_{(t,s)} = G_v(\gamma(t),\nu(t,s))\dot{\gamma}(t)\,\mathrm{d}t = G_v(\gamma(t),\nu(t,s))s\,\nu(t,s)\,\mathrm{d}t = s\,\mathrm{d}t. \qquad \square$$

From now on, the annulus $S^1 \times [-1,1]$ will be implicitly equipped with the Euclidean area form $\mathrm{d}s \wedge \mathrm{d}t$. By means of Lemma 5.2, we will always consider the Birkhoff map of a simple closed geodesic γ as a homeomorphism $\psi : S^1 \times [-1,1] \to S^1 \times [-1,1]$ that restricts to a symplectomorphism of $(S^1 \times (-1,1), \mathrm{d}s \wedge \mathrm{d}t)$ and acts on the boundary as $\psi(t,\pm 1) = (t_{\pm 2}, \pm 1)$.

5.2. **Periodic points of twist maps.** Let $\psi : S^1 \times [-1,1] \to S^1 \times [-1,1]$ be an area preserving homeomorphism preserving the boundary components $S^1 \times \{-1\}$ and $S^1 \times \{1\}$. Such a ψ is called a **twist map** when it admits a lift

$$\widetilde{\psi} : \mathbb{R} \times [-1,1] \to \mathbb{R} \times [-1,1], \qquad \widetilde{\psi}(t,s) = (a(t,s), b(t,s)), \qquad (5.3)$$

satisfying the twist conditions $a(t,1) < t$ and $a(t,-1) > t$ for all $t \in \mathbb{R}$. If ψ is the Birkhoff map of a simple closed geodesic γ of (S^2, F), the set of its periodic orbits is in one-to-one correspondence with the set of closed geodesics intersecting γ (other than γ itself). In particular, the existence of infinitely many periodic points of ψ implies the existence of infinitely many closed geodesics on (S^2, F)

By the celebrated Poincaré-Birkhoff theorem [Bir66], any twist map has at least two fixed points in the interior of the annulus. Indeed, more is true: any lift (5.3) satisfying the twist condition has at least two fixed points. A simple argument

due to Neumann [Neu77] further implies that any twist map ψ has infinitely many periodic points. Indeed, consider the translation

$$\tau : \mathbb{R} \times [-1, 1] \to \mathbb{R} \times [-1, 1], \qquad \tau(t, s) = (t + 1, s).$$

For each integer $q > 0$ there exists another relatively prime integer $p > 0$ that is large enough so that

$$p \min_{x \in \mathbb{R}} \big(x - a(x, 1) \big) > q.$$

This condition guarantees that $\widetilde{\phi} := \widetilde{\psi}^p \circ \tau^q$ is a lift of $\phi = \psi^p$ satisfying the twist condition, and therefore has at least a fixed point $z \in \mathbb{R} \times [-1, 1]$. Such a z projects to a p-periodic point $[z]$ of ψ, and since p, q are relatively prime the minimal period of $[z]$ is p.

Let us now apply this results to the Birkhoff map of a simple closed geodesic γ of (S^2, F). For each $t \in S^1$, we denote $t \cdot \gamma := \gamma(t + \cdot)$ the closed geodesic γ with the parametrization translated by t, and by $\mathrm{ind}_\Omega(t \cdot \gamma)$ the Morse index of $t \cdot \gamma$ in its corresponding based loop space $\Omega_t = \{\zeta \in \Lambda \mid \zeta(0) = \gamma(t)\}$ (see Section 4.2).

Theorem 5.3. *If the simple closed geodesic γ satisfies $\mathrm{ind}_\Omega(t \cdot \gamma) \geq 2$ for all $t \in S^1$ and has a well defined Birkhoff map ψ, then ψ is a twist map, and in particular (S^2, F) has infinitely many closed geodesics.*

Proof. Let us consider the family of Jacobi fields η_t introduced in (5.1). As we already mentioned in the proof of Proposition 4.2, the classical Morse index theorem [Mor96] allows to relate $\mathrm{ind}_\Omega(t \cdot \gamma)$ to the zeros of η_t by

$$\mathrm{ind}_\Omega(t \cdot \gamma) = \#\{s \in (t, t+1) \mid \eta_t(s) = 0\} = \#\{s \in (t-1, t) \mid \eta_t(s) = 0\}.$$

Therefore, $\mathrm{ind}_\Omega(t \cdot \gamma) \geq 2$ is equivalent to $t_2 - t < 1$ and $t - t_{-2} < 1$. If this holds for all $t \in \mathbb{R}$, we claim that the Birkhoff map ψ is a twist map. Indeed, ψ can be lifted to a continuous map

$$\widetilde{\psi} : \mathbb{R} \times [-1, 1] \to \mathbb{R} \times [-1, 1],$$

as follows. Let $\sigma : S^1 \times [-1, 1] \to \overline{A}$, $\sigma(t, s) = (\gamma(t), \nu(t, s))$ be the homeomorphism of Lemma 5.2. For each $(t, s) \in \mathbb{R} \times (-1, 1)$, we consider the geodesic ray ζ starting at $\zeta(0) = \gamma(t)$ with speed $\dot{\zeta}(0) = \nu(t, s)$. Let $a', a'' \in (0, 1]$ be the such that the first intersection of ζ at positive time with with γ is at $\gamma(t + a')$, and the second one is at $\gamma(t + a' + a'')$. Let $0 < b' < b''$ be the first positive times such that $\zeta(b') = \gamma(t + a')$ and $\zeta(b'') = \gamma(t + a' + a'')$. We denote by $i', i'' \in \mathbb{Z}$ the algebraic count of self-intersections of the geodesics $\zeta|_{(0, b')}$ and $\zeta|_{(b', b'')}$ respectively; here a double-point intersection is counted positively if and only if ζ crosses itself from left to right (up to isotoping $\zeta|_{[0, b'']}$ without moving $\zeta(0)$, $\zeta(b')$, and $\zeta(b'')$, we can assume that all the self-intersections of $\zeta|_{[0, b'']}$ are double points). We define the lift

$$\widetilde{\psi}(t, s) = (a(t, s), b(t, s)),$$

by setting the first component to be

$$a(t, s) = t + a' + a'' + i' + i'' - 1.$$

It is straightforward to verify that such a function a is continuous. Since $\mathrm{ind}_\Omega(t \cdot \gamma) \geq 2$, if $|s|$ is close to 1 (that is, if $\dot{\zeta}(0)$ is close to $\pm \dot{\gamma}(0)$), we have $i' = i'' = 0$; if s is
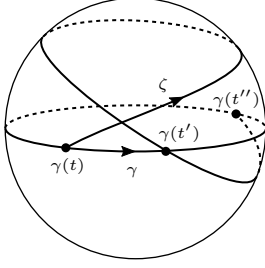
FIGURE 2. Example of geodesic ray ζ on a reversible Finsler sphere intersecting the simple closed geodesic γ at subsequent points $\gamma(t)$, $\gamma(t')$ and $\gamma(t'')$, with $t' = t + a' \in (t, t+1]$ and $t'' = t' + a'' \in (t', t'+1]$, such that $i' = 1$ and $i'' = 0$.

close to 1 we have $a' + a'' \in (0, 1)$, whereas if s is close to -1 we have $a' + a'' \in (1, 2)$. Therefore

$$a(t, 1) - t = t_2 - t - 1 \in (-1, 0), \qquad a(t, -1) - t = 1 - (t - t_{-2}) \in (0, 1),$$

namely $\widetilde{\psi}$ satisfies the twist condition. Since ψ is a twist map, it has infinitely many periodic points corresponding to infinitely many closed geodesics of (S^2, F).   □

5.3. **Hingston's theorems.** A celebrated theorem due to Hingston [Hin93], that extends previous results of Bangert [Ban80, Ban93], implies the existence of infinitely many closed geodesics on (S^2, F) when there is a simple closed geodesic with non-zero local homology in degree 3 and a Birkhoff map not of twist type. Hingston's original proof was phrased for Riemannian manifolds, but is valid as well in the Finsler setting, and indeed even in the non-reversible Finsler setting. We include the full argument here for the reader's convenience.

Theorem 5.4. *Let (M, F) be a closed Finsler manifold of dimension $d \geq 2$, and γ a closed geodesic such that*

   (i) *The local homology $C_i(\gamma)$ with coefficient in an arbitrary field is non-zero in degree $i = \mathrm{ind}(\gamma) + \mathrm{nul}(\gamma)$.*

   (ii) *$\mathrm{ind}(\gamma^m) + \mathrm{nul}(\gamma^m) \leq m(\mathrm{ind}(\gamma) + \mathrm{nul}(\gamma)) - (d-1)(m-1)$ for all $m \in \mathbb{N}$.*

Then (M, F) has infinitely many closed geodesics.

Proof. Condition (ii), together with Lemma 4.1(ii,v), implies that

$$\mathrm{ind}_\Omega(\gamma) + \mathrm{nul}_\Omega(\gamma) \leq \tfrac{1}{m}\big(\mathrm{ind}_\Omega(\gamma^m) + \mathrm{nul}_\Omega(\gamma^m)\big) \leq \mathrm{ind}(\gamma) + \mathrm{nul}(\gamma) - \tfrac{m-1}{m}(d-1).$$

In the limit $m \to \infty$ the latter term converges to $d - 1$. This, together with the opposite inequality provided by Proposition 4.2(iv), implies

$$\mathrm{ind}_\Omega(\gamma) + \mathrm{nul}_\Omega(\gamma) = \mathrm{ind}(\gamma) + \mathrm{nul}(\gamma) - (d-1). \tag{5.4}$$

Let $\Sigma \subset M$ be an embedded hypersurface diffeomorphic to a compact $(d-1)$-dimensional disk intersecting in its interior the closed geodesic γ transversely at $\gamma(0)$. As in Section 4.3, we introduce the space

$$\Sigma_k := \{\zeta \in \Lambda_k \mid \zeta(0) \in \Sigma\},$$

for an integer k large enough so that $\gamma \in \Lambda_k$. Therefore, γ is an isolated critical point of the restricted energy functional $E|_{\Sigma_k}$ with non-trivial local homology

$$H_i(\Sigma_k^{<\ell} \cup \{\gamma\}, \Sigma_k^{<\ell}) \cong H_i(\Lambda^{<\ell} \cup \{\gamma\}, \Lambda^{<\ell}) = C_i(\gamma),$$

where $\ell^2 := E(\gamma)$. Since $i = \operatorname{ind}(\gamma) + \operatorname{nul}(\gamma)$, we can find a generator of this local homology that is an embedded ball $B \subset \Sigma_k$ of dimension i containing γ in its interior, and such that $E|_{B \setminus \{\gamma\}} < \ell^2$. This can be easily seen by considering a tubular neighborhood $N \subset \Sigma_k$ of B diffeomorphic to the normal bundle of B in Σ_k, since the restriction of the energy functional E to any fiber F of N has a non-degenerate local minimizer at $F \cap B$; thus, the local homology of E at γ is isomorphic to the local homology of $E|_B$ at its local maximizer γ, and the local homology at a local maximizer is generated by the relative cycle covering the whole domain (see, e.g., [Maz13, Proposition 2.6] for a detailed proof of this general Morse-theoretic fact). This argument is independent of the choice of the coefficient field, and in particular $[B] \neq 0$ in $C_i(\gamma; \mathbb{Q})$ as well.

We consider the evaluation map $\operatorname{ev} : B \to \Sigma$, $\operatorname{ev}(\zeta) = \zeta(0)$, whose differential has the form

$$\operatorname{d}\operatorname{ev}(\gamma) : \operatorname{T}_\gamma B \to \operatorname{T}_{\gamma(0)}\Sigma, \qquad \operatorname{d}\operatorname{ev}(\gamma)\xi = \xi(0).$$

We claim that $\operatorname{d}\operatorname{ev}(\gamma)$ is surjective. Indeed, if as usual $\Omega = \{\zeta \in \Lambda \mid \zeta(0) = \gamma(0)\}$ denotes the based loop space, we have

$$\ker(\operatorname{d}\operatorname{ev}(\zeta)) = \operatorname{T}_\gamma\Omega \cap \operatorname{T}_\gamma B.$$

Since the Hessian $\operatorname{d}^2 E(\gamma)$ is negative semi-definite on $\operatorname{T}_\gamma B$, Equation (5.4) implies

$$\dim \ker(\operatorname{d}\operatorname{ev}(\zeta)) \leq \operatorname{ind}_\Omega(\gamma) + \operatorname{nul}_\Omega(\gamma)$$
$$\leq \operatorname{ind}(\gamma) + \operatorname{nul}(\gamma) - (d-1)$$
$$= \dim(B) - (d-1).$$

Since $\dim(\Sigma) = d - 1$, we infer that $\operatorname{d}\operatorname{ev}(\gamma)$ is surjective. By the implicit function theorem, up to shrinking B around γ, we find a diffeomorphism $\phi : \Sigma \times U \to B$ such that $\operatorname{ev} \circ \phi(x, y) = x$.

If $\zeta_i : [0, \tau_i] \to M$ are continuous paths such that $\zeta_i(\tau_i) = \zeta_{i+1}(0)$ for all $i \in \mathbb{Z}_m$, we define $\zeta := \zeta_0 * \ldots * \zeta_{m-1} \in \Lambda$ to be the 1-periodic curve obtained by first concatenating the ζ_i's with their original parametrization, and then by linearly reparametrizing the resulting curve so that it becomes 1-periodic. Namely,

$$\zeta(t) = \tilde{\zeta}((\tau_0 + \ldots \tau_{m-1})t),$$

where

$$\tilde{\zeta}(\tau_0 + \ldots \tau_{i-1} + u) = \zeta_i(u), \qquad \forall u \in [0, \tau_i].$$

If the ζ_i's are $W^{1/2}$ paths, the energy of $\zeta_0 * \ldots * \zeta_{m-1}$ is

$$E(\zeta_0 * \ldots * \zeta_{m-1}) = (\tau_0 + \ldots \tau_{m-1}) \sum_{i=0}^{m-1} \int_0^{\tau_i} F(\zeta_i, \dot{\zeta}_i)^2 \, \mathrm{d}t. \qquad (5.5)$$

We now employ ϕ to construct a relative cycle representing a non-zero element of the local homology group of γ^m. We first define the smooth embedding

$$\phi_m : \Sigma \times U^{\times m} \hookrightarrow \Sigma_{mk}, \qquad \phi_m(x, y_0, \ldots, y_{m-1}) = \phi(x, y_0) * \ldots * \phi(x, y_{m-1}),$$

where $U^{\times m} = U \times \ldots \times U$ denotes the m-fold cartesian product. The fact that that ϕ_m is a smooth embedding can be easily seen if we identify the loops $\zeta_i =$

$\phi_i(x, y_i) \in \Sigma_k$ with the tuple $\boldsymbol{x}_i = (\zeta_i(0), \zeta_i(\frac{1}{k}), ..., \zeta_i(\frac{k-1}{k}))$ as explained in Section 4.1: indeed, the curve $\phi_m(x, y_0, ..., y_{m-1}) \in \Sigma_{mk}$ is then identified with the juxtaposition $(\boldsymbol{x}_0, ..., \boldsymbol{x}_{m-1})$. The image of ϕ_m is an embedded ball

$$B_m := \phi_m(\Sigma \times U^{\times m}) \subset \Sigma_{mk}$$

containing γ^m in its interior. By assumption (ii) of the lemma, its dimension is bounded from below as

$$\dim(B_m) = d - 1 + m(i - (d-1)) \geq \mathrm{ind}(\gamma^m) + \mathrm{nul}(\gamma^m). \qquad (5.6)$$

Since our ζ_i's are 1-periodic loops (that is, we consider them as closed paths parametrized on $[0, 1]$), Equation (5.5) reduces to

$$E(\zeta_0 * ... * \zeta_{m-1}) = m\big(E(\zeta_0) + ... + E(\zeta_{m-1})\big).$$

Since $E(\zeta_i) < E(\gamma)$, we have

$$E|_{B_m \setminus \{\gamma^m\}} < E(\gamma^m) = m^2 \ell^2.$$

This, together with (5.6), implies that $\dim(B_m) = \mathrm{ind}(\gamma^m) + \mathrm{nul}(\gamma^m) =: i_m$, and B_m is a generator of the local homology group

$$H_{i_m}(\Sigma_{mk}^{<m\ell} \cup \{\gamma^m\}, \Sigma_{mk}^{<m\ell}; \mathbb{Q}) \cong C_{i_m}(\gamma^m; \mathbb{Q}).$$

We claim that, for each $\epsilon > 0$ sufficiently small, there exists $\overline{m} = \overline{m}_\epsilon \in \mathbb{N}$ such that, for all integers $m \geq \overline{m}$, the homomorphism

$$C_{i_m}(\gamma^m; \mathbb{Q}) \to H_{i_m}(\Lambda^{<m\ell+\epsilon/\ell}, \Lambda^{<m\ell}; \mathbb{Q})$$

induced by the inclusion is the zero one. Indeed, let $\epsilon \in (0, 1)$ be small enough so that

$$\max_{\Sigma \times \partial U} E \circ \phi < \ell^2 - \epsilon.$$

If needed, we shrink Σ around $\gamma(0)$ so that

$$\mathrm{diam}(\Sigma) := \max_{x_1, x_2 \in \Sigma} d(x_1, x_2) < \frac{\epsilon}{2(\ell^2 + 2)},$$

where $d : M \times M \to [0, \infty)$ denotes the (possibly non-symmetric) distance (4.1) induced by the Finsler metric F. Let $\delta > 0$ be such that

$$\max_{\partial \Sigma \times U} E \circ \phi = \ell^2 - \delta,$$

and notice that

$$\max_{\partial \Sigma \times U^{\times m}} E \circ \phi_m = m^2 (\ell^2 - \delta).$$

We define the continuous map

$$\psi_m : \Sigma \times \Sigma \times U^{\times \lfloor m/2 \rfloor} \times U^{\times \lceil m/2 \rceil} \to \Lambda,$$

$$\psi_m(x_1, x_2, y_1, y_2) = \phi_{\lfloor m/2 \rfloor}(x_1, y_1) * \gamma_{x_1 x_2} * \phi_{\lceil m/2 \rceil}(x_2, y_2) * \gamma_{x_2 x_1},$$

where $\gamma_{x_i x_j} : [0, d(x_i, x_j)] \to M$ is the shortest geodesic parametrized with unit speed joining x_i and x_j. Let us compute the composition $E \circ \psi_m$. If we set

$$\zeta_1 := \phi_{\lfloor m/2 \rfloor}(x_1, y_1), \quad \zeta_2 := \phi_{\lceil m/2 \rceil}(x_2, y_2), \quad \zeta := \psi_m(x_1, x_2, y_1, y_2),$$
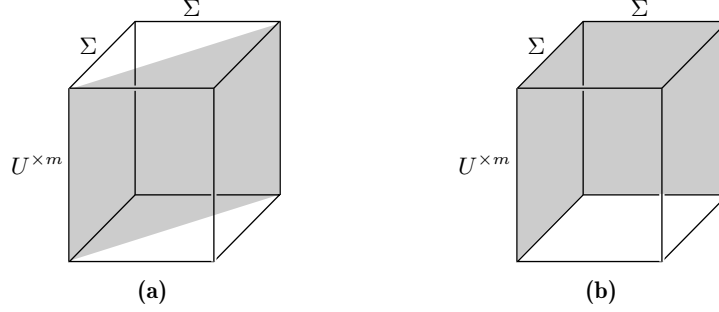
FIGURE 3. The shaded region in **(a)** is the relative cycle $\mathrm{diag}_\Sigma \times U^m$, which is null-homologous in $H_*(\Sigma \times \Sigma \times U^{\times m}, \partial(\Sigma \times \Sigma \times U^{\times m}); \mathbb{Q})$, for instance because it is homologous to the shaded region in **(b)**.

we have

$$E(\zeta) = \big(m + d(x_1, x_2) + d(x_2, x_1)\big) \left(\frac{E(\zeta_1)}{\lfloor \frac{m}{2} \rfloor} + d(x_1, x_2) + \frac{E(\zeta_2)}{\lceil \frac{m}{2} \rceil} + d(x_2, x_1) \right)$$

$$\leq \big(m + 2\,\mathrm{diam}(\Sigma)\big)\big(m\ell^2 + 2\,\mathrm{diam}(\Sigma)\big)$$

$$< m^2\ell^2 + 2\,\mathrm{diam}(\Sigma)(m + \ell^2 m + 2\,\mathrm{diam}(\Sigma))$$

$$< m^2\ell^2 + m\epsilon < (m\ell + \epsilon/\ell)^2.$$

If $y_1 \in \partial U^{\times \lfloor m/2 \rfloor}$ or $y_2 \in \partial U^{\times \lceil m/2 \rceil}$, we have the estimate

$$E(\zeta) < \big(m + 2\,\mathrm{diam}(\Sigma)\big)\big(m\ell^2 - \epsilon + 2\,\mathrm{diam}(\Sigma)\big)$$

$$= m^2\ell^2 + m\big(2\,\mathrm{diam}(\Sigma)(1 + \ell^2) - \epsilon\big) + 2\,\mathrm{diam}(\Sigma)\big(2\,\mathrm{diam}(\Sigma) - \epsilon\big)$$

$$< m^2\ell^2.$$

If instead $x_1 \in \partial\Sigma$ or $x_2 \in \partial\Sigma$, we have

$$E(\zeta) \leq \big(m + 2\,\mathrm{diam}(\Sigma)\big)\big(m\ell^2 - \lfloor \tfrac{m}{2} \rfloor \delta + 2\,\mathrm{diam}(\Sigma)\big)$$

$$\leq m^2\ell^2 + \big(2\,\mathrm{diam}(\Sigma)(2 + \ell^2) - \lfloor \tfrac{m}{2} \rfloor \delta\big)m$$

$$\leq m^2\ell^2 + \underbrace{\big(\epsilon - \lfloor \tfrac{m}{2} \rfloor \delta\big)}_{(*)} m,$$

and the term $(*)$ is negative for $m \geq \overline{m} = 2\epsilon/\delta + 2$. Summing up, our map ψ_m satisfies

$$E \circ \psi_m|_{\partial(\Sigma \times \Sigma \times U^m)} < m^2\ell^2. \tag{5.7}$$

The relative cycle $\mathrm{diag}_\Sigma \times U^m$ is null-homologous in

$$H_*(\Sigma \times \Sigma \times U^m, \partial(\Sigma \times \Sigma \times U^m); \mathbb{Q}),$$

since it is homologous to a relative cycle contained in $\partial(\Sigma \times \Sigma \times U^m)$, see Figure 3. Therefore, (5.7) implies that the relative cycle $B_m = \phi_m(\Sigma \times U^m) = \psi_m(\mathrm{diag}_\Sigma \times U^m)$ is null-homologous in $H_{i_m}(\Lambda^{<m\ell+\epsilon/\ell}, \Lambda^{<m\ell}; \mathbb{Q})$, i.e.

$$[B_m] = 0 \text{ in } H_{i_m}(\Lambda^{<m\ell+\epsilon/\ell}, \Lambda^{<m\ell}; \mathbb{Q}). \tag{5.8}$$

We now employ Morse theory. We choose a sequence of positive numbers $\epsilon_n \to 0$, and for each of them a large enough prime number m_n such that (5.8) holds for $m =$

m_n and $\epsilon = \epsilon_n$. We can assume in particular that $m_n \to \infty$. Since each m_n is prime, the inclusion induces an injective homomorphism $C_*(\gamma^{m_n}; \mathbb{Q}) \hookrightarrow C_*(S^1 \cdot \gamma^{m_n}; \mathbb{Q})$. Therefore, the commutative diagram

$$
\begin{array}{ccc}
C_{i_{m_n}}(\gamma^{m_n}; \mathbb{Q}) & \hookrightarrow & C_{i_{m_n}}(S^1 \cdot \gamma^{m_n}; \mathbb{Q}) \\
& \searrow{\scriptstyle 0} & \downarrow{\scriptstyle \iota_*} \\
& & H_{i_{m_n}}(\Lambda^{<m_n \ell + \epsilon/\ell}, \Lambda^{<m_n \ell}; \mathbb{Q})
\end{array}
$$

whose homomorphisms are all induced by the inclusion, implies that ι_* is not injective. This, in turn, implies that there exists a closed geodesic $\zeta_n \in \mathrm{crit}(E)$ with length

$$
L(\zeta_n) = E(\zeta_n)^{1/2} \in (m_n \ell, m_n \ell + \epsilon_n/\ell].
$$

We claim that the sequence $\{\zeta_n \mid n \in \mathbb{N}\}$ contains infinitely many geometrically distinct closed geodesics. Indeed, let us assume by contradiction that there exists a closed geodesic $\zeta \in \mathrm{crit}(E)$ of length $\ell_1 := E(\zeta)^{1/2}$ and a sequence $q_n \in \mathbb{N}$ such that $\zeta^{q_n} = \zeta_n$ for infinitely many $n \in \mathbb{N}$. This implies that $m_n \ell < q_n \ell_1 \leq m_n \ell + \epsilon_n/\ell$, and therefore

$$
\left\lceil \frac{m_n \ell}{\ell_1} \right\rceil < q_n \leq \left\lfloor \frac{m_n \ell}{\ell_1} + \frac{\epsilon_n}{\ell\,\ell_1} \right\rfloor.
$$

However, these two inequalities cannot simultaneously hold for n large enough so that $\epsilon_n/(\ell\,\ell_1) < 1$. $\qquad\square$

We now derive the two corollaries that we will need for proving Theorem 1.3.

Corollary 5.5. *Let (M, F) be an orientable Finsler surface, and $\gamma \in \mathrm{crit}(E)$ a closed geodesic such that $\mathrm{ind}_\Omega(\gamma) = \mathrm{nul}_\Omega(\gamma) = 1$ and the local homology $C_3(\gamma)$ with some coefficient field is non-zero. Then (M, F) has infinitely many closed geodesics.*

Proof. Since $C_3(\gamma)$ is non-trivial, we have

$$
\mathrm{ind}(\gamma) \leq 3 \leq \mathrm{ind}(\gamma) + \mathrm{nul}(\gamma).
$$

We now employ Proposition 4.2. Since $\mathrm{nul}_\Omega(\gamma) = 1$, we have

$$
\mathrm{ind}_\Omega(\gamma^m) = m\,\mathrm{ind}_\Omega(\gamma) + (m-1)\mathrm{nul}_\Omega(\gamma) = 2m - 1, \qquad \mathrm{nul}_\Omega(\gamma^m) = 1.
$$

Moreover

$$
\mathrm{ind}(\gamma) \leq \mathrm{ind}_\Omega(\gamma) + 1 = 2, \qquad \mathrm{nul}(\gamma) \geq 3 - \mathrm{ind}(\gamma) \geq 1.
$$

If $\mathrm{nul}(\gamma) = 2$, then $\mathrm{nul}(\gamma^m) = 2$ and $\mathrm{ind}(\gamma^m)$ is odd for all $m \in \mathbb{N}$; since $\mathrm{ind}_\Omega(\gamma^m) \leq \mathrm{ind}(\gamma^m) \leq \mathrm{ind}_\Omega(\gamma^m) + 1$, we infer

$$
\mathrm{ind}(\gamma^m) = \mathrm{ind}_\Omega(\gamma^m) = 2m - 1.
$$

If instead $\mathrm{nul}(\gamma) = 1$, then $\mathrm{ind}(\gamma) = 2$, and

$$
\begin{aligned}
\mathrm{ind}(\gamma^m) + \mathrm{nul}(\gamma^m) &\leq \mathrm{ind}_\Omega(\gamma^m) + \mathrm{nul}_\Omega(\gamma^m) + 1 \\
&= 2m + 1 \\
&= m(\mathrm{ind}(\gamma) + \mathrm{nul}(\gamma)) - (m - 1).
\end{aligned}
$$

In both cases, γ satisfies the assumptions of Theorem 5.4, and we infer that (S^2, F) has infinitely many closed geodesics. $\qquad\square$

The second corollary of Theorem 5.4 was established in the Riemannian case by Bangert [Ban80, Ban93]. Even though we present it here as a corollary of Theorem 5.4, Bangert's proof came historically earlier than [Hin93]. Let us recall, once again, the classical notion of conjugate points: two points $\gamma(t)$ and $\gamma(s)$ along a geodesic $\gamma : [t, s] \to M$ are conjugate when there exists a Jacobi field along γ that is not identically zero, but vanishes at both $\gamma(t)$ and $\gamma(s)$. When $\dim(M) = 2$ this condition can be expressed in terms of the Jacobi field η_t introduced in Equation (5.1): $\gamma(t)$ and $\gamma(s)$ are conjugate points if and only if $\eta_t(s) = 0$. A closed geodesic γ on a Finsler surface has no conjugate points if and only if $\mathrm{ind}_\Omega(t \cdot \gamma^m) = \mathrm{nul}_\Omega(t \cdot \gamma^m) = 0$ for all $t \in S^1$ and $m \in \mathbb{N}$; equivalently, $\mathrm{ind}(\gamma^m) = 0$ for all $m \in \mathbb{N}$, according to Lemma 4.1(iii) and Proposition 4.2(iii).

Corollary 5.6. *Any reversible Finsler 2-sphere with a simple closed geodesic without conjugate points possesses infinitely many closed geodesics.*

Proof. Let $\gamma \in \mathrm{crit}(E) \cap E^{-1}(\ell^2)$ be a simple closed geodesic without conjugate points in the reversible Finsler 2-sphere (S^2, F). We claim that there exists a neighborhood $\mathcal{U} \subset \Lambda$ of the critical circle $S^1 \cdot \gamma := \{t \cdot \gamma \mid t \in S^1\}$ such that every $\zeta \in \mathcal{U}$ that intersects the curve γ has energy $E(\zeta) \geq E(\gamma)$. Indeed, if this were not true, we could find a sequence $\zeta_n \in \Lambda$ such that $\zeta_n(0) = \gamma(t_n)$, $E(\zeta_n) < E(\gamma)$, $t_n \to t$ and $\zeta_n \to t \cdot \gamma$ as $n \to \infty$. We consider the based loop spaces

$$\Omega_s := \{\zeta \in \Lambda \mid \zeta(0) = \gamma(s)\}, \qquad s \in S^1,$$

and the space of broken closed geodesics Λ_k introduced in Section 4.1. Here $k \in \mathbb{N}$ must be large enough so that $\gamma \in \Lambda_k$. Since γ has no conjugate points, $\mathrm{ind}_\Omega(s \cdot \gamma) = \mathrm{nul}_\Omega(s \cdot \gamma) = 0$. Therefore, every $s \cdot \gamma$ is a non-degenerate local minimizer of $E|_{\Lambda_k \cap \Omega_t}$. Since $E_k := E|_{\Lambda_k}$ is smooth in a neighborhood of the critical circle of γ, we can apply the parametric Morse lemma, which provides an $\epsilon > 0$ and an open neighborhood $U \subset \Lambda_k$ of γ such that, for all $t \in (-\epsilon, \epsilon)$, $t \cdot \gamma$ is the unique global minimizer of $E|_{U \cap \Omega_t}$. Let $\gamma_n \in \Lambda_k \cap \Omega_{t_n}$ be the sequence of broken closed geodesics such that $\gamma_n(i/k) = \zeta_n(i/k)$ for all $i \in \mathbb{Z}_k$, which have energy $E(\gamma_n) \leq E(\zeta_n) < E(\gamma)$. Since $\zeta \to t \cdot \gamma$ in Λ, we would have that $\gamma_n \to \gamma$ in Λ_k, and in particular $\gamma_n \in U$ for all $n \in \mathbb{N}$ large enough, contradicting the fact that $E|_{U \cap \Omega_{t_n}}$ has a strict global minimizer at $t_n \cdot \gamma$.

We denote by B_0 and B_1 the connected components of $S^2 \setminus \gamma$, and by $\mathcal{B}_i \subset \Lambda$ the open subset of those $\zeta \in \Lambda$ such that $\zeta(S^1) \subset B_i$. Since we are looking for infinitely many closed geodesics, we can assume that γ is an isolated closed geodesic (i.e., the critical circle of each iterate γ^m is isolated in $\mathrm{crit}(E)$). We set $\ell^2 := E(\gamma)$. We have two possible cases, which we deal with separately.

Case 1: For every open neighborhood $\mathcal{V} \subset \Lambda$ of $S^1 \cdot \gamma$, the intersections $\mathcal{V} \cap \mathcal{B}_0 \cap \Lambda^{<\ell}$ and $\mathcal{V} \cap \mathcal{B}_1 \cap \Lambda^{<\ell}$ are both non-empty. If we choose $\mathcal{V}$ to be connected and contained in $\mathcal{U}$, we infer that $\mathcal{V}^{<\ell} := \mathcal{V} \cap \Lambda^{<\ell}$ is the disjoint union of the open subsets $\mathcal{V}^{<\ell} \cap \mathcal{B}_0$ and $\mathcal{V}^{<\ell} \cap \mathcal{B}_1$. In particular $\mathcal{V}^{<\ell}$ is not connected. Since $\mathcal{V}$ can be chosen arbitrarily small, we infer that the local homology $C_1(\gamma; \mathbb{Q})$ is non-zero (see Section 4.3). Since $\mathrm{ind}(\gamma^m) = 0$, Proposition 4.2(vi) implies that $\mathrm{nul}(\gamma^m) < 2$ for all $m \in \mathbb{N}$. Since the local homology $C_1(\gamma; \mathbb{Q})$ is non-zero, we must have $\mathrm{ind}(\gamma) + \mathrm{nul}(\gamma) \geq 1$, and thus $\mathrm{nul}(\gamma) = 1$. Since $\mathrm{nul}(\gamma) \leq \mathrm{nul}(\gamma^m) < 2$, we infer that $\mathrm{nul}(\gamma^m) = 1$ for all $m \in \mathbb{N}$. Therefore, γ satisfies the assumptions of Theorem 5.4, which implies that (S^2, F) has infinitely many closed geodesics.
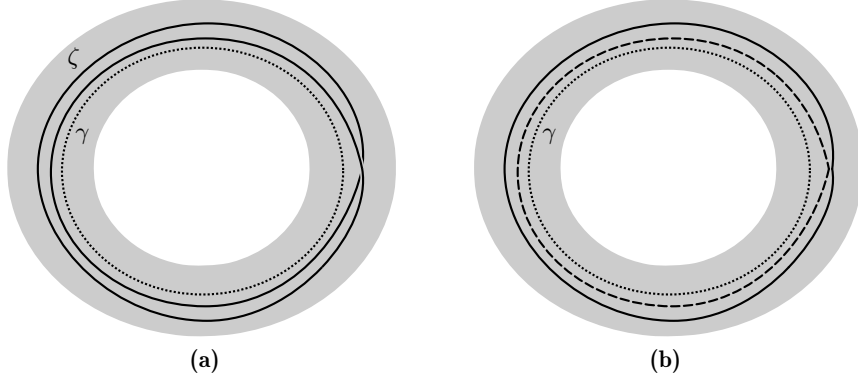
(a)   (b)

FIGURE 4. **(a)** The simple closed geodesic γ (dotted), and a curve $\zeta \in \Lambda$ close to γ^2 with a self-intersection. **(b)** The support of the curve ζ can be decomposed as the union of ζ_1 (dashed curve) and ζ_2 (solid curve), both close to γ.

Case 2: For some $i \in \{0,1\}$, there exists an open neighborhood $\mathcal{V} \subset \Lambda$ of γ such that $\mathcal{V} \cap \mathcal{B}_i \cap \Lambda^{<\ell} = \varnothing$. This implies the analogous property for γ^m: there exists an open neighborhood $\mathcal{V}_m \subset \Lambda$ of γ^m such that $\mathcal{V}_m \cap \mathcal{B}_i \cap \Lambda^{<\ell} = \varnothing$. Indeed, since we are on an orientable surface, a tubular neighborhood of the simple closed geodesic γ is diffeomorphic to the annulus $S^1 \times (-1,1)$, γ being its zero section $S^1 \times \{0\}$; therefore, any curve ζ sufficiently close to the iterated curve γ^m has at least $m-1$ self-intersections counted with multiplicity (see Figure 4), and its support can be decomposed as the union of the supports of $\zeta_1, ..., \zeta_m \in \mathcal{V}$, and $E(\zeta) = m(E(\zeta_1) + ... + E(\zeta_m))$; if $E(\zeta) < E(\gamma^m) = m^2 E(\gamma)$, we would have $E(\zeta_j) < E(\gamma)$ for some j, contradicting the fact that $\mathcal{V} \cap \mathcal{B}_i \cap \Lambda^{<\ell} = \varnothing$.

Since γ is an isolated closed geodesic, if we choose $\mathcal{V}_m$ to be small enough we have that $E(\zeta) > E(\gamma^m)$ for all $\zeta \in \mathcal{V}_m \cap \mathcal{B}_i$. Let us fix a homotopy $u : [0,1] \to \mathcal{B}_i \cup \{\gamma^m\}$, $u(t) = u_t$, such that $u_0 = \gamma^m$ and $E(u_1) = 0$; namely, u_t defines a contraction of γ to a point within the disk B_i. We choose an integer

$$k \geq \frac{\max\{E \circ u\}}{\mathrm{injrad}(S^2, F)^2},$$

and consider the space of broken closed geodesics Λ_k and the restricted energy functional $E_k = E|_{\Lambda_k}$. We have the associated retraction

$$r : \Lambda^{<\sqrt{k}\,\mathrm{injrad}(S^2,F)} \to \Lambda_k, \qquad r(\zeta) = \tilde{\zeta},$$

where $\tilde{\zeta} \in \Lambda_k$ is the broken closed geodesics such that $\tilde{\zeta}(i/k) = \zeta(i/k)$ for all $i \in \mathbb{Z}_k$. We recall that $E_k \circ r \leq E$. Since the boundary of the disk B_i is geodesic, we readily see that r preserves $\mathcal{B}_i$. Consider an open neighborhood $V \subset \Lambda_k$ of γ^m such that $\overline{V} \subset \mathcal{V}_m$. Since ∂V is compact, we have

$$b^2 := \min_{\partial V} E_k > E_k(\gamma^m).$$

Let $W \subset V$ be a small enough open neighborhood of γ^m such that

$$E_k(\gamma^m) < a^2 := \max_{\overline{W}} E_k < b^2.$$

We consider

$$c^2 := \inf_v \max\{E_k \circ v\},$$

where the infimum ranges over all homotopies $v : [0,1] \to \Lambda_k \cap \mathcal{B}_i \cup \{\gamma^m\}$, $v(t) = v_t$, such that $v_0 = \gamma^m$ and $E_k(v_1) = 0$. Notice that the space of such homotopies is non-empty, as it contains $r \circ u$. We have $c > b$, since every such a homotopy v must eventually intersect ∂W. We fix an arbitrary $d \in (c, k \operatorname{injrad}(S^2, F)^{1/2})$. Notice that $E_k^{-1}[a^2, d^2] \cap \mathcal{B}_i$ is compact, since it is a connected component of the compact set $E_k^{-1}[a^2, d^2] \setminus W$. Therefore, the classical min-max theorem implies that c^2 is a critical value of E_k. Since we are looking for infinitely many closed geodesics, we can assume that (S^2, F) has only isolated closed geodesic (i.e. any critical circle is isolated in $\operatorname{crit}(E) \cap E^{-1}(0, \infty)$). Under this assumption, there exists at least one closed geodesic $\zeta_m \in \operatorname{crit}(E_k) \cap E_k^{-1}(c^2) \cap \mathcal{B}_i$ such that every connected open neighborhood $U \subset \Lambda_k \cap \mathcal{B}_i$ of it has a non-connected intersection $U^{<c} = U \cap \Lambda_k^{<c}$. Indeed, if no closed geodesic in $\operatorname{crit}(E_k) \cap E_k^{-1}(c^2) \cap \mathcal{B}_i$ satisfied this property, we could find a homotopy v as above such that $\max\{E_k \circ v\} < c^2$, contradicting the definition of the min-max value c^2. Since U can be chosen arbitrarily small, we infer that the local homology $C_1(\zeta_m; \mathbb{Q})$ is non-trivial.

Now, either the family $\{\zeta_m \mid m \in \mathbb{N}\}$ that we found contains infinitely many geometrically distinct closed geodesics, or there exists a non-iterated closed geodesic ζ, an infinite subset $\mathbb{K} \subset \mathbb{N}$, and a function $\mu : \mathbb{K} \to \mathbb{N}$ such that $\mu(m) \to \infty$ as $m \to \infty$ and $\zeta_m = \zeta^{\mu(m)}$ for all $m \in \mathbb{N}$. Since every iterate $\zeta^{\mu(m)}$ has non-trivial local homology $C_1(\zeta^{\mu(m)})$, we have $\operatorname{ind}(\zeta^{\mu(m)}) \leq 1$ for all $m \in \mathbb{K}$, and therefore $\operatorname{ind}(\zeta^m) = 0$ for all $m \in \mathbb{N}$ according to Lemma 4.1(iii). We cannot have $\operatorname{nul}(\zeta) = 0$, for otherwise ζ would be a local minimizer of E_k, and the same would be true for all its iterates according to analogous argument of Figure 4. Therefore $1 \leq \operatorname{nul}(\zeta) \leq \operatorname{nul}(\zeta^m) \leq 2$ for all $m \in \mathbb{N}$. By Proposition 4.2(vi), since the Morse indices $\operatorname{ind}(\zeta^m)$ vanish, we must have $\operatorname{nul}(\zeta^m) = 1$ for all $m \in \mathbb{N}$.

Since $\operatorname{ind}(\zeta) = \operatorname{ind}(\zeta^m)$ and $\operatorname{nul}(\zeta) = \operatorname{nul}(\zeta^m)$ for all $m \in \mathbb{N}$, we have an isomorphism of local homology groups $C_*(\zeta; \mathbb{Q}) \cong C_*(\zeta^m; \mathbb{Q})$. In particular $C_1(\zeta; \mathbb{Q})$ does not vanish. Therefore, ζ satisfies the assumptions of Theorem 5.4, which implies that (S^2, F) has infinitely many closed geodesics. $\square$

5.4. **Bangert's theorem.** As it turns out, the statements proved so far allow us to conclude the existence of infinitely many closed geodesics on any reversible (S^2, F), except when none of its simple closed geodesics has a well-defined Birkhoff map. We recall that a simple closed geodesic γ of a reversible (S^2, F) does not have a well defined Birkhoff map when, for some $x = \gamma(t)$ and $v \in T_x S^2$ transverse to $\dot\gamma(t)$, the geodesic $\zeta(t) = \exp_x(tv)$ does not intersect γ at any positive time $t > 0$. In this section, we show that this last case is covered by Corollary 5.6. For Riemannian 2-spheres, this is a theorem due to Bangert [Ban93].

Theorem 5.7. *Any reversible Finsler 2-sphere having a simple closed geodesic without a well-defined Birkhoff map possesses infinitely many closed geodesics.*

The proof is based on the following two lemmas of independent interest.

Lemma 5.8. *Let (M, F) be a (not necessarily reversible) Finsler surface, and $\gamma : [-T, T] \to M$ a geodesic parametrized with constant speed. If there exists a sequence of geodesics $\gamma_n : [-T, T] \to M$ parametrized with constant speed, not intersecting γ,*

and such that $(\gamma_n(0), \dot{\gamma}_n(0)) \to (\gamma(0), \dot{\gamma}(0))$ in TM, then $\gamma|_{(-T,T)}$ has no conjugate points.

Proof. Since the problem is local to γ, we can assume without loss of generality that $M = \mathbb{R}^2$ and $\gamma(t) = (t, 0)$ for all $t \in [-T, T]$, so that we can write expressions in coordinates. Without loss of generality, we can assume that $F(\gamma, \dot{\gamma}) \equiv 1$. We reparametrize the geodesics γ_n so that they have speed $F(\gamma_n, \dot{\gamma}_n) \equiv 1$. By doing this, we change the interval of definition of γ_n: the reparametrized curve has the form $\gamma_n : [-T_n, T_n] \to M$ with $T_n \to T$ as $n \to \infty$.

We set $(x, v) = (\gamma(0), \dot{\gamma}(0)) = (0, \dot{\gamma}(0))$, $G = \frac{1}{2}F^2$, and consider the line

$$\Sigma := \{y \in \mathbb{R}^2 \mid y \in \ker G_v(x, v)\}.$$

For all $n \in \mathbb{N}$ large enough, the geodesic $\gamma_n : [-T_n, T_n] \to \mathbb{R}^2$ intersects Σ in a unique point x_n, and clearly $x_n \to x$. We shift the parametrization of γ_n, so that we have a sequence of geodesics $\gamma_n : [-T_n + \epsilon_n, T_n + \epsilon_n] \to \mathbb{R}^2$ not intersecting γ and such that $\epsilon_n \to 0$ and $v_n := \dot{\gamma}_n(0) \to v$. Up to extracting a subsequence, we can assume that each γ_n lies on the same side of γ. Therefore we have a well defined non-zero vector

$$w' := \frac{x_n - x}{\|x_n - x\|} \in \Sigma$$

independent of n. Here, $\|\cdot\|$ denotes the Euclidean norm. We consider the vectors

$$z_n := \frac{v_n - v}{\|x_n - x\|}.$$

Since the geodesics γ_n and γ do not intersect on the time interval $[-T/2, T/2]$ and the second derivative of $\gamma_n - \gamma$ is uniformly bounded on $[-T/2, T/2]$ independently of n, we readily obtain that the sequence $\|z_n\|$ is uniformly bounded from above. In particular, up to extracting a subsequence, we have $z_n \to z'$ as $n \to \infty$.

We set $\lambda := (1 + \|z'\|^2)^{-1/2}$, $w := \lambda w'$, and $z := \lambda z'$, so that

$$\lim_{n \to \infty} \frac{(x_n, v_n) - (x, v)}{\|(x_n, v_n) - (x, v)\|} = \lim_{n \to \infty} \frac{(x_n, v_n) - (x, v)}{\|x_n - x\|\sqrt{1 + \|z_n\|^2}} = (w, z) \in T_{(x,v)}S\mathbb{R}^2,$$

where $S\mathbb{R}^2 = \{(x', v') \in T\mathbb{R}^2 \mid F(x', v') = 1\}$. Therefore,

$$\lim_{n \to \infty} \frac{\phi_t(x_n, v_n) - \phi_t(x, v)}{\|(x_n, v_n) - (x, v)\|} = \mathrm{d}\phi_t(x, v)(w, z).$$

The Jacobi field $\zeta : (-T, T) \to \mathbb{R}^2$ along γ defined by

$$(\zeta(t), \dot{\zeta}(t)) = \mathrm{d}\phi_t(x, v)(w, z)$$

satisfies $G_v(\gamma(t), \dot{\gamma}(t))\zeta(t) = 0$.

We claim that ζ is nowhere vanishing. Indeed, let $\nu : (-T, T) \to \mathbb{R}^2$ the smooth vector field along γ defined by $\nu(t) \in \ker G_v(\gamma(t), \dot{\gamma}(t))$, $\|\nu(t)\| = 1$, and $\nu(t)$ pointing to the the side of γ containing the γ_n's. We can write $\zeta(t) = z(t)\nu(t)$ for some continuous function $z = (-T, T) \to \mathbb{R}$. Notice that

$$\zeta(t) = \lim_{n \to \infty} \frac{\gamma_n(t) - \gamma(t)}{\|(x_n, v_n) - (x, v)\|}.$$

If $\zeta(t) = 0$ for some t, since $\gamma_n(t) - \gamma(t)$ and $\nu(t)$ point to the same side of γ, we readily obtain that $z(t) = 0$ and $\dot{z}(t) = 0$. But this would imply that $\zeta(t) =$

$\dot{\zeta}(t) = 0$, and since ζ is a Jacobi field we would conclude that ζ vanishes identically, contradicting $\zeta(0) = w$.

If γ had conjugate points $\gamma(t_1), \gamma(t_2)$ for some $-T < t_1 < t_2 < T$, there would exists a Jacobi field $\eta : [t_1, t_2] \to \mathbb{R}^2$ such that $\eta(t_1) = \eta(t_2) = 0$, $\dot{\eta}(t_1) \neq 0$, and $\eta(t) \in \ker G_v(\gamma(t), \dot{\gamma}(t))$ for all $t \in [t_1, t_2]$. Since we are on a surface, $\ker G_v(\gamma(t), \dot{\gamma}(t))$ is 1-dimensional. Therefore, by the Sturm separation theorem, η and ζ would have alternating zeroes, contradicting the fact that ζ is nowhere vanishing. $\square$

Proof of Theorem 5.7. Let $\gamma_0 \in \mathrm{crit}(E) \cap E^{-1}(0, \infty)$ be a simple closed geodesic that does not have a well defined Birkhoff map. We only need to consider the case in which γ_0 has conjugate points (i.e. $\mathrm{ind}_\Omega(\gamma^m) > 0$ for some integer $m \geq 1$), for otherwise the existence of infinitely many closed geodesics is already provided by Corollary 5.6. The fact that γ_0 does not have a well defined Birkhoff map means that, for some $x_0 = \gamma_0(t_0)$ and $v_0 \in S_x S^2$ transverse to $\dot{\gamma}_0(t_0)$, the geodesic $\zeta : (0, \infty) \to S^2$, $\zeta(t) = \exp_{x_0}(tv_0)$ does not intersect γ_0 in positive time, and therefore stays trapped in a connected component $B \subset S^2 \setminus \gamma(S^1)$. We consider the compact subset

$$K := \bigcap_{t > 0} \overline{\zeta[t, \infty)} \subset \overline{B}.$$

We claim that $K \cap \gamma_0(S^1) = \varnothing$. Otherwise we can find a sequence $t_n \to \infty$ and $s \in \mathbb{R}$ such that $\zeta(t_n) \to \gamma(s)$. Since ζ does not intersect γ in positive time, up to extracting a subsequence we must have $\dot{\zeta}(t_n) \to \dot{\gamma}_0(s)$. Since γ_0 has conjugate points, there exists $\delta > 0$ such that $\gamma_0|_{(s-\delta, s+\delta)}$ has conjugate points. Lemma 5.8 thus provides a contradiction: since $\zeta|_{[t_n - \delta, t_n + \delta]}$ does not intersect $\gamma_0|_{[s-\delta, s+\delta]}$, $\gamma_0|_{(s-\delta, s+\delta)}$ cannot have conjugate points.

Let $U \subset B \setminus K$ be the connected component whose closure contains $\gamma_0(S^1)$. One would expect this open set to be locally geodesically convex. We prove a slightly weaker convexity: for all $x, y \in U$ that can be joined by means of an absolutely continuous curve in U of length strictly less than $\rho := \mathrm{injrad}(S^2, F)$, the shortest geodesic joining x and y is entirely contained in U. Indeed, let $\gamma_{x,y} : [0, 1] \to S^2$, $\gamma_{x,y}(t) = \exp_x(t \exp_x^{-1}(y))$ be such a geodesic, and assume by contradiction that some $z = \gamma_{x,y}(s)$ belongs to K. Then, by the definition of K, there exists a sequence $t_n \to \infty$ such that $\zeta(t_n) \to z$. Up to extracting a subsequence, the sequence $\dot{\zeta}(t_n)$ converges to some $w \in S_z S^2$ that is transverse to $\dot{\gamma}_{x,y}(s)$, since the geodesic $\theta : \mathbb{R} \to S^2$, $\theta(t) = \exp_z(tw)$ is entirely contained in K. We denote the geodesic balls centered at z by

$$B(z, r) := \left\{ z' \in S^2 \mid d(z, z') < \mathrm{injrad}(S^2, F)/2 \right\}, \qquad r > 0,$$

where $d : S^2 \times S^2 \to [0, \infty)$ is the distance (4.1) induced by the Finsler metric F. The points x and y are contained in different connected components of $B(z, \rho) \setminus \zeta(-\rho, \rho)$. Therefore, every continuous curve $\theta : [0, 1] \to U$ such that $\theta(0) = x$ and $\theta(1) = y$ must leave the geodesic ball $B(z, \rho)$; since $d(x, z) + d(z, y) < \rho$, we readily obtain that the length of such a θ is larger than ρ, contradicting our assumption on x, y.

We consider the space

$$W := \left\{ \gamma \in \Lambda \mid \gamma(S^1) \subset U, \ \gamma \text{ not contractible in } U, \ E(\gamma) < E(\gamma_0) \right\}$$

We claim that W is not empty. Indeed, since γ_0 has conjugate points, by Proposition 4.2(vii) there exists a nowhere vanishing 1-periodic vector field ξ along γ such that $\mathrm{d}^2 E(\gamma)[\xi, \xi] < 0$, and $\xi(t)$ points inside U for all $t \in S^1$. We define $\gamma_s \in \Lambda$ by

$$\gamma_s(t) = \exp_{\gamma_0(t)}(s\xi(t)).$$

If $s > 0$ is small enough, then γ_s is contained in U, non-contractible in U (since it is homotopic to γ_0 within $U \cup \gamma_0(S^1)$), and since

$$E(\gamma_s) \leq E(\gamma_0) + \tfrac{1}{2}s^2\, \mathrm{d}^2 E(\gamma)[\xi, \xi] + o(s^2)$$

we have $E(\gamma_s) < E(\gamma_0)$. Thus any such γ_s belongs to W.

We fix $k \in \mathbb{N}$ large enough so that γ_0 is contained in the space of broken closed geodesics $\Lambda_k \subset \Lambda$. We define the continuous map $r : W \to \Lambda_k$ by $r(\gamma)(\frac{i}{k}) = \gamma(\frac{i}{k})$ for all $i \in \mathbb{Z}_k$. The above convexity property of U implies that, for each $\gamma \in W$, $r(\gamma)$ is a curve contained in U and homotopic to γ within U. Therefore, r is a retraction $r : W \to W \cap \Lambda_k$. Since $E(r(\gamma)) \leq E(\gamma)$, we have

$$\ell^2 := \inf_W E = \inf_{W \cap \Lambda_k} E.$$

We choose a sequence $\gamma_n \in W \cap \Lambda_k$ such that $E(\gamma_n) \to \ell^2$. We can assume that each γ_n is without self-intersection. Indeed, if γ_n has self-intersections, we can find an interval $[a, b] \subsetneq [0, 1]$ such that $\gamma_n|_{[a,b]}$ is a non-contractible loop. If i_0, i_1 are positive integers such that

$$[\tfrac{i_0+1}{k}, \tfrac{i_1-1}{k}] \subseteq [a, b] \subseteq [\tfrac{i_0}{k}, \tfrac{i_1}{k}],$$

we define $\tilde{\gamma}_n \in W \cap \Lambda_k$ by setting $\tilde{\gamma}_n(\frac{i}{k}) = \gamma_n(a)$ for all $i \in \{0, ..., i_0\} \cup \{i_1, ..., k-1\}$, and $\tilde{\gamma}_n(\frac{i}{k}) = \gamma_n(\frac{i}{k})$ for all $i \in \{i_0 + 1, ..., i_1 - 1\}$. The curve $\tilde{\gamma}_n$ has less self-intersections than γ_n, and energy $E(\tilde{\gamma}_n) \leq E(\gamma_n)$. Since a broken closed geodesic has only finitely many self intersections, by repeating this procedure a finite number of times we eliminate all of them.

Since $\overline{W \cap \Lambda_k}$ is compact, up to extracting a subsequence we have that

$$\gamma_n \to \gamma \in \overline{W \cap \Lambda_k},$$

and $E(\gamma) = \ell^2$. We claim that γ is a closed geodesic. This is clear if γ is contained in $W \cap \Lambda_k$, for in this case it would be a critical point of the energy functional E. Assume now that $\gamma \in \partial(W \cap \Lambda_k)$, and consider the unique $\theta, \theta_n \in \Lambda_k$ such that

$$\theta(\tfrac{i}{k}) = \gamma(\tfrac{i+1/2}{k}), \quad \theta_n(\tfrac{i}{k}) = \gamma_n(\tfrac{i+1/2}{k}), \qquad \forall i \in \mathbb{Z}_k.$$

Clearly, $\theta_n \to \theta$. Moreover, $E(\theta) \leq E(\gamma)$ with equality if and only if γ is a closed geodesic. The above convexity property of U implies that $\theta_n \in W \cap \Lambda_k$. Therefore $E(\theta_n) \geq \inf E|_W = E(\gamma)$ and $E(\theta) = E(\gamma)$, and we conclude that γ is a closed geodesic.

Since the approximating loops γ_n are without self-intersections, γ is a simple closed geodesic. Therefore, the union $\gamma_0(S^1) \cup \gamma(S^1)$ bounds an open annulus $A \subset U$. Since $E(\gamma) = \inf E|_W$, in particular there is no $\tilde{\gamma} \in \Lambda$ with energy $E(\tilde{\gamma}) < E(\gamma)$ and support $\tilde{\gamma}(S^1) \subset A$. Therefore, by applying Proposition 4.2(vii) as above, we infer that γ has no conjugate points. Corollary 5.6 implies that (S^2, F) has infinitely many closed geodesics. $\qquad\square$

Proof of Theorem 1.3. By Theorem 1.2, if (S^2, F) has only finitely many simple closed geodesics, there exists at least one simple closed geodesic $\gamma \in \mathrm{crit}(E)$ with non-zero local homology $C_3(\gamma; \mathbb{Z}_2)$. If γ does not have a well defined Birkhoff map,

Theorem 5.7 implies that there are infinitely many closed geodesics. Assume now that γ has a well defined Birkhoff map. Since $C_3(\gamma; \mathbb{Z}_2)$ is non-zero, $C_3(t \cdot \gamma; \mathbb{Z}_2)$ is non-zero as well, and

$$\mathrm{ind}(t \cdot \gamma) \leq 3 \leq \mathrm{ind}(t \cdot \gamma) + \mathrm{nul}(t \cdot \gamma), \qquad \forall t \in S^1.$$

By Proposition 4.2(iv), we have

$$\mathrm{ind}_\Omega(t \cdot \gamma) + \mathrm{nul}_\Omega(t \cdot \gamma) \geq \mathrm{ind}(t \cdot \gamma) + \mathrm{nul}(t \cdot \gamma) - 1 \geq 2, \qquad \forall t \in S^1.$$

Since $\mathrm{nul}_\Omega(t \cdot \gamma) \leq 1$, this inequality implies that $\mathrm{ind}_\Omega(t \cdot \gamma) \geq 1$. If $\mathrm{ind}_\Omega(t \cdot \gamma) \geq 2$ for all $t \in S^1$, Theorem 5.3 implies that there are infinitely many closed geodesics. If instead $\mathrm{ind}_\Omega(t \cdot \gamma) = 1$ for some $t \in S^1$, the above inequality implies that $\mathrm{nul}_\Omega(t \cdot \gamma) = 1$, and Corollary 5.5 implies that there are infinitely many closed geodesics. $\square$

REFERENCES

[AM18]   L. Asselle and M. Mazzucchelli, *Closed geodesics with local homology in maximal degree on non-compact manifolds*, Differential Geom. Appl. **58** (2018), 17–51.

[Ang90] S. Angenent, *Parabolic equations for curves on surfaces. I. Curves with p-integrable curvature*, Ann. of Math. (2) **132** (1990), no. 3, 451–483.

[Ang91] _____, *Parabolic equations for curves on surfaces. II. Intersections, blow-up and generalized solutions*, Ann. of Math. (2) **133** (1991), no. 1, 171–215.

[Ang08] _____, *Self-intersecting geodesics and entropy of the geodesic flow*, Acta Math. Sin. (Engl. Ser.) **24** (2008), no. 12, 1949–1952.

[Bal78] W. Ballmann, *Der Satz von Lusternik und Schnirelmann*, Beiträge zur Differentialgeometrie, Heft 1, Bonner Math. Schriften, vol. 102, Univ. Bonn, Bonn, 1978, pp. 1–25.

[Ban80] V. Bangert, *Closed geodesics on complete surfaces*, Math. Ann. **251** (1980), no. 1, 83–96.

[Ban93] _____, *On the existence of closed geodesics on two-spheres*, Internat. J. Math. **4** (1993), no. 1, 1–10.

[Bir66] G. D. Birkhoff, *Dynamical systems*, American Mathematical Society Colloquium Publications, Vol. IX, American Mathematical Society, Providence, R.I., 1966.

[BL10]   V. Bangert and Y. Long, *The existence of two closed geodesics on every Finsler 2-sphere*, Math. Ann. **346** (2010), no. 2, 335–366.

[Bot56] R. Bott, *On the iteration of closed geodesics and the Sturm intersection theory*, Comm. Pure Appl. Math. **9** (1956), 171–206.

[BTZ82] W. Ballmann, G. Thorbergsson, and W. Ziller, *Closed geodesics on positively curved manifolds*, Ann. of Math. **116** (1982), no. 2, 213–247.

[CJM11] E. Caponio, M. Á. Javaloyes, and A. Masiello, *On the energy functional on Finsler manifolds and applications to stationary spacetimes*, Math. Ann. **351** (2011), no. 2, 365–392.

[Fra92] J. Franks, *Geodesics on S^2 and periodic points of annulus homeomorphisms*, Invent. Math. **108** (1992), no. 2, 403–418.

[GM69a] D. Gromoll and W. Meyer, *On differentiable functions with isolated critical points*, Topology **8** (1969), 361–369.

[GM69b] _____, *Periodic geodesics on compact Riemannian manifolds*, J. Differ. Geom. **3** (1969), 493–510.

[Gra89] M. A. Grayson, *Shortening embedded curves*, Ann. Math. **129** (1989), 71–111.

[Had98] M. Hadamard, *Les surfaces á courbures opposées et leurs lignes géodésiques*, J. Math. Pures Appl. **5** (1898), no. 4, 27–74.

[Hin93] N. Hingston, *On the growth of the number of closed geodesics on the two-sphere*, Internat. Math. Res. Notices (1993), no. 9, 253–262.

[Kat73] A. B. Katok, *Ergodic perturbations of degenerate integrable Hamiltonian systems*, Izv. Akad. Nauk SSSR Ser. Mat. **37** (1973), 539–576.

[Kli95] W. P. A. Klingenberg, *Riemannian geometry*, second ed., De Gruyter Studies in Mathematics, vol. 1, Walter de Gruyter & Co., Berlin, 1995.

[Lju66] L. A. Ljusternik, *The topology of the calculus of variations in the large*, Translated from the Russian by J. M. Danskin. Translations of Mathematical Monographs, Vol. 16, American Mathematical Society, Providence, R.I., 1966.

[LS34]   L. Lusternik and L. Schnirelmann, *Méthodes topologiques dans les problèmes variation-nels*, Hermann, Paris, 1934.

[Lu15]   G. Lu, *Methods of infinite dimensional Morse theory for geodesics on Finsler manifolds*, Nonlinear Anal. **113** (2015), 230–282.

[Maz13] M. Mazzucchelli, *Symplectically degenerate maxima via generating functions*, Math. Z. **275** (2013), no. 3-4, 715–739.

[Maz16] _____, *The Morse index of Chaperon's generating families*, Publ. Mat. Urug. **16** (2016), 81–125.

[MM12]   C. Mantegazza and L. Martinazzi, *A note on quasilinear parabolic equations on mani-folds*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **11** (2012), no. 4, 857–874.

[Mor96] M. Morse, *The calculus of variations in the large*, American Mathematical Society Colloquium Publications, vol. 18, American Mathematical Society, Providence, RI, 1996, Reprint of the 1932 original.

[MS18]   M. Mazzucchelli and S. Suhr, *A characterization of Zoll Riemannian metrics on the 2-sphere*, Bull. Lond. Math. Soc. **50** (2018), 997–1006.

[Neu77] W. D. Neumann, *Generalizations of the Poincaré Birkhoff fixed point theorem*, Bull. Austral. Math. Soc. **17** (1977), no. 3, 375–389.

[Oak94] J. A. Oaks, *Singularities and self-intersections of curves evolving on surfaces*, Indiana Univ. Math. J. **43** (1994), no. 3, 959–981.

[Poi05] H. Poincaré, *Sur les lignes géodésiques des surfaces convexes*, Trans. Amer. Math. Soc. **6** (1905), 237–274.

[Rad92] H.-B. Rademacher, *Morse-Theorie und geschlossene Geodätische*, Bonner Mathematische Schriften, 229, Universität Bonn Mathematisches Institut, Bonn, 1992.

[VPS76] M. Vigué-Poirrier and D. Sullivan, *The homology theory of the closed geodesic problem*, J. Differential Geometry **11** (1976), no. 4, 633–644.

[Zil83] W. Ziller, *Geometry of the Katok examples*, Ergodic Theory Dynam. Systems **3** (1983), no. 1, 135–157.

Guido De Philippis
Courant Institute of Mathematical Sciences, New York University
251 Mercer Street, New York, N.Y. 10012-1185, United States
Email address: guido@cims.nyu.edu

Michele Marini
Scuola Internazionale Superiore di Studi Avanzati
via Bonomea 265, 34136 Trieste, Italy
Email address: mmarini@sissa.it

Marco Mazzucchelli
CNRS, École Normale Supérieure de Lyon, UMPA
46 allée d'Italie, 69364 Lyon Cedex 07, France
Email address: marco.mazzucchelli@ens-lyon.fr

Stefan Suhr
Ruhr Universität Bochum, Fakultät für Mathematik
Universitätsstrasse 150, 44801 Bochum, Germany
Email address: stefan.suhr@rub.de