A systolic inequality for geodesic flows on the two-sphere

Alberto Abbondandolo, Barney Bramham,
Umberto L. Hryniewicz and Pedro A. S. Salomão

Tuesday 28th October, 2014

Abstract

Given a smooth Riemannian two-sphere (S^2, g), consider $\ell_{\min}(S^2, g)$ defined as the minimum of all lengths of non-constant closed geodesics. Our main result asserts that if g is δ-pinched for some $\delta > (4+\sqrt{7})/8 = 0.8307\ldots$ then the systolic inequality $\ell_{\min}(S^2, g)^2 \leq \pi\ \mathrm{Area}(S^2, g)$ holds, with equality if and only if (S^2, g) is Zoll. The proof is based on Toponogov's comparison theorem and on a theorem relating the Calabi invariant to the action of fixed points for certain area-preserving annulus maps admitting a generating function.

Contents

Introduction

The 1-systole $\mathrm{Sys}_1(M,g)$ of a Riemannian manifold (M,g) is the infimum of lengths of non-contractible closed loops. The origin of systolic geometry can be traced back to a classical result due to Loewner, asserting that for every Riemannian 2-torus $(\mathbb{T}^2,g)$

$$\mathrm{Sys}_1(\mathbb{T}^2,g)^2 \leq \frac{2}{\sqrt{3}}\ \mathrm{Area}(\mathbb{T}^2,g),$$

and to Pu's inequality, asserting that the inequality

$$\mathrm{Sys}_1(\mathbb{RP}^2,g)^2 \leq \frac{\pi}{2}\ \mathrm{Area}(\mathbb{RP}^2,g)$$

holds for every Riemannian metric g on the real projective plane $\mathbb{RP}^2$.

In the eighties Gromov [Gro83] introduced the filling radius of a Riemannian manifold and showed that there exists a constant $C_n > 0$ such that

$$\mathrm{Sys}_1(M,g)^n \leq C_n\ \mathrm{Vol}(M,g)$$

holds for all n-dimensional closed aspherical Riemannian manifolds. This theorem also holds on so-called essential manifolds.

The number Sys_1 is a critical value of the length functional. It is also interesting to analyze critical values of the length on simply connected Riemannian manifolds, which may not be attained by local minima.

Given a Riemannian metric g on the two-sphere S^2, a relevant quantity is

$$\ell_{\min}(S^2,g) = \text{minimum of lengths of non-constant closed geodesics on } (S^2,g).$$

A deep result by Croke [Cro88] asserts the existence of a number $C > 0$ such that

$$\ell_{\min}(S^2,g)^2 \leq C\ \mathrm{Area}(S^2,g),$$

for every metric g on S^2. In other words, the *systolic ratio*

$$\rho_{\mathrm{sys}}(g) := \frac{\ell_{\min}(S^2,g)^2}{\mathrm{Area}(S^2,g)}$$

is bounded from above. The value of the supremum of ρ is not known, but it was shown to be not larger than 32 by Rotman [Rot06], who improved the previous estimates due to Croke [Cro88], Nabutowski and Rotman [NR02], and Sabourau [Sab04].

The naïve conjecture that the round metric g_{round} on S^2 maximizes ρ_{sys} is false. Indeed,

$$\rho_{\mathrm{sys}}(g_{\mathrm{round}}) = \pi,$$

while, by studying suitable metrics approximating a singular metric constructed by gluing two flat equilateral triangles along their boundaries, one sees that

$$\sup \rho_{\mathrm{sys}} \geq 2\sqrt{3} > \pi.$$

This singular example is known as the Calabi-Croke sphere. Actually, it is conjectured that the supremum of ρ_{sys} is $2\sqrt{3}$ and that it is not attained. See [Bal10] and [Sab10] for two different proofs of the fact that the Calabi-Croke sphere can be seen as a local maximum of ρ_{sys}.

The round metric can be seen as a critical point of ρ_{sys}, and in [Bal06] Balacheff asked if it is a local maximizer (in [BM13, Question 8.7.2] this question is attributed to Babenko). Certainly, g_{round} is not a strict local maximiser of ρ_{sys}, even after modding out rescaling, because in any neighbourhood of it there are infinitely many non-isometric Zoll metrics, i.e. Riemannian metrics on S^2 all of whose geodesics are closed and have the same length, and ρ_{sys} is constantly equal to π on them (see [Gui76]). Evidence in favour of the local maximality of the round metric is given in [APB14], where Álvarez Paiva and Balacheff prove that ρ_{sys} strictly decreases under infinitesimal deformations of the round metric which are not tangent with infinite order to the space of Zoll metrics.

The aim of this paper is to give a positive answer to Balacheff's question. We recall that a Riemannian metric g on S^2 is δ-pinched, for some $\delta \in (0,1]$, if its Gaussian curvature K is positive and satisfies

$$\min K \geq \delta \max K.$$

Our main theorem asserts that, if the metric g on S^2 is sufficiently pinched, then

$$\rho_{\mathrm{sys}}(g) \leq \rho_{\mathrm{sys}}(g_{\mathrm{round}}) = \pi,$$

and that the equality holds if and only if g is Zoll. More precisely, we shall prove the following:

THEOREM. *Let g be a δ-pinched smooth Riemannian metric on S^2, with*

$$\delta > \frac{4+\sqrt{7}}{8} = 0.8307\ldots$$

Then

$$\ell_{\min}(S^2,g)^2 \leq \pi \,\mathrm{Area}(S^2,g),$$

and the equality holds if and only if g is Zoll.

We conclude this introduction with an informal description of the proof of this theorem. We start by looking at a closed geodesic γ on (S^2,g) of minimal length $L = \ell_{\min}(S^2,g)$, parametrized by arc length. When g is δ-pinched for some $\delta > 1/4$, one can show that γ is a simple curve.

Then we consider a Birkhoff annulus Σ_γ^+ which is associated to γ: Σ_γ^+ is the set of all unit tangent vectors to S^2 which are based at points of $\gamma(\mathbb{R})$ and point in the direction of one of the two disks which compose $S^2 \setminus \gamma(\mathbb{R})$. The set Σ_γ^+ is a closed annulus, and its boundary consists of the unit vectors $\dot\gamma(t)$ and $-\dot\gamma(t)$, for $t \in \mathbb{R}/L\mathbb{Z}$.

By a famous result of Birkhoff, the positivity of the curvature K guarantees that the orbit of any v in the interior part of Σ_γ^+ under the geodesic flow on the unit tangent bundle T^1S^2 of (S^2,g) hits again Σ_γ^+ at some positive time. This allows us to consider the first return time function

$$\tau : \mathrm{int}(\Sigma_\gamma^+) \to (0,+\infty), \qquad \tau(v) := \inf\{t>0 \mid \phi_t(v) \in \Sigma_\gamma^+\},$$

and the first return time map

$$\varphi : \mathrm{int}(\Sigma_\gamma^+) \to \mathrm{int}(\Sigma_\gamma^+), \qquad \varphi(v) := \phi_{\tau(v)}(v),$$

where $\phi_t : T^1S^2 \to T^1S^2$ denotes the geodesic flow induced by g. The function τ and the map φ are smooth and, as we will show, have a unique smooth extension to the boundary of Σ_γ^+.

The map φ preserves the two-form $d\lambda$, where λ is the restriction to Σ_γ^+ of the standard contact form on T^1S^2. The two-form $d\lambda$ is an area-form in the interior of Σ_γ^+, but vanishes on the boundary, due to the fact that the geodesic flow is not transverse to the boundary. Indeed, if we consider the coordinates

$$(x, y) \in \mathbb{R}/L\mathbb{Z} \times [0, \pi]$$

on Σ_γ^+ given by the arc parameter x on the geodesic γ and the angle y which a unit tangent vector makes with $\dot{\gamma}$, the one-form λ and its differential have the form

$$\lambda = \cos y \, dx, \qquad d\lambda = \sin y \, dx \wedge dy. \tag{1}$$

By lifting the first return map φ to the strip $S = \mathbb{R} \times [0, \pi]$, we obtain a diffeomorphism $\Phi : S \to S$ which preserves the two-form $d\lambda$ given by (1), maps each boundary component into itself, and satisfies

$$\Phi(x + L, y) = (L, 0) + \Phi(x, y), \qquad \forall (x, y) \in S.$$

As we shall see, diffeomorphisms of S with these properties have a well defined *flux* and, when the flux vanishes, a well defined *Calabi invariant*. The *flux* of Φ is its average horizontal displacement. We shall prove that, if g is δ-pinched with $\delta > 1/4$, one can find a lift Φ of φ having zero flux. For diffeomorphisms Φ with zero flux, the action and the Calabi invariant can be defined in the following way. The *action* of Φ is the unique function

$$\sigma : S \to \mathbb{R},$$

such that

$$d\sigma = \Phi^*\lambda - \lambda \qquad \text{on } S,$$

and whose value at each boundary point $w \in \partial S$ coincides with the integral of λ on the arc from w to $\Phi(w)$ along ∂S. The *Calabi invariant* of Φ is the average of the action, that is, the number

$$\mathrm{CAL}(\Phi) = \frac{1}{2L} \iint_{[0,L]\times[0,\pi]} \sigma \, d\lambda.$$

We shall prove that, still assuming g to be δ-pinched with $\delta > 1/4$, the action and the Calabi invariant of Φ are related to the geometric quantities we are interested in by the identities

$$\begin{aligned} \tau \circ p &= L + \sigma, && (2)\\ \pi \,\mathrm{Area}(S^2, g) &= L^2 + L \,\mathrm{CAL}(\Phi), && (3) \end{aligned}$$

where

$$p : S = \mathbb{R} \times [0, \pi] \to \Sigma_\gamma^+ = \mathbb{R}/L\mathbb{Z} \times [0, \pi]$$

is the standard projection. The δ-pinching assumption on g with $\delta > (4 + \sqrt{7})/8$ implies that the map Φ is *monotone*, meaning that, writing

$$\Phi(x, y) = (X(x, y), Y(x, y)),$$

the strict inequality $D_2 Y > 0$ holds on S. This is proved by using an upper bound on the perimeter of convex geodesic polygons which follows from Topogonov's comparison theorem. This upper bound plays an important role also in the proof of some of the facts stated above, and we discuss it in the appendix which concludes this article. The monotonicity of Φ allows us to express it in terms of a generating function. By using such a generating function, we shall prove the following fixed point theorem: If a monotone map Φ with vanishing flux is not the identity and satisfies $\mathrm{CAL}(\Phi) \leq 0$, then Φ has an interior fixed point with negative action.

Our main theorem is now a consequence of the latter fixed point theorem and of the identities (2) and (3). First one observes that Φ is the identity if and only if g is Zoll. Assume that g is not Zoll. If, by contradiction, the inequality

$$L^2 = \ell_{\min}(S^2, g)^2 \geq \pi \,\mathrm{Area}(S^2, g)$$

holds, (3) implies that $\mathrm{CAL}(\Phi) \leq 0$, so Φ has a fixed point $w \in \mathrm{int}(S)$ with $\sigma(w) < 0$. But then (2) implies that the closed geodesic which is determined by $p(w) \in \Sigma_\gamma^+$ has length $\tau(p(w)) < L$, which is a contradiction, because L is the minimal length of a closed geodesic. This shows that when g is not Zoll, the strict inequality

$$\ell_{\min}(S^2, g)^2 < \pi \,\mathrm{Area}(S^2, g)$$

holds, concluding the proof.

A final remark on the pinching hypothesis. Since the smooth metrics approximating the Calabi-Croke sphere can be chosen to have positive curvature, the above theorem cannot hold for every positively curved metric, and it is natural to ask under which value of the pinching constant δ it may fail. Our pinching constant $(4 + \sqrt{7})/8$ is most probably not optimal. Indeed, in many of our arguments it is enough to assume that $\delta > 1/4$, a condition which has clear geometrical implications: For instance, under this assumption, every geodesic ray emanating from the Birkhoff annulus does not intersect itself before hitting the annulus again. The stronger pinching assumption $\delta > (4 + \sqrt{7})/8$ is used only to guarantee that the lift of the first return map is monotone, and hence has a generating function. Proving a suitable generalization of our fixed point theorem to non-monotone maps would allow one to prove the systolic optimality of the round sphere among all δ-pinched metrics with $\delta > 1/4$.

Acknowledgments. The present work is part of the first author's activities within CAST, a Research Network Program of the European Science Foundation. U.H. is grateful to Samuel Senti for endless interesting discussions about the relations between systolic inequalities and ergodic theory, and for his interest in the paper. U.H. also acknowledges support from CNPq Grant 309983/2012-6. P.S. is partially supported by CNPq Grant no. 301715/2013-0 and FAPESP Grant no. 2013/20065-0

1 A class of self-diffeomorphisms of the strip preserving a two-form

We denote by S the closed strip

$$S := \mathbb{R} \times [0, \pi],$$

on which we consider coordinates (x, y), $x \in \mathbb{R}$, $y \in [0, \pi]$. The smooth two-form

$$\omega(x, y) := \sin y \, dx \wedge dy$$

is an area form on the interior of S and vanishes on its boundary. Fix some $L > 0$, and let $\mathcal{D}_L(S, \omega)$ be the group of all diffeomorphisms $\Phi : S \to S$ such that:

(i) $\Phi(x + L, y) = (L, 0) + \Phi(x, y)$ for every $(x, y) \in S$.

(ii) Φ maps each component of ∂S into itself.

(iii) Φ preserves the two-form ω.

The elements of $\mathcal{D}_L(S, \omega)$ are precisely the maps which are obtained by lifting to the universal cover

$$S \to A := \mathbb{R}/L\mathbb{Z} \times [0, \pi]$$

self-diffeomorphisms of A which preserve the two-form ω on A and map each boundary component into itself.

By conjugating an element Φ of $\mathcal{D}_L(S, \omega)$ by the homeomorphism

$$S \to \mathbb{R} \times [-1, 1], \qquad (x, y) \mapsto (x, -\cos y),$$

one obtains a self-homeomorphism of the strip $\mathbb{R} \times [-1, 1]$ which preserves the standard area form $dx \wedge dy$. Such a homeomorpshism is in general not continuously differentiable up to the boundary. Since we find it more convenient to work in the smooth category, we prefer not to use the above conjugacy and to deal with the non-standard area-form ω vanishing on the boundary.

1.1 The flux and the Calabi invariant

In this section, we define the flux on $\mathcal{D}_L(S, \omega)$ and the Calabi homomorphism on the kernel of the flux. These real valued homomorphisms were introduced by Calabi in [Cal70] for the group of compactly supported symplectic diffeomorphisms of symplectic manifolds of arbitrary dimension. See also [MS98, Chapter 10]. In this paper we need to extend these definitions to the surface with boundary S. Our presentation is self-contained.

Definition 1.1. *The flux of a map $\Phi \in \mathcal{D}_L(S, \omega)$, $\Phi(x, y) = (X(x, y), Y(x, y))$, is the real number*

$$\mathrm{FLUX}(\Phi) := \frac{1}{2L} \iint_{[0,L]\times[0,\pi]} (X(x, y) - x)\, \omega(x, y).$$

In other words, the flux of Φ is the average shift in the horizontal direction (notice that $2L$ is the total area of $[0, L] \times [0, \pi]$ with respect to the area form ω). Using the fact that the elements of $\mathcal{D}_L(S, \omega)$ preserve ω, it is easy to show that the function $\mathrm{FLUX} : \mathcal{D}_L(S, \omega) \to \mathbb{R}$ is a homomorphism.

Proposition 1.2. *Let $\alpha_0 : [0, \pi] \to S$ be the path $\alpha_0(t) := (0, t)$. Then*

$$\mathrm{FLUX}(\Phi) = \frac{1}{2} \int_{\Phi(\alpha_0)} x \sin y \, dy,$$

for every Φ in $\mathcal{D}_L(S, \omega)$.

Proof. Let $\Theta : S \to S$ be the covering transformation $(x, y) \mapsto (x + L, y)$, and set $Q := [0, L] \times [0, \pi]$. With its natural orientation, $Q \subset S$ is the region whose signed boundary is $\Theta(\alpha_0) - \alpha_0$ plus pieces that lie in ∂S. Since $\Phi \in \mathcal{D}_L(S, \omega)$ commutes with Θ, we have

$$\Phi(Q) - Q = \Theta(R) - R \tag{4}$$

as simplicial 2-chains in S, where $R \subset S$ is an oriented region whose signed boundary consists of $\Phi(\alpha_0) - \alpha_0$ plus two additional pieces in ∂S that we do not need to label. Therefore,

$$\mathrm{FLUX}(\Phi) = \frac{1}{2L} \int_Q (X - x)\,\omega = \frac{1}{2L} \int_Q \big(\Phi^*(x\,\omega) - x\,\omega\big) = \frac{1}{2L} \int_R \big(\Theta^*(x\,\omega) - x\,\omega\big),$$

using (4) for the last equality. Since

$$\Theta^*(x\,\omega) - x\,\omega = L\,\omega = L\,d\big(x \sin y\, dy\big),$$

by Stokes theorem we conclude that

$$\mathrm{FLUX}(\Phi) = \frac{1}{2} \int_{\partial R} x \sin y\, dy = \frac{1}{2} \int_{\Phi(\alpha_0) - \alpha_0} x \sin y\, dy = \frac{1}{2} \int_{\Phi(\alpha_0)} x \sin y\, dy.$$

□

REMARK 1.3. *More generally, it is not difficult to show that if α is any smooth path in S with the first end-point in $\mathbb{R} \times \{0\}$ and the second one in $\mathbb{R} \times \{\pi\}$, then*

$$\mathrm{FLUX}(\Phi) = \frac{1}{2} \int_{\Phi(\alpha)} x \sin y\, dy - \frac{1}{2} \int_\alpha x \sin y\, dy,$$

for every Φ in $\mathcal{D}_L(S, \omega)$.

Now we fix the following primitive of ω on S

$$\lambda := \cos y\, dx.$$

Notice that λ is invariant with respect to translations in the x-direction. Let Φ be an element of $\mathcal{D}_L(S, \omega)$. Since Φ preserves $\omega = d\lambda$, the one-form

$$\Phi^*\lambda - \lambda$$

is closed. Since S is simply connected, there exists a unique smooth function

$$\sigma : S \to \mathbb{R}$$

such that

$$d\sigma = \Phi^*\lambda - \lambda \qquad \text{on } S, \tag{5}$$

and

$$\sigma(0, 0) = \int_{\gamma_0} \lambda - \mathrm{FLUX}(\Phi), \tag{6}$$

where γ_0 is a smooth path in ∂S going from $(0, 0)$ to $\Phi(0, 0)$. Of course, the value of the integral in (6) does not depend on the choice of γ_0, but only on its end-points.

Notice that the function σ is L-periodic in the first variable: This follows from the fact that $\Phi^*\lambda - \lambda$ is L-periodic in the first variable and its integral on the path $\beta_0 : [0, L] \to S$, $\beta_0(t) = (t, 0)$, vanishes:

$$\int_{\beta_0} (\Phi^*\lambda - \lambda) = \int_{\Phi(\beta_0)} \lambda - \int_{\beta_0} \lambda = \int_{\Phi(0,0)+\beta_0} \lambda - \int_{\beta_0} \lambda = 0,$$

thanks to the invariance of λ with respect to horizontal translations (here, the L-periodicity of λ in the first variable would have sufficed).

Notice also that, thanks to (5), the same normalization condition (6) holds for every point in the lower component of the boundary of S: For every x in $\mathbb{R}$ there holds

$$\sigma(x, 0) = \int_{\gamma_x} \lambda - \mathrm{FLUX}(\Phi), \tag{7}$$

where γ_x is a smooth path in ∂S going from $(x, 0)$ to $\Phi(x, 0)$. Indeed, if ξ_x is a smooth path in ∂S from $(0, 0)$ to $(x, 0)$, then the paths $\gamma_0 \#(\Phi \circ \xi_x)$ and $\xi_x \# \gamma_x$ in ∂S have the same end-points. Thus,

$$\int_{\gamma_0} \lambda + \int_{\xi_x} \Phi^*\lambda = \int_{\xi_x} \lambda + \int_{\gamma_x} \lambda,$$

and equations (5) and (6) imply

$$\sigma(x, 0) = \sigma(0, 0) + \int_{\xi_x} d\sigma = \int_{\gamma_0} \lambda - \mathrm{FLUX}(\Phi) + \int_{\xi_x} (\Phi^*\lambda - \lambda) = \int_{\gamma_x} \lambda - \mathrm{FLUX}(\Phi).$$

Therefore, we can give the following definitions.

Definition 1.4. *Let $\Phi \in \mathcal{D}_L(S, \omega)$. The unique smooth function $\sigma : S \to \mathbb{R}$ which satisfies (5) and (6) (or, equivalently, (5) and (7)) is called action of Φ.*

Definition 1.5. *Let $\Phi \in \ker \mathrm{FLUX}$ and let σ be the action of Φ. The Calabi invariant of Φ is the real number*

$$\mathrm{CAL}(\Phi) = \frac{1}{2L} \iint_{[0,L]\times[0,\pi]} \sigma\, \omega.$$

In other words, the Calabi invariant of Φ is its average action. The following remark explains why we define the Calabi invariant only for diffeomorphisms having zero flux.

Remark 1.6. *The action σ depends on the choice of the primitive λ of ω. Let λ' be another primitive of ω, still L-periodic in the first variable. Then one can easily show that $\lambda' = \lambda + df + c\, dx$, where $f : S \to \mathbb{R}$ is a smooth function which is L-periodic in the first variable and c is a real number, and that the action σ' of Φ with respect to λ' is given by*

$$\sigma'(x, y) = \sigma(x, y) + f \circ \Phi(x, y) - f(x, y) + c(X(x, y) - x),$$

where $\Phi = (X, Y)$. If Φ has zero flux, then the integrals of $\sigma'\, \omega$ and of $\sigma\, \omega$ on $[0, L] \times [0, \pi]$ coincide, so the Calabi invariant of Φ does not depend on the choice of the periodic primitive of ω. Moreover, this formula also shows that the value of the action at a fixed point of Φ is independent on the choice of the primitive of ω. Since $\Phi^*\lambda$ is *another periodic primitive of ω, the above facts imply that* $\mathrm{CAL} : \ker \mathrm{FLUX} \to \mathbb{R}$ *is a homomorphsim. In this paper, we work always with the chosen primitive λ of ω and do not need the homomorphsim property of* CAL*, so we leave these verifications to the reader. See [Fat80] and [GG95] for interesting equivalent definitions of the Calabi invariant in the case of compactly supported area preserving diffeomorphisms of the plane.*

In our definition of the action, we have chosen to normalize σ by looking at the lower component of ∂S. The following result describes what happens on the upper component.

PROPOSITION 1.7. *Let $\Phi \in \mathcal{D}_L(S, \omega)$ and let $\sigma : S \to \mathbb{R}$ be its action. Let δ_x be a smooth path in ∂S going from (x, π) to $\Phi(x, \pi)$. Then*

$$\sigma(x, \pi) = \int_{\delta_x} \lambda + \mathrm{FLUX}(\Phi).$$

Proof. The same argument used in the paragraph above Definition 1.4 shows that it is enough to check the formula for $x = 0$. In this case, by integrating over the path $\alpha_0 : [0, \pi] \to S$, $\alpha_0(t) := (0, t)$, we find by Stokes theorem

$$\begin{aligned}\sigma(0, \pi) &= \sigma(0, 0) + \int_{\alpha_0} d\sigma = \int_{\gamma_0} \lambda - \mathrm{FLUX}(\Phi) + \int_{\alpha_0} (\Phi^*\lambda - \lambda) \\ &= \int_{\gamma_0} \lambda - \mathrm{FLUX}(\Phi) + \int_{\Phi(\alpha_0)} \lambda + \int_{\alpha_0^{-1}} \lambda = \int_{\delta_0} \lambda - \mathrm{FLUX}(\Phi) + \iint_R h^*(d\lambda),\end{aligned}$$

where $h : R \to S$ is a smooth map on a closed rectangle R whose restriction to the boundary is given by the concatenation $\gamma_0 \#(\Phi \circ \alpha_0) \# \delta_0^{-1} \# \alpha_0^{-1}$. By using again Stokes theorem with the primitive $x \sin y \, dy$ of $\omega = d\lambda$, we get

$$\iint_R h^*(d\lambda) = \int_{\gamma_0 \#(\Phi \circ \alpha_0) \# \delta_0^{-1} \# \alpha_0^{-1}} x \sin y \, dy = \int_{\Phi(\alpha_0)} x \sin y \, dy.$$

By Proposition 1.2, the latter quantity coincides with twice the flux of Φ, and the conclusion follows. □

1.2 Generating functions

As it is well known, area-preserving self-diffeomorphisms of the strip which satisfy a suitable monotonicity condition can be represented in terms of a generating function. See for instance [MS98, Chapter 9]. Here we need to review these facts in the case of diffeomorphims preserving the special two-form $\omega = \sin y \, dx \wedge dy$.

DEFINITION 1.8. *The diffeomorphism $\Phi = (X, Y)$ in $\mathcal{D}_L(S, \omega)$ is said to be monotone if $D_2 Y(x, y) > 0$ for every $(x, y) \in S$.*

Assume that $\Phi = (X, Y) \in \mathcal{D}_L(S, \omega)$ is a monotone map. Then the map

$$\Psi : S \to S, \qquad \Psi(x, y) = \big(x, Y(x, y)\big)$$

is a diffeomorphism: This follows from the fact that its differential at every point is invertible, thanks to the monotonicity assumption, and from the fact that Ψ fixes the boundary. Denoting by y the second component of the inverse of Ψ, we can work with coordinates (x, Y) on S and consider the one-form

$$\eta(x, Y) = (\cos Y - \cos y)\, dx + (X - x) \sin Y\, dY \qquad \text{on } S.$$

From the fact that Φ preserves ω we find

$$\begin{aligned} d\eta &= \sin Y\, dx \wedge dY - \sin y\, dx \wedge dy + \sin Y\, dX \wedge dY - \sin Y\, dx \wedge dY \\ &= - \sin y\, dx \wedge dy + \sin Y\, dX \wedge dY = 0, \end{aligned}$$

so η is closed. Let $W = W(x, Y)$ be a primitive of η. Then also $(x, y) \mapsto W(x + L, y)$ is a primitive of η, and hence

$$W(x + L, Y) - W(x, Y) = c, \qquad \forall (x, Y) \in S,$$

for some real number c. Since the integral of η on any path in ∂S connecting $(0, 0)$ to $(L, 0)$ vanishes, the constant c must be zero, and hence any primitive W of η is L-periodic. By writing

$$dW(x, Y) = D_1 W(x, Y)\, dx + D_2 W(x, Y)\, dY,$$

and using the definition of η, we obtain the following:

Proposition 1.9. *Assume that Φ in $\mathcal{D}_L(S, \omega)$ is a monotone map. Then there exists a smooth function $W : S \to \mathbb{R}$ such that the following holds: $\Phi(x, y) = (X, Y)$ if and only if*

$$\begin{aligned} (X - x) \sin Y &= D_2 W(x, Y), &(8)\\ \cos Y - \cos y &= D_1 W(x, Y). &(9) \end{aligned}$$

The function W is L-periodic in the first variable. It is uniquely defined up to the addition of a real constant.

A function W as above is called a generating function of Φ. Equation (9) implies that W is constant on each of the two connected components of the boundary of S. The difference between these two constant values coincides with twice the flux of Φ:

Proposition 1.10. *If W is a generating function of the monotone map $\Phi \in \mathcal{D}_L(S, \omega)$, then*

$$\mathrm{FLUX}(\Phi) = \frac{1}{2} \left(W|_{\mathbb{R} \times \{\pi\}} - W|_{\mathbb{R} \times \{0\}} \right).$$

Proof. By Proposition 1.2 and (8) we compute

$$\begin{aligned} \mathrm{FLUX}(\Phi) &= \frac{1}{2} \int_{\Phi(\alpha_0)} x \sin y\, dy = \frac{1}{2} \int_{\alpha_0} X \sin Y\, dY = \frac{1}{2} \int_{\alpha_0} (X - x) \sin Y\, dY \\ &= \frac{1}{2} \int_{\alpha_0} D_2 W(x, Y)\, dY = \frac{1}{2} \left(W|_{\mathbb{R} \times \{\pi\}} - W|_{\mathbb{R} \times \{0\}} \right), \end{aligned}$$

where we have used the fact that $x = 0$ on the path α_0 which is defined in Proposition 1.2. □

By the above proposition, we can choose the free additive constant of the generating function W in such a way that:

$$W|_{\mathbb{R}\times\{0\}} = -\mathrm{FLUX}(\Phi), \qquad W|_{\mathbb{R}\times\{\pi\}} = \mathrm{FLUX}(\Phi). \tag{10}$$

We conclude this section by expressing the action and the Calabi invariant of a monotone element of $\mathcal{D}_L(S,\omega)$ in terms of its generating function, normalized by the above condition.

PROPOSITION 1.11. *Let $\Phi = (X,Y) \in \mathcal{D}_L(S,\omega)$ be a monotone map, and denote by W the generating function of Φ normalized by (10). Then we have:*

(i) The action of Φ is the function

$$\sigma(x,y) = W(x,Y(x,y)) + D_2W(x,Y(x,y)) \cot Y(x,y).$$

(ii) If moreover $\mathrm{FLUX}(\Phi) = 0$*, then the Calabi invariant of Φ is the number*

$$\mathrm{CAL}(\Phi) = \frac{1}{2L}\iint_{[0,L]\times[0,\pi]} \big(W(x,y) + W(x,Y(x,y))\big)\, \omega(x,y).$$

The formula for σ in (i) is valid only in the interior of S, because the cotangent function diverges at 0 and π. Since D_2W vanishes on the boundary of S, thanks to (8), this formula defines a smooth function on S by setting

$$\sigma(x,0) = W(x,0) + D_{22}W(x,0), \qquad \sigma(x,\pi) = W(x,\pi) + D_{22}W(x,\pi),$$

for every $x \in \mathbb{R}$.

Proof. Let us check that the function σ which is defined in (i) coincides with the action of Φ. By (8) we have

$$\sigma = W + D_2W \cot Y = W + (X-x)\cos Y \tag{11}$$

on $\mathrm{int}(S)$. By continuity, this formula for σ is valid on the whole S. By differentiating it and using again (8) together with (9), we obtain

$$\begin{aligned} d\sigma &= dW - (X-x)\sin Y\, dY + \cos Y(dX - dx) \\ &= dW - D_2W\, dY + \cos Y(dX-dx) = D_1W\, dx + \cos Y(dX-dx) \\ &= (\cos Y - \cos y)\, dx + \cos Y(dX-dx) = \cos Y\, dX - \cos y\, dx = \Phi^*\lambda - \lambda. \end{aligned}$$

Therefore, σ satisfies (5). Evaluating (11) in $(0,0)$ we find

$$\sigma(0,0) = W(0,0) + X(0,0) = -\mathrm{FLUX}(\Phi) + X(0,0) = -\mathrm{FLUX}(\Phi) + \int_{\gamma_0} \lambda,$$

where γ_0 is a path in ∂S going from $(0,0)$ to $\Phi(0,0)$. We conclude that σ satisfies also (6), and hence coincides with the action of Φ. This proves (i).

We now use (i) in order to compute the integral of the two form $\sigma\,\omega$ on $[0,L]\times[0,\pi]$. We start from the identity

$$\begin{aligned} \iint_{[0,L]\times[0,\pi]} \sigma\,\omega = &\iint_{[0,L]\times[0,\pi]} W(x,Y(x,y))\,\omega(x,y) \\ &+ \iint_{[0,L]\times[0,\pi]} D_2W(x,Y(x,y)) \cot Y(x,y) \sin y\, dx\wedge dy, \end{aligned} \tag{12}$$

and we manipulate the last integral. By differentiating (9), that is, the identity

$$\cos Y(x,y) - \cos y = D_1 W(x, Y(x,y)),$$

we obtain

$$\sin y\, dy = \sin Y\, dY + D_{11} W\, dx + D_{12} W\, dY.$$

By the above formula, the integrand in the last integral in (12) can be rewritten as

$$\begin{aligned} D_2 W \cot Y \sin y\, dx \wedge dy &= D_2 W \cot Y\, dx \wedge (\sin Y\, dY + D_{12} W\, dY) \\ &= D_2 W \cos Y\, dx \wedge dY + D_2 W D_{12} W \cot Y\, dx \wedge dY. \end{aligned} \tag{13}$$

We integrate the above two forms separately. By the L-periodicity in x, the integral of the first two-form can be manipulated as follows:

$$\begin{aligned} &\iint_{[0,L]\times[0,\pi]} D_2 W(x, Y(x,y)) \cos Y(x,y)\, dx \wedge dY(x,y) \\ &\quad = \iint_{[0,L]\times[0,\pi]} D_2 W(x,Y) \cos Y\, dx \wedge dY \\ &\quad = \int_0^L \left(\int_0^\pi D_2 W(x,Y) \cos Y\, dY \right) dx \\ &\quad = \int_0^L \left(\Big[W(x,Y) \cos Y \Big]_{Y=0}^{Y=\pi} + \int_0^\pi W(x,Y) \sin Y\, dY \right) dx \\ &\quad = -L\big(W|_{\mathbb{R}\times\{\pi\}} + W|_{\mathbb{R}\times\{0\}} \big) + \iint_{[0,L]\times[0,\pi]} W(x,Y) \sin Y\, dx \wedge dY \\ &\quad = -L\big(-\mathrm{FLUX}(\Phi) + \mathrm{FLUX}(\Phi) \big) + \iint_{[0,L]\times[0,\pi]} W(x,y)\, \sin y\, dx \wedge dy \\ &\quad = \iint_{[0,L]\times[0,\pi]} W(x,y)\, \omega(x,y), \end{aligned} \tag{14}$$

where we have used the normalization condition (10). The integral of the second form in the right-hand side of (13) vanishes, because

$$\begin{aligned} &\iint_{[0,L]\times[0,\pi]} D_2 W D_{12} W \cot Y\, dx \wedge dY \\ &\qquad = \frac{1}{2} \iint_{[0,L]\times[0,\pi]} D_1 (D_2 W)^2 \cot Y\, dx \wedge dY \\ &\qquad = \frac{1}{2} \int_0^\pi \cot Y \left(\int_0^L D_1 (D_2 W)^2\, dx \right) dY = 0, \end{aligned} \tag{15}$$

by L-periodicity in x. By (12), (13), (14) and (15) we obtain

$$\iint_{[0,L]\times[0,\pi]} \sigma\, \omega = \iint_{[0,L]\times[0,\pi]} \big(W(x, Y(x,y)) + W(x,y) \big)\, \omega(x,y),$$

and (ii) follows. $\square$

1.3 The Calabi invariant and the action at fixed points

We are now in the position to prove the main result of this first part.

THEOREM 1.12. *Let Φ be a monotone element of $\mathcal{D}_L(S,\omega)$ which is different from the identity and has zero flux. If* $\mathrm{CAL}(\Phi) \leq 0$*, then Φ has an interior fixed point with negative action.*

Proof. Let W be the generating function of Φ normalised by the condition (10). Since Φ has zero flux, this condition says that W is zero on the boundary of S. Since Φ is not the identity, W is not identically zero. Then the condition $\mathrm{CAL}(\Phi) \leq 0$ and the formula of Proposition 1.11 (ii) for $\mathrm{CAL}(\Phi)$ imply that W is somewhere negative. Being a continuous periodic function, W achieves its minimum at some interior point $(x, Y) \in \mathrm{int}(S)$. Since the differential of W vanishes at (x, Y), equations (8) and (9) imply that $(x, y) := (x, Y)$ is a fixed point of Φ. By Proposition 1.11 (i),

$$\sigma(x, y) = W(x, Y) < 0.$$

Therefore, (x, y) is an interior fixed point of Φ with negative action. $\square$

2 The geodesic flow on a positively curved two-sphere

Throughout this section, a smooth oriented Riemannian two-sphere (S^2, g) is fixed. The associated unit tangent bundle is

$$T^1S^2 := \{v \in TS^2 \mid g_{\pi(v)}(v, v) = 1\},$$

where $\pi : TS^2 \to S^2$ denotes the bundle projection. For each $v \in T^1S^2$, we denote by $v^\perp \in T_{\pi(v)}S^2$ the unit vector perpendicular to v such that $\{v, v^\perp\}$ is a positive basis of $T_{\pi(v)}S^2$.

We shall deal always with Riemannian metrics g having positive Gaussian curvature K and shall often use Klingenberg's lower bound on the injectivity radius $\mathrm{inj}(g)$ of the metric g from [Kli59], that is,

$$\mathrm{inj}(g) \geq \frac{\pi}{\sqrt{\max K}}, \tag{16}$$

see also [Kli82, Theorem 2.6.9].

2.1 Extension and regularity of the Birkhoff map

Let $\gamma : \mathbb{R}/L\mathbb{Z} \to S^2$ be a simple closed geodesic of length L parametrized by arc-length, i.e. satisfying $g_\gamma(\dot\gamma, \dot\gamma) \equiv 1$. The smooth unit vector field $\dot\gamma^\perp$ along γ determines the Birkhoff annuli

$$\begin{aligned} \Sigma_\gamma^+ &:= \{\cos y\ \dot\gamma(x) + \sin y\ \dot\gamma^\perp(x) \in T^1S^2 \mid (x, y) \in \mathbb{R}/L\mathbb{Z} \times [0, \pi]\}, \\ \Sigma_\gamma^- &:= \{\cos y\ \dot\gamma(x) + \sin y\ \dot\gamma^\perp(x) \in T^1S^2 \mid (x, y) \in \mathbb{R}/L\mathbb{Z} \times [-\pi, 0]\}. \end{aligned} \tag{17}$$

These sets are embedded closed annuli and (x, y) are smooth coordinates on them. The annuli Σ_γ^+ and Σ_γ^- intersect along their boundaries $\partial\Sigma_\gamma^+ = \partial\Sigma_\gamma^-$. This common boundary has two components, one containing unit vectors $\dot\gamma$ and the other containing unit vectors $-\dot\gamma$. We denote the open annuli by

$$\mathrm{int}(\Sigma_\gamma^+) := \Sigma_\gamma^+ \setminus \partial\Sigma_\gamma^+, \qquad \mathrm{int}(\Sigma_\gamma^-) := \Sigma_\gamma^- \setminus \partial\Sigma_\gamma^-.$$

Let ϕ_t be the geodesic flow on T^1S^2. We define the functions

$$\begin{aligned} &\tau_+ : \mathrm{int}(\Sigma_\gamma^+) \to (0,+\infty], \qquad \tau_+(v) := \inf\{t>0 \mid \phi_t(v) \in \mathrm{int}(\Sigma_\gamma^-)\},\\ &\tau_- : \mathrm{int}(\Sigma_\gamma^-) \to (0,+\infty], \qquad \tau_-(v) := \inf\{t>0 \mid \phi_t(v) \in \mathrm{int}(\Sigma_\gamma^+)\},\end{aligned}$$

where the infimum of the empty set is $+\infty$. The functions τ_+ and τ_- are the transition times to go from the interior of Σ_γ^+ to the interior of Σ_γ^- and the other way round. The first return time to Σ_γ^+ is instead the function

$$\tau : \mathrm{int}(\Sigma_\gamma^+) \to (0,+\infty], \qquad \tau(v) := \inf\{t>0 \mid \phi_t(v) \in \mathrm{int}(\Sigma_\gamma^+)\}.$$

Recall the following celebrated theorem due to Birkhoff (see also [Ban93]):

THEOREM 2.1 (Birkhoff [Bir27]). *If the Gaussian curvature of g is everywhere positive then the functions τ_+, τ_- and τ are everywhere finite.*

Thanks to the above result, we have the transition maps

$$\begin{aligned} &\varphi_+ : \mathrm{int}(\Sigma_\gamma^+) \to \mathrm{int}(\Sigma_\gamma^-), \qquad \varphi_+(v) := \phi_{\tau_+(v)}(v),\\ &\varphi_- : \mathrm{int}(\Sigma_\gamma^-) \to \mathrm{int}(\Sigma_\gamma^+), \qquad \varphi_-(v) := \phi_{\tau_-(v)}(v),\end{aligned}$$

and the first return map

$$\varphi : \mathrm{int}(\Sigma_\gamma^+) \to \mathrm{int}(\Sigma_\gamma^+), \qquad \varphi(v) := \phi_{\tau(v)}(v).$$

By construction,

$$\varphi = \varphi_- \circ \varphi_+, \tag{18}$$
$$\tau = \tau_+ + \tau_- \circ \varphi_+. \tag{19}$$

Using the implicit function theorem and the fact that the geodesic flow is transverse to both $\mathrm{int}(\Sigma_\gamma^+)$ and $\mathrm{int}(\Sigma_\gamma^-)$, one easily proves that the functions τ^+, τ_- and τ are smooth. These functions have smooth extensions to the closure of their domains. More precisely, we have the following statement.

PROPOSITION 2.2. *Assume that the Gaussian curvature of (S^2,g) is everywhere positive. Then:*

(i) The functions τ_+ and τ_- can be smoothly extended to Σ_γ^+ and Σ_γ^-, respectively, as follows: $\tau_+(\dot\gamma(x)) = \tau_-(\dot\gamma(x))$ is the time to the first conjugate point along the geodesic ray $t \in [0,+\infty) \mapsto \gamma(x+t)$, and $\tau_+(-\dot\gamma(x)) = \tau_-(-\dot\gamma(x))$ is the time to the first conjugate point along the geodesic ray $t \in [0,+\infty) \mapsto \gamma(x-t)$.

(ii) The function τ can be smoothly extended to Σ_γ^+ as follows: $\tau(\dot\gamma(x))$ is the time to the second conjugate point along the geodesic ray $t \in [0,+\infty) \mapsto \gamma(x+t)$, and $\tau(-\dot\gamma(x))$ is the time to the second conjugate point along the geodesic ray $t \in [0,+\infty) \mapsto \gamma(x-t)$.

The smooth extensions of τ^+, τ_- and τ are denoted by the same symbols. The above proposition has the following consequence:

Corollary 2.3. *Suppose that the Gaussian curvature of (S^2, g) is everywhere positive. Then the formulas*

$$v \mapsto \phi_{\tau_+(v)}(v), \qquad v \mapsto \phi_{\tau_-(v)}(v) \quad \textit{and} \quad v \mapsto \phi_{\tau(v)}(v)$$

define smooth extensions of the maps φ_+, φ_- and φ to diffeomorphisms

$$\varphi_+ : \Sigma_\gamma^+ \to \Sigma_\gamma^-, \qquad \varphi_- : \Sigma_\gamma^- \to \Sigma_\gamma^+ \quad \textit{and} \quad \varphi : \Sigma_\gamma^+ \to \Sigma_\gamma^+,$$

which still satisfy (18) and (19).

Proof. The smoothness of the geodesic flow ϕ and of the functions τ_+, τ_- and τ imply that φ_+, φ_- and φ are smooth. Since the inverses of these maps on the interior of their domains have analogous definitions, such as for instance

$$\varphi_+^{-1}(v) = \phi_{\hat\tau_+(v)}(v), \qquad \text{where} \qquad \hat\tau_+(v) := \sup\{t < 0 \mid \phi_t(v) \in \mathrm{int}(\Sigma_\gamma^+)\},$$

the maps φ_+^{-1}, φ_-^{-1} and φ^{-1} have also smooth extensions to the closure of their domains, and hence φ_+, φ_- and φ are diffeomorphisms. □

For sake of completeness, we include a proof of Proposition 2.2. A proof of statement (ii) has recently appeared in [Sch14]. This proof is based on a technical lemma about return time functions of a certain class of flow, which we now introduce. Consider coordinates $(x, q, p) \in \mathbb{R}/\mathbb{Z} \times \mathbb{R}^2$ and a smooth tangent vector field X on $\mathbb{R}/\mathbb{Z} \times \mathbb{R}^2$ satisfying

$$X(x, 0, 0) = (1, 0, 0), \qquad \forall x \in \mathbb{R}/\mathbb{Z}. \tag{20}$$

If we denote by ψ_t the flow of X then

$$\psi_t(x, 0, 0) = (x + t, 0, 0), \qquad \forall x \in \mathbb{R}/\mathbb{Z},$$

and $P := \mathbb{R}/\mathbb{Z} \times 0$ is a 1-manifold invariant by the flow. We assume also that for every $x \in \mathbb{R}/\mathbb{Z}$ and $t \in \mathbb{R}$ the subspace $\{0\} \times \mathbb{R}^2 \subset \mathbb{R}^3$ is preserved by the differential of the flow, i.e.

$$D\psi_t(x, 0, 0)\big[\{0\} \times \mathbb{R}^2\big] = \{0\} \times \mathbb{R}^2, \qquad \forall x \in \mathbb{R}/\mathbb{Z},\ \forall t \in \mathbb{R}. \tag{21}$$

For each $\delta \in (0, \infty]$ consider the annuli

$$A_\delta^+ := \mathbb{R}/\mathbb{Z} \times [0, \delta), \qquad A_\delta^- := \mathbb{R}/\mathbb{Z} \times (-\delta, 0],$$

both equipped with the coordinates (x, y). To each point $(x, y) \in \mathrm{int}(A_\delta^+)$ one may try to associate the point $\varphi_+(x, y) \in \mathrm{int}(A_\delta^-)$ given by the formula

$$\varphi_+(x, y) = \psi_{\tau_+(x,y)}(x, y, 0) \tag{22}$$

where $\tau_+(x, y)$ is a tentative "first hitting time of A_δ^-", that is,

$$\tau_+(x, y) = \inf\big\{t > 0 \ \mid\ \psi_t(x, y, 0) \in \mathrm{int}(A_\infty^-) \times \{0\}\big\}. \tag{23}$$

Of course, in general τ_+ and φ_+ may not be well-defined, even for small δ. Our purpose below is to give a sufficient condition on the vector field X to guarantee that, if δ is small enough, τ_+ and φ_+ are well-defined smooth functions on $\mathrm{int}(A_\delta^+)$ which extend smoothly to A_δ^+. In the following definition and in the proof of the lemma below, we identify $\mathbb{R}^2$ with $\mathbb{C}$.

Definition 2.4. *Fix some $x \in \mathbb{R}/\mathbb{Z}$ and $v \in \mathbb{R}^2 \setminus \{0\}$. By (21) the image of $(0, v)$ by the differential of ψ_t at $(x, 0, 0)$ has the form*

$$D\psi_t(x,0,0)\big[(0,v)\big] = (0, \rho(t)e^{i\theta(t)}),$$

for suitable smooth functions $\rho > 0$ and θ, where ρ is unique and θ is unique up to the addition of an integer multiple of 2π. We say that the linearized flow along P has a positive twist if for every choice of $x \in \mathbb{R}/\mathbb{Z}$ and $v \in \mathbb{R}^2 \setminus \{0\}$ the function θ which is defined above satisfies $\theta'(t) > 0$ for all $t \in \mathbb{R}$.

Lemma 2.5. *If the linearized flow along P has a positive twist, then there exists $\delta_0 > 0$ such that τ_+ is a well-defined smooth function on $\mathrm{int}(A^+_{\delta_0})$ which extends smoothly as a positive function on $A^+_{\delta_0}$. Moreover, this extension is described by the formula*

$$\tau^+(x,0) = \inf\, \{t > 0 \ \mid \ D\psi_t(x,0,0)[\partial_y] \in \mathbb{R}^- \partial_y\}, \tag{24}$$

where $\partial_y := (0, 1, 0)$.

Proof. Write $w = y + iz$ and $Y = X_2 + iX_3$, where (X_1, X_2, X_3) are the components of the vector field X. Then

$$X(x,w) = \big(X_1(x,w), Y(x,w)\big).$$

By (20) we have $X_1(x, 0) = 1$ and $Y(x, 0) = 0$. Consider $W(x,w) \in \mathcal{L}_{\mathbb{R}}(\mathbb{C})$ defined by

$$W(x,w) = \int_0^1 D_2Y(x,sw)\,ds,$$

where D_2Y denotes derivative with respect to the second variable. Then

$$W(x,0) = D_2Y(x,0), \qquad Y(x,w) = W(x,w)w.$$

We shall now translate the assumption that the linearized flow along P has a positive twist into properties of $W(x, 0)$. Choose $v_0 \in \mathbb{C} \setminus 0$. Using (21) we find a smooth non-vanishing complex valued function v such that

$$D\psi_t(x,0)[(0,v_0)] = (0, v(t)).$$

From

$$\frac{d}{dt} D\psi_t = (DX \circ \psi_t) D\psi_t,$$

and from (21) we get the linear ODE

$$\dot v(t) = D_2Y(x+t,0)v(t) = W(x+t,0)v(t).$$

Writing $v(t) = r(t)e^{i\theta(t)}$ with smooth functions $r > 0$ and θ, we know that

$$\begin{aligned} \theta' &= \mathrm{Re}\left(\frac{\dot v}{iv}\right) = \mathrm{Re}\left(\frac{W(x+t,0)v}{iv}\,\frac{\overline{iv}}{\overline{iv}}\right) \\ &= \frac{\langle W(x+t,0)v, iv\rangle}{|v|^2} = \left\langle W(x+t,0)e^{i\theta}, ie^{i\theta}\right\rangle, \end{aligned} \tag{25}$$

where $\langle\cdot,\cdot\rangle$ denotes the Hermitian product on $\mathbb{C}$. Since x, t and $v(t)$ can take arbitrary values, we conclude from the above formula and the assumptions of the lemma that

$$\langle W(x,0)u, iu\rangle > 0, \qquad \forall u \in \mathbb{C}\setminus\{0\},\ \forall x \in \mathbb{R}/\mathbb{Z}. \tag{26}$$

Consider polar coordinates $(r,\theta) \in [0,+\infty)\times\mathbb{R}/2\pi\mathbb{Z}$ in the w-plane given by $w = y + iz = re^{i\theta}$. The map

$$(x,r,\theta) \mapsto X(x, re^{i\theta})$$

is smooth. Using the formulas

$$\partial_y = \frac{y}{r}\partial_r - \frac{z}{r^2}\partial_\theta, \qquad \partial_z = \frac{z}{r}\partial_r + \frac{y}{r^2}\partial_\theta,$$

we obtain that the vector field X pulls back by this change of coordinates to a smooth vector field

$$Z = (Z_1, Z_2, Z_3),$$

which is given by

$$\begin{cases} Z_1(x,r,\theta) = X_1(x, re^{i\theta}), \\ Z_2(x,r,\theta) = \cos\theta\ X_2(x, re^{i\theta}) + \sin\theta\ X_3(x, re^{i\theta}), \\ Z_3(x,r,\theta) = \dfrac{1}{r}\left(\cos\theta\ X_3(x, re^{i\theta}) - \sin\theta\ X_2(x, re^{i\theta})\right). \end{cases} \tag{27}$$

Indeed, the smoothness of Z_1 and Z_2 follows immediately from the above formulas, while that of Z_3 needs a little more care. Since X_2, X_3 vanish on $\mathbb{R}/\mathbb{Z}\times\{0\}$, we can find smooth functions $X_{2,2}, X_{2,3}, X_{3,2}, X_{3,3}$ such that

$$\begin{aligned} X_2(x, y+iz) &= yX_{2,2}(x, y+iz) + zX_{2,3}(x, y+iz), \\ X_3(x, y+iz) &= yX_{3,2}(x, y+iz) + zX_{3,3}(x, y+iz), \end{aligned}$$

where

$$\begin{aligned} X_{2,2}(x,0) &= D_2X_2(x,0,0), \qquad & X_{2,3}(x,0) &= D_3X_2(x,0,0), \\ X_{3,2}(x,0) &= D_2X_3(x,0,0), \qquad & X_{3,3}(x,0) &= D_3X_3(x,0,0), \end{aligned}$$

and

$$W(z,w) = \begin{bmatrix} X_{2,2}(x,w) & X_{2,3}(x,w) \\ X_{3,2}(x,w) & X_{3,3}(x,w) \end{bmatrix}.$$

Substituting $y = r\cos\theta$, $z = r\sin\theta$ we find

$$Z_3(x,r,\theta) = \left\langle W(x, re^{i\theta})e^{i\theta}, ie^{i\theta}\right\rangle. \tag{28}$$

Thus Z_3 is a smooth function of (x,r,θ) and

$$Z_3(x,0,\theta) > 0, \qquad \forall x \in \mathbb{R}/\mathbb{Z},\ \forall \theta \in \mathbb{R}/2\pi\mathbb{Z}, \tag{29}$$

thanks to (26).

From now on we lift the variable θ from $\mathbb{R}/2\pi\mathbb{Z}$ to the universal covering $\mathbb{R}$ and think of the vector field Z as a smooth vector field defined on $\mathbb{R}/\mathbb{Z}\times[0,+\infty)\times\mathbb{R}$, having components 2π-periodic in θ. Clearly this vector field is tangent to $\{r=0\}$.

Let ζ_t denote the flow of Z. After changing coordinates and lifting, we see that the conclusions of the lemma will follow if we check that

$$\tau_+(x,r) = \inf\{t > 0 \mid \theta \circ \zeta_t(x,r,0) = \pi\} \tag{30}$$

defines a smooth function of $(x,r) \in \mathbb{R}/\mathbb{Z} \times [0,\delta)$ when δ is small enough. By (29) we see that if δ_0 is fixed small enough then $\tau_+(x,r)$ is a well-defined, uniformly bounded and strictly positive function of $(x,r) \in \mathbb{R}/\mathbb{Z} \times [0,\delta_0)$. Here we used that Z is tangent to $\{r = 0\}$. Perhaps after shrinking δ_0, we may also assume that

$$Z_3(\zeta_t(x,r,0)) > 0, \qquad \forall (x,r) \in \mathbb{R}/\mathbb{Z} \times [0,\delta_0),\ \forall t \in [0,\tau_+(z,r)]. \tag{31}$$

Continuity and smoothness properties of τ_+ remain to be checked. This is achieved with the aid of the implicit function theorem. In fact, consider the smooth function

$$F : \mathbb{R} \times \mathbb{R}/\mathbb{Z} \times [0,+\infty) \to \mathbb{R}, \qquad F(\tau,x,r) := \theta \circ \zeta_\tau(x,r,0).$$

Since

$$D_1F(\tau,x,r) = d\theta\big[Z(\zeta_\tau(x,r,0))\big] = Z_3(\zeta_\tau(x,r,0)),$$

it follows from (31) and from the implicit function theorem that the equation

$$F(\tau_+,x,r) = \pi$$

determines $\tau_+ = \tau_+(x,r)$ as a smooth function of $(x,r) \in \mathbb{R}/\mathbb{Z} \times [0,\delta_0)$.

We now check formula (24) for $\tau_+(x,0)$. From the above equations one sees that $\theta(t) = \theta \circ \zeta_t(x,0,0)$ satisfies the differential equation

$$\theta'(t) = \left\langle D_2Y(x+t,0)e^{i\theta}, ie^{i\theta} \right\rangle,$$

with initial condition $\theta(0) = 0$. Thanks to (25), this is exactly the same initial value problem for the argument $\hat{\theta}(t)$ of the solution $v(t) = \rho(t)e^{i\hat{\theta}(t)}$ of the linearized flow starting at the base point $(x,0)$ applied to the vector ∂_y. □

In order to prove Proposition 2.2, it is enough to show that coordinates can be arranged in such a way that the geodesic flow near a simple closed geodesic γ meets the assumptions of Lemma 2.5 when the Gaussian curvature is positive along γ. We will assume for simplicity, and without loss of generality, that $L = 1$. We start by recalling basic facts from Riemannian geometry and fixing some notation.

Given $v \in TS^2$, let $\mathcal{V}_v \subset T_vTS^2$ be the vertical subspace, which is defined as $\mathcal{V}_v := \ker d\pi(v)$. The isomorphism

$$i_{\mathcal{V}_v} : T_{\pi(v)}S^2 \to \mathcal{V}_v$$

is defined as

$$i_{\mathcal{V}_v}(w) := \frac{d}{dt}(v+tw)\Big|_{t=0}, \qquad \forall w \in T_{\pi(v)}S^2.$$

The Levi-Civita connection of g determines a bundle map $K : TTS^2 \to TS^2$ satisfying $\nabla_Y X = K(dX \circ Y)$, where X, Y are vector fields on S^2 seen as maps $S^2 \to TS^2$.

The horizontal subspace $\mathcal{H}_v := \ker K|_{T_vTS^2}$ satisfies $T_vTS^2 = \mathcal{V}_v \oplus \mathcal{H}_v$. There is an isomorphism

$$i_{\mathcal{H}_v} : T_{\pi(v)}S^2 \to \mathcal{H}_v, \qquad i_{\mathcal{H}_v}(w) := \frac{d}{dt}V(t)\Big|_{t=0}, \qquad \forall w \in T_{\pi(v)}S^2,$$

where V is the parallel vector field along the geodesic $\beta(t)$ satisfying $\dot\beta(0) = w$ with initial condition $V(0) = v$, seen as a curve in TTS^2. The isomorphism $i_{\mathcal{H}_v}$ satisfies

$$d\pi(v)\big[i_{\mathcal{H}_v}(w)\big] = w, \qquad \forall w \in T_{\pi(v)}S^2. \tag{32}$$

For each $v \in T^1S^2$ we have

$$T_vT^1S^2 = \operatorname{span}\{i_{\mathcal{V}_v}(v^\perp), i_{\mathcal{H}_v}(v^\perp), i_{\mathcal{H}_v}(v)\}.$$

The Hilbert form λ_H on TS^2 is given by

$$\lambda_H(v)[\zeta] := g_{\pi(v)}\big(v, d\pi(v)[\zeta]\big), \qquad \forall \zeta \in T_vS^2, \tag{33}$$

and restricts to a contact form α on T^1S^2. The contact structure $\xi := \ker\alpha$ is trivial since

$$\xi_v = \operatorname{span}\{i_{\mathcal{V}_v}(v^\perp), i_{\mathcal{H}_v}(v^\perp)\}.$$

The Reeb vector field R_α of α coincides with $i_{\mathcal{H}_v}(v)$, and $\{i_{\mathcal{V}_v}(v^\perp), i_{\mathcal{H}_v}(v^\perp)\}$ forms a symplectic basis for $d\alpha|_{\xi_v}$, because

$$d\alpha(v)\big[i_{\mathcal{V}_v}(v^\perp), i_{\mathcal{H}_v}(v^\perp)\big] = 1.$$

If (x, y) are the standard coordinates on $\Sigma_\gamma^\pm$ given by

$$v = \cos y\ \dot\gamma(x) + \sin y\ \dot\gamma(x)^\perp,$$

then the tangent vectors ∂_x and ∂_y in $T_v\Sigma_\gamma^\pm$ are

$$\begin{aligned} \partial_x &= i_{\mathcal{H}_v}(\dot\gamma(x)) = \cos y\ i_{\mathcal{H}_v}(v) - \sin y\ i_{\mathcal{H}_v}(v^\perp), \\ \partial_y &= i_{\mathcal{V}_v}(v^\perp). \end{aligned} \tag{34}$$

Proof of Proposition 2.2. It is enough to prove statement (i) for the function τ_+. In fact, the case of τ_- follows by inverting the orientation of γ, and statement (ii) is then a direct consequence of the identity (19).

By (34) the vector field $R_\alpha = i_{\mathcal{H}_v}(v)$ is transverse to the interior of $\Sigma_\gamma^\pm$. The smooth vector field

$$i_{\mathcal{H}_v}(\dot\gamma^\perp) = \sin y\ i_{\mathcal{H}_v}(v) + \cos y\ i_{\mathcal{H}_v}(v^\perp)$$

along $\Sigma_\gamma^+ \cup \Sigma_\gamma^-$ is transverse to it near $\dot\gamma$. To obtain the desired coordinates near $\dot\gamma$ we proceed as follows: let $\bar g$ be the Riemannian metric on T^1S^2 defined by

$$\bar g_v(\zeta_1, \zeta_2) := \alpha(\zeta_1)\alpha(\zeta_2) + d\alpha\big(\pi_\xi(\zeta_1), J\pi_\xi(\zeta_2)\big),$$

where $J : \xi \to \xi$ is the $d\lambda$-compatible complex structure determined by

$$J\big(i_{\mathcal{V}_v}(v^\perp)\big) = i_{\mathcal{H}_v}(v^\perp),$$

$\pi_\xi : T^1S^2 \to \xi$ is the projection along R_α, and $\zeta_1, \zeta_2 \in T_vT^1S^2$ are arbitrary. Note that ξ is orthogonal to $\mathbb{R}R_\alpha$ with respect to $\bar{g}$ and $\bar{g}(i_{\mathcal{V}_v}(v^\perp), i_{\mathcal{H}_v}(v^\perp)) = 0$.

Denote by Exp the exponential map of $\bar{g}$. Then for all $\delta > 0$ sufficiently small, the map

$$\begin{aligned}
&\mathbb{R}/\mathbb{Z} \times (-\delta, \delta) \times (-\delta, \delta) \to \mathcal{U} \\
&(x, y, z) \mapsto \mathrm{Exp}_{v=\cos y \dot{\gamma}(x)+\sin y \dot{\gamma}^\perp(x)} \Big(z(\sin y \; i_{\mathcal{H}_v}(v) + \cos y \; i_{\mathcal{H}_v}(v^\perp)) \Big)
\end{aligned}$$

is a diffeomorphism, where $\mathcal{U} \subset T^1S^2$ is a small tubular neighborhood of $\dot{\gamma}$. In coordinates (x, y, z), we have

$$\begin{aligned}
\dot{\gamma} &\equiv \mathbb{R}/\mathbb{Z} \times \{(0,0)\} \\
\Sigma_\gamma^+ &\equiv \{z = 0, y \geq 0\} \\
\Sigma_\gamma^- &\equiv \{z = 0, y \leq 0\} \\
R_\alpha|_{\dot{\gamma}} &\equiv (1,0,0)|_{\mathbb{R}/\mathbb{Z}\times\{(0,0)\}} \\
\xi|_{\dot{\gamma}} &\equiv \{0\} \times \mathbb{R}^2|_{\mathbb{R}/\mathbb{Z}\times\{(0,0)\}} \\
i_{\mathcal{V}_{\dot{\gamma}}}(\dot{\gamma}^\perp) &\equiv \partial_y|_{\mathbb{R}/\mathbb{Z}\times\{(0,0)\}} \\
i_{\mathcal{H}_{\dot{\gamma}}}(\dot{\gamma}^\perp) &\equiv \partial_z|_{\mathbb{R}/\mathbb{Z}\times\{(0,0)\}}.
\end{aligned} \tag{35}$$

Denote by $X = (X_1, X_2, X_3)$ the Reeb vector field R_α in these coordinates and by ψ_t its flow. Then $X(x, 0, 0) = (1, 0, 0)$ and since ψ_t preserves the contact structure, we have

$$D\psi_t(x, 0, 0)\big[\{0\} \times \mathbb{R}^2\big] = \{0\} \times \mathbb{R}^2.$$

A linearized solution $\zeta(t) = a_1(t)\partial_y + a_2(t)\partial_z$ along $\psi_t(x, 0, 0) = (x + t, 0, 0)$ satisfies

$$\begin{pmatrix} a_1'(t) \\ a_2'(t) \end{pmatrix} = \begin{pmatrix} 0 & -K(t) \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1(t) \\ a_2(t) \end{pmatrix},$$

where $K(t)$ is the Gaussian curvature at $\gamma(x+t)$. Writing in complex polar coordinates $a_1(t) + ia_2(t) = \rho(t)e^{i\theta(t)}$, for smooth functions $\rho \geq 0$ and θ, we can easily check that

$$\theta'(t) = \cos^2\theta(t) + K(t)\sin^2\theta(t), \qquad \forall t \in \mathbb{R}.$$

Therefore, the positivity of the Gaussian curvature along γ implies the twist condition. We have finished checking that X meets all the assumptions of Lemma 2.5. Proposition 2.2 follows readily from an application of that lemma. □

2.2 The contact volume, the return time and the Riemannian area

As we have seen in the previous section, the Hilbert form λ_H defined in (33) induces by restriction a contact form α on T^1S^2. A further restriction produces the one-form λ on the Birkhoff annulus Σ_γ^+. By using the standard smooth coordinates $(x, y) \in \mathbb{R}/L\mathbb{Z} \times [0, \pi]$ on Σ_γ^+, we express a vector $v \in \Sigma_\gamma^+$ as

$$v = \cos y \; \dot{\gamma}(x) + \sin y \; \dot{\gamma}(x)^\perp, \tag{36}$$

and we find, using (33) and (34), together with (32),

$$\begin{aligned}
\lambda(v)[\partial_x] &= g_{\pi(v)}\big(v, d\pi(v)[\cos y\ i_{\mathcal{H}_v}(v) - \sin y\ i_{\mathcal{H}_v}(v^\perp)]\big)\\
&= g_{\pi(v)}(v, \cos y\ v - \sin y\ v^\perp) = \cos y,\\
\lambda(v)[\partial_y] &= g_{\pi(v)}\big(v, d\pi(v)[i_{\mathcal{V}_v}(v^\perp)]\big) = g_{\pi(v)}(v,0) = 0
\end{aligned}$$

Therefore, the expression of λ in the coordinates (x, y) is

$$\lambda = \cos y\, dx,$$

and its differential reads

$$d\lambda = \sin y\, dx \wedge dy.$$

Thus, the forms λ and $\omega = d\lambda$ are the ones considered in part 1 on the universal cover S of $\mathbb{R}/L\mathbb{Z} \times [0, \pi]$.

Since the geodesic flow ϕ_t preserves α for all t, we have for any v in $\mathrm{int}(\Sigma_\gamma^+)$ and ζ in $T_v\Sigma_\gamma^+$

$$\begin{aligned}
(\varphi^*\lambda)(v)[\zeta] &= \lambda(\varphi(v))\big[d\varphi(v)[\zeta]\big]\\
&= \lambda\big(\phi_{\tau(v)}(v)\big)\big[d\phi_{\tau(v)}(v)[\zeta] + d\tau(v)[\zeta]R_\alpha(\phi_{\tau(v)}(v))\big]\\
&= \lambda(v)[\zeta] + d\tau(v)[\zeta]
\end{aligned}$$

on $\mathrm{int}(\Sigma_\gamma^+)$, and hence on its closure Σ_γ^+ since all the objects here are smooth. Here, R_α is the Reeb vector field on the contact manifold (T^1S^2, α), which coincides with the generator of the geodesic flow. Therefore,

$$d\tau = \varphi^*\lambda - \lambda \qquad \text{on } \Sigma_\gamma^+.$$

Now let

$$\Psi : \mathrm{int}(\Sigma_\gamma^+) \times \mathbb{R} \to T^1S^2 \setminus \big(\dot\gamma(\mathbb{R}) \cup -(\dot\gamma(\mathbb{R})\big)$$

be defined as $\Psi(v, t) := \phi_t(v)$. Then

$$\begin{aligned}
\Psi^*\alpha(v,t)[(\zeta, s)] &= \alpha(\phi_t(v))\big[d\phi_t(v)[\zeta] + sR_\alpha(\phi_t(v))\big]\\
&= \alpha(v)[\zeta] + s = \lambda(v)[\zeta] + s,
\end{aligned}$$

that is,

$$\Psi^*\alpha = \lambda + dt.$$

Again, we used the preservation of α by ϕ_t. Since $\lambda \wedge d\lambda = 0$, being a three-form on a two-dimensional manifold, we deduce that

$$\Psi^*(\alpha \wedge d\alpha) = dt \wedge d\lambda.$$

Denoting by K the subset

$$K := \{(v, t) \in \mathrm{int}(\Sigma_\gamma^+) \times \mathbb{R} \mid v \in \mathrm{int}(\Sigma_\gamma^+),\ t \in [0, \tau(x)]\},$$

we can relate the contact volume $\mathrm{Vol}(T^1S^2, \alpha)$ with the function τ as follows

$$\begin{aligned}
\mathrm{Vol}(T^1S^2, \alpha) &= \iiint_{T^1S^2\setminus\left(\dot\gamma(\mathbb{R})\cup(-\dot\gamma(\mathbb{R}))\right)} \alpha \wedge d\alpha = \iiint_K \Psi^*(\alpha \wedge d\alpha)\\
&= \iiint_K dt \wedge d\lambda = \iint_{\Sigma_\gamma^+} \left(\int_0^{\tau(v)} dt\right) d\lambda(v) = \iint_{\Sigma_\gamma^+} \tau\, d\lambda.
\end{aligned}$$

Summarizing, we have proved the following:

PROPOSITION 2.6. *The restriction λ of the contact form α of T^1S^2 to Σ_γ^+ has the form*

$$\lambda = \cos y\, dx$$

in the standard coordinates $(x,y) \in \mathbb{R}/L\mathbb{Z} \times [0,\pi]$. The first return map $\varphi : \Sigma_\gamma^+ \to \Sigma_\gamma^+$ preserves $d\lambda$. Moreover, the first return time $\tau : \Sigma_\gamma^+ \to \mathbb{R}$ satisfies

$$d\tau = \varphi^*\lambda - \lambda \qquad \text{on } \Sigma_\gamma^+.$$

Finally

$$\mathrm{Vol}(T^1S^2, \alpha) = \iint_{\Sigma_\gamma^+} \tau \; d\lambda.$$

For completeness we state and prove below a well known fact.

PROPOSITION 2.7. *The contact volume of (T^1S^2, α) and the Riemannian area of (S^2, g) are related by the identity*

$$\mathrm{Vol}(T^1S^2, \alpha) = 2\pi\, \mathrm{Area}(S^2, g).$$

Proof. Take isothermal coordinates $(x, y) \in U \subset \mathbb{R}^2$ on an embedded closed disk $U' \subset S^2$. In these coordinates, the metric g takes the form

$$ds^2 = a(x,y)^2(dx^2 + dy^2),$$

for a smooth positive function a. Any unit tangent vector $v \in T^1U' \subset T^1S^2$ can be written as

$$v = \frac{\cos\theta}{a}\partial_x + \frac{\sin\theta}{a}\partial_y, \qquad \text{with} \qquad \theta \in \mathbb{R}/2\pi\mathbb{Z},$$

where $a = |\partial_x|_g = |\partial_y|_g$. Thus $(x, y, \theta) \in U \times \mathbb{R}/2\pi\mathbb{Z}$ can be taken as coordinates on T^1U', and the bundle projection becomes $\pi(x, y, q) = (x, y)$. With respect to these coordinates, the contact form

$$\alpha(v)[\zeta] = g_{\pi(v)}\big(v, d\pi(v)[\zeta]\big)$$

has the expression

$$\alpha = a(\cos\theta\, dx + \sin\theta\, dy).$$

Differentiation yields

$$d\alpha = da \wedge (\cos\theta\, dx + \sin\theta\, dy) + a(-\sin\theta\, d\theta \wedge dx + \cos\theta\, d\theta \wedge dy).$$

Hence

$$\begin{aligned} \alpha \wedge d\alpha &= a\, da \wedge (\cos\theta \sin\theta\, dx \wedge dy + \sin\theta\cos\theta\, dy \wedge dx) \\ &\quad + a^2(\cos^2\theta\, dx \wedge d\theta \wedge dy - \sin^2\theta\, dy \wedge d\theta \wedge dx) \\ &= a^2\, dx \wedge d\theta \wedge dy = -a^2\, dx \wedge dy \wedge d\theta. \end{aligned}$$

Therefore, the orientation of T^1U' which is induced by $\alpha \wedge d\alpha$ is opposite to the standard orientation of $U \times \mathbb{R}/2\pi\mathbb{Z}$, and we get

$$\begin{aligned} \mathrm{Vol}(T^1U', \alpha) &= \iiint_{T^1U'} \alpha \wedge d\alpha = \iiint_{U \times \mathbb{R}/2\pi\mathbb{Z}} a^2 dx \wedge dy \wedge d\theta \\ &= \iint_U a^2(x,y) \left(\int_0^{2\pi} d\theta \right) dxdy = 2\pi \iint_U a^2(x,y)\; dxdy \\ &= 2\pi \iint_U \sqrt{\det(g)}\; dxdy = 2\pi\, \mathrm{Area}(U', g). \end{aligned}$$

Taking two embedded disks $U', U'' \subset S^2$ with disjoint interiors and coinciding boundaries, we get

$$\begin{aligned}\mathrm{Vol}(T^1S^2, \alpha) &= \mathrm{Vol}(T^1U', \alpha) + \mathrm{Vol}(T^1U'', \alpha)\\ &= 2\pi(\mathrm{Area}(U', g) + \mathrm{Area}(U'', g))\\ &= 2\pi \, \mathrm{Area}(S^2, g).\end{aligned}$$

□

2.3 The flux and the Calabi invariant of the Birkhoff return map

By using the standard smooth coordinates (x, y) given by (36), we can identify the Birkhoff annulus Σ_γ^+ with $\mathbb{R}/L\mathbb{Z} \times [0, \pi]$. Its universal cover is the natural projection

$$p : S \to \Sigma_\gamma^+,$$

where S is the strip $\mathbb{R} \times [0, \pi]$. The first return map $\varphi : \Sigma_\gamma^+ \to \Sigma_\gamma^+$ preserves the two-form $\omega = d\lambda$ and maps each boundary component into itself. Therefore, φ can be lifted to a diffeomorphism in the group $\mathcal{D}_L(S, \omega)$ which is considered in part 1. The aim of this section is to prove the following result, which relates the objects of this part with those of part 1.

Theorem 2.8. *Assume that the metric g on S^2 is δ-pinched with $\delta > 1/4$. Let γ be a simple closed geodesic of length L on (S^2, g). Then the first return map $\varphi : \Sigma_\gamma^+ \to \Sigma_\gamma^+$ has a lift $\Phi : S \to S$ which belongs to $\mathcal{D}_L(S, \omega)$ and has the following properties:*

(i) *Φ has zero flux.*

(ii) *The first return time $\tau : \Sigma_\gamma^+ \to \mathbb{R}$ is related to the action $\sigma : S \to \mathbb{R}$ of Φ by the identity*

$$\tau \circ p = L + \sigma \qquad \text{on } S.$$

(iii) *The area of (S^2, g) is related to the Calabi invariant of Φ by the identity*

$$\pi \, \mathrm{Area}(S^2, g) = L^2 + L \, \mathrm{CAL}(\Phi).$$

The proof of this theorem requires an auxiliary lemma, which will play an important role also in the next section.

Lemma 2.9. *Assume that (S^2, g) is δ-pinched for some $\delta > 1/4$. Fix some v in $\Sigma_\gamma^\pm$ and denote by α the geodesic satisfying $\dot\alpha(0) = v$. Then the geodesic arc $\alpha|_{[0,\tau^\pm(v)]}$ is injective.*

Proof. We consider the case of Σ_γ^+, the case of Σ_γ^- being completely analogous. Up to the multiplication of g by a positive number, we may assume that $1 \leq K < 4$.

Let $x^* \in \mathbb{R}$ be such that $\alpha(0) = \gamma(x^*)$ and let $y^* \in [0, \pi]$ be the angle between $\dot\gamma(x^*)$ and $v = \dot\alpha(0)$. Consider the family of unit speed geodesics α_y with $\alpha_y(0) = \alpha(0) = \gamma(x^*)$ such that the angle from $\dot\gamma(x^*)$ to $v_y := \dot\alpha_y(0)$ is y, for $y \in [0, \pi]$. In particular, $\alpha_{y^*} = \alpha$ and $v_{y^*} = v$. By Proposition 2.2 (i),

$$\{\alpha_y|_{[0,\tau_+(v_y)]}\}_{y \in [0,\pi]}$$

is a smooth family of geodesic arcs, parametrized on a family of intervals whose length varies smoothly.

We claim that $\tau_+(v_0) < L$ and $\tau_+(v_\pi) < L$. In order to prove this, first notice that the length L of the closed geodesic γ satisfies

$$L \geq \frac{2\pi}{\sqrt{\max K}} > \frac{2\pi}{\sqrt{4}} = \pi, \tag{37}$$

thanks to the lower bound (16) on the injectivity radius and to the inequality $K < 4$. Moreover, by Proposition 2.2 (i) the number $\tau_+(v_0)$ is the first positive zero of the solution u of the Jacobi equation

$$u''(t) + K(\gamma(x^* + t))u(t) = 0, \qquad u(0) = 0, \qquad u'(0) = 1.$$

Writing the complex function $u' + iu$ in polar coordinates as $u' + iu = re^{i\theta}$, for smooth real functions $r > 0$ and θ satisfying $r(0) = 1$, $\theta(0) = 0$, a standard computation gives

$$\theta'(t) = \cos^2\theta(t) + K(\gamma(x^* + t))\sin^2\theta(t).$$

Since $K \geq 1$, we have $\theta' \geq 1$ and hence $\theta(L) \geq L > \pi$. This implies that $\tau_+(v_0) < L$. The case of $\tau_+(v_\pi)$ follows by applying the previous case to the geodesic $t \mapsto \gamma(-t)$.

Let Y_0 be the subset of $[0,\pi]$ consisting of those y for which $\alpha_y|_{[0,\tau_+(v_y)]}$ is injective. The set Y_0 is open in $[0,\pi]$, and by the above claim 0 and π belong to Y_0. Let Y_1 be the subset of $(0,\pi)$ consisting of those y for which $\alpha_y|_{[0,\tau_+(v_y)]}$ has an interior self-intersection: There exist $0 < s < t < \tau_+(v_y)$ such that $\alpha_y(s) = \alpha_y(t)$. Such an interior self-intersection must be transverse, so the fact that S^2 is two-dimensional implies that also Y_1 is open in $[0,\pi]$. It is enough to show that $Y_0 \cup Y_1 = [0,\pi]$: Indeed, if this is so, the fact that $[0,\pi]$ is connected implies that only one of the two open sets Y_0 and Y_1 can be non-empty, and we have already checked that Y_0 contains 0 and π. The conclusion is that $[0,\pi] = Y_0$, and in particular $\alpha = \alpha_{y^*}$ is injective.

If y belongs to the complement of $Y_0 \cup Y_1$ in $[0,\pi]$, then $y \in (0,\pi)$ and $\alpha_y|_{[0,\tau_+(v_y)]}$ has a self-intersection only at its endpoints: $\alpha|_{[0,\tau_+(v_y))}$ is injective and $\alpha_y(\tau_+(v_y)) = \alpha_y(0)$. Denote by $l > 0$ the length of the geodesic loop $\alpha_y|_{[0,\tau_+(v_y)]}$. Together with the closed curve γ, this geodesic loop forms a two-gon with perimeter equal to $L + l$. By Theorem A.12 and the inequality $K \geq 1$, its perimeter $L + l$ satisfies

$$L + l \leq \frac{2\pi}{\sqrt{\min K}} \leq 2\pi.$$

By using the bound (37) and the analogous bound $l > \pi$ for the geodesic loop $\alpha_y|_{[0,\tau_+(v_y)]}$, we obtain

$$L + l > 2\pi.$$

The above two estimates contradict each other, and this shows that the complement of $Y_0 \cup Y_1$ is empty, concluding the proof. □

Proof of Theorem 2.8. Given $v \in T^1S^2$, we denote by α_v the geodesic parametrised by arc length such that $\dot\alpha_v(0) = v$. Let $v \in \Sigma_\gamma^+$ with $\pi(v) = \gamma(x)$. Then we know from Lemma 2.9 that the geodesic arc $\alpha_v|_{[0,\tau_+(v)]}$ is injective. In particular, $\alpha_v(\tau_+(v))$ is distinct from $\alpha_v(0) = \gamma(x)$, so there exists a unique number

$$\rho_+(v) \in (0, L)$$

such that

$$\alpha_v(\tau_+(v)) = \gamma(x + \rho_+(v)).$$

By the continuity of the geodesic flow and of the function τ_+, the function

$$\rho_+ : \Sigma_\gamma^+ \to (0, L)$$

is continuous. The restriction of τ_+ to the boundary of Σ_γ^+ satisfies

$$\rho_+(\dot\gamma(x)) = \tau_+(\dot\gamma(x)) \qquad \text{and} \qquad \rho_+(-\dot\gamma(x)) = L - \tau_+(-\dot\gamma(x)), \qquad \forall x \in \mathbb{R}. \tag{38}$$

Similarly, there exists a unique continuous function

$$\rho_- : \Sigma_\gamma^- \to (0, L)$$

such that, if $v \in \Sigma_\gamma^-$ is based at $\gamma(x)$, we have

$$\alpha_v(\tau_-(v)) = \gamma(x + \rho_-(v)).$$

As before,

$$\rho_-(\dot\gamma(x)) = \tau_-(\dot\gamma(x)) \qquad \text{and} \qquad \rho_-(-\dot\gamma(x)) = L - \tau_-(-\dot\gamma(x)), \qquad \forall x \in \mathbb{R}. \tag{39}$$

Define the function

$$\rho : \Sigma_\gamma^+ \to (0, 2L)$$

by

$$\rho := \rho_+ + \rho_- \circ \varphi_+.$$

By construction, we have for every $v \in \Sigma_\gamma^+$ with $\pi(v) = \gamma(x)$,

$$\pi(\varphi(v)) = \gamma(x + \rho(v)), \tag{40}$$

and, by (38) and (39), together with (19),

$$\rho(\dot\gamma(x)) = \tau(\dot\gamma(x)) \qquad \text{and} \qquad \rho(-\dot\gamma(x)) = 2L - \tau(-\dot\gamma(x)), \qquad \forall x \in \mathbb{R}. \tag{41}$$

Using the standard coordinates $(x, y) \in \mathbb{R}/L\mathbb{Z} \times [0, \pi]$ on Σ_γ^+, we can see ρ and τ as functions on $\mathbb{R}/L\mathbb{Z} \times [0, \pi]$ or, equivalently, as functions on $\mathbb{R} \times [0, \pi]$ which are L-periodic in the first variable. Thanks to (40) we can fix a lift $\Phi = (X, Y) \in \mathcal{D}_L(S, \omega)$ of φ by requiring its first component to be given by

$$X(x, y) = x + \rho(x, y) - L. \tag{42}$$

By (41) we have

$$X(x, 0) - x = \tau(x, 0) - L, \qquad X(x, \pi) - x = L - \tau(x, \pi), \qquad \forall x \in \mathbb{R}. \tag{43}$$

By definition, the action $\sigma : S \to \mathbb{R}$ of Φ is uniquely determined by the conditions

$$d\sigma = \Phi^*\lambda - \lambda,$$

$$\sigma(x, 0) + \mathrm{FLUX}(\Phi) = \int_{\gamma_x} \lambda = X(x, 0) - x, \qquad \forall x \in \mathbb{R}.$$

where γ_x is a path in ∂S connecting $(x,0)$ to $\Phi(x,0) = (X(x,0),0)$. By the first identity in (43) we have

$$\sigma(x,0) + \mathrm{FLUX}(\Phi) = \tau(x,0) - L, \qquad \forall x \in \mathbb{R}.$$

By Proposition 2.6, also the $(L,0)$-periodic function $\tau : S \to \mathbb{R}$ satisfies $d\tau = \Phi^*\lambda - \lambda$, so the above identity implies that

$$\sigma(x,y) + \mathrm{FLUX}(\Phi) = \tau(x,y) - L, \qquad \forall (x,y) \in S. \tag{44}$$

By Proposition 1.7 and the second identity in (43) we have

$$\sigma(x,\pi) - \mathrm{FLUX}(\Phi) = \int_{\delta_x} \lambda = -X(x,\pi) + x = \tau(x,\pi) - L, \qquad \forall x \in \mathbb{R},$$

where δ_x is a path in ∂S connecting (x,π) to $\Phi(x,\pi) = (X(x,\pi),\pi)$. Together with (44) this implies that $\mathrm{FLUX}(\Phi) = 0$, thus proving statement (i). Statement (ii) now follows from (44).

By Propositions 2.7 and 2.6, we have

$$\begin{aligned}\pi\,\mathrm{Area}(S^2,g) &= \frac{1}{2}\,\mathrm{Vol}(T^1S^2,\alpha) = \frac{1}{2}\iint_{\mathbb{R}/L\mathbb{Z}\times[0,\pi]} \tau\, d\lambda = \frac{1}{2}\iint_{[0,L]\times[0,\pi]} (L+\sigma)\, d\lambda \\ &= L^2 + \frac{1}{2}\iint_{[0,L]\times[0,\pi]} \sigma\, d\lambda = L^2 + L\ \mathrm{CAL}(\Phi),\end{aligned}$$

and (iii) is proved. □

2.4 Proof of the monotonicity property

As we have seen, the first return map φ can be lifted to a diffeomorphism Φ in the class $\mathcal{D}_L(S,\omega)$. The aim of this section is to prove that, if the curvature is sufficiently pinched, then this lift is a monotone map, in the sense of Definition 1.8 (notice that the monotonicity does not depend on the choice of the lift).

Proposition 2.10. *If g is δ-pinched for some $\delta > (4+\sqrt{7})/8$, then any lift $\Phi : S \to S$ of the first return map $\varphi : \Sigma_\gamma^+ \to \Sigma_\gamma^+$ is monotone.*

Proof. We may assume that the values of the curvature lie in the interval $[\delta,1]$, where $\delta > (4+\sqrt{7})/8$.

Fix some $x^* \in \mathbb{R}$. In order to simplify the notation in the next computations, we set for every $y \in [0,\pi]$

$$l_y := \tau(x^*,y), \qquad t_y := X(x^*,y), \qquad \tilde{y}(y) := Y(x^*,y),$$

where τ is seen as a $(L,0)$-periodic function on S and X and Y are the components of the fixed lift $\Phi = (X,Y)$ of φ. Our aim is to show that the derivative of the function $\tilde{y}$ is positive on $[0,\pi]$.

Consider the 1-parameter geodesic variation

$$\alpha_y(t) := \exp_{\gamma(x^*)}[t(\cos y\ \dot{\gamma}(x^*) + \sin y\ \dot{\gamma}(x^*)^\perp)],$$

where $y \in [0, \pi]$. For each $y \in (0, \pi)$, l_y is the second time $\alpha_y(t)$ hits $\gamma(\mathbb{R})$ or, equivalently, the first time $\dot\alpha_y(t)$ hits Σ^+_γ. Moreover, $\alpha_0(t) = \gamma(x^* + t)$, and l_0 is the time to the second conjugate point to $\alpha_0(0)$ along α_0; analogously, $\alpha_\pi(t) = \gamma(x^* - t)$, and l_π is the time to the second conjugate point to $\alpha_\pi(0)$ along α_π. By construction

$$\alpha_y(l_y) = \gamma(t_y),$$

and

$$\begin{aligned} \dot\alpha_y(l_y) &= \cos\tilde y\ \dot\gamma(t_y) + \sin\tilde y\ \dot\gamma(t_y)^\perp, \\ \dot\alpha_y(l_y)^\perp &= -\sin\tilde y\ \dot\gamma(t_y) + \cos\tilde y\ \dot\gamma(t_y)^\perp, \end{aligned} \tag{45}$$

for every $y \in [0, \pi]$, where the function $\tilde y$ is evaluated at y. Since γ is a geodesic,

$$\frac{D}{dy}\dot\gamma \circ t_y = \frac{D}{dt}\dot\gamma(t_y)\frac{\partial t_y}{\partial y} = 0,$$

and since the vector field $\dot\gamma^\perp$ along γ is parallelly transported,

$$\frac{D}{dy}\dot\gamma^\perp \circ t_y = \frac{D}{dt}\dot\gamma(t_y)^\perp\frac{\partial t_y}{\partial y} = 0.$$

Notice that $V(y) := \dot\alpha_y(l_y)$ is a vector field along the smooth curve $y \mapsto \gamma(t_y)$. Using that γ is a geodesic we obtain from (45)

$$\begin{aligned} \frac{DV}{dy}(y) &= -\tilde y' \sin\tilde y\ \dot\gamma(t_y) + \cos\tilde y\ \frac{D}{dy}\dot\gamma \circ t_y + \tilde y' \cos\tilde y\ \dot\gamma(t_y)^\perp + \sin\tilde y\ \frac{D}{dy}\dot\gamma^\perp \circ t_y \\ &= -\tilde y' \sin\tilde y\ \dot\gamma(t_y) + \tilde y' \cos\tilde y\ \dot\gamma(t_y)^\perp \\ &= \tilde y'(y)\ \dot\alpha_y(l_y)^\perp. \end{aligned} \tag{46}$$

The geodesic variation $\{\alpha_y\}$ at $y = y^*$ corresponds to the Jacobi field J along α_{y^*} given by

$$J(t) := \left.\frac{\partial}{\partial y}\right|_{y=y^*} \alpha_y(t). \tag{47}$$

From the initial conditions $J(0) = 0$ and

$$\frac{DJ}{dt}(0) = \left.\frac{D}{dy}\right|_{y=y^*} \dot\alpha_y(0) = \left.\frac{d}{dy}\right|_{y=y^*} \dot\alpha_y(0) = \dot\alpha_{y^*}(0)^\perp,$$

we find a smooth real function u such that

$$J(t) = u(t)\dot\alpha_{y^*}(t)^\perp, \qquad \frac{DJ}{dt}(t) = u'(t)\dot\alpha_{y^*}(t)^\perp, \qquad \forall t \in \mathbb{R},$$

and

$$u(0) = 0, \qquad u'(0) = 1. \tag{48}$$

Moreover

$$\left.\frac{D}{dy}\right|_{y=y^*} \dot\alpha_y(t) = \frac{D}{dt}J(t) = u'(t)\dot\alpha_{y^*}(t)^\perp, \qquad \forall t \in \mathbb{R}. \tag{49}$$

Recall that the covariant derivative of a vector field v along a curve δ on S^2 is the full derivative of the corresponding curve (δ, v) on TS^2 projected back to TS^2 by the connection operator $K : TTS^2 \to TS^2$. More precisely, K projects this full derivative $(\delta, v)'$ onto the vertical subspace $\mathcal{V}_{(\delta,v)} \subset T_{(\delta,v)}TS^2$ along the horizontal subspace $\mathcal{H}_{(\delta,v)} \subset T_{(\delta,v)}TS^2$, and then brings it to $T_\delta S^2$ via the inverse of the isomorphism $i_{\mathcal{V}_v}$, see the discussion after the proof of Lemma 2.5. In (46) we find the covariant derivative of the vector field $y \mapsto \dot\alpha_y(l_y)$ along the curve $y \mapsto \alpha_y(l_y)$. In (49) we see the covariant derivative of the vector field $y \mapsto \dot\alpha_y(t)$ along the curve $y \mapsto \alpha_y(t)$ for fixed t. Since α_y is a geodesic for all y, by using the above description of the covariant derivative we get from (46) and (49)

$$\begin{aligned}\tilde y'(y^*)\dot\alpha_{y^*}(l_{y^*})^\perp &= \frac{DV}{dy}(y^*) = \left.\frac{D}{dy}\right|_{y=y^*} \dot\alpha_y(l_{y^*}) + l_y'(y^*) \left.\frac{D}{dt}\right|_{t=l_{y^*}} \dot\alpha_{y^*}(t)\\ &= \left.\frac{D}{dy}\right|_{y=y^*} \dot\alpha_y(l_{y^*}) = u'(l_{y^*})\dot\alpha_{y^*}(l_{y^*})^\perp,\end{aligned}$$

for every $y^* \in [0,\pi]$, from which we derive the important identity

$$\tilde y'(y^*) = u'(l_{y^*}), \qquad \forall y^* \in [0,\pi]. \tag{50}$$

Write

$$l_{y^*} = l + l'$$

where $l > 0$ is the first time $\alpha_{y^*}(t)$ hits γ, that is,

$$l = \tau_+(\dot\alpha_{y^*}(0)), \qquad l' = \tau_-\big(\varphi_+(\dot\alpha_{y^*}(0))\big).$$

By Lemma 2.9, $\alpha_{y^*}|_{[0,l]}$ is injective and, in particular, its end-points are distinct points of γ, dividing it into two segments γ_1, γ_2 with lengths $l_1, l_2 > 0$, respectively, and $l_1 + l_2 = L$. Therefore, $\alpha_{y^*}|_{[0,l]}$ and γ_1 determine a geodesic two-gon. The same holds with $\alpha_{y^*}|_{[0,l]}$ and γ_2. It follows from Theorem A.12 that

$$l_1 + l \le \frac{2\pi}{\sqrt\delta} \qquad \text{and} \qquad l_2 + l \le \frac{2\pi}{\sqrt\delta}.$$

Theorem A.12 also implies that $L \le 2\pi/\sqrt\delta$. From Klingenberg's lower bound (16) on the injectivity radius of g, we must have $l_1 + l \ge 2\pi$, $l_2 + l \ge 2\pi$, and $L \ge 2\pi$. Putting these inequalities together, we obtain

$$2\pi \le l_i + l \le \frac{2\pi}{\sqrt\delta}, \qquad i = 1,2, \tag{51}$$

$$2\pi \le L = l_1 + l_2 \le \frac{2\pi}{\sqrt\delta}. \tag{52}$$

By adding the inequalities (51), we obtain

$$4\pi \le 2l + L \le \frac{4\pi}{\sqrt\delta}. \tag{53}$$

Together with (52), the above inequality implies

$$2\pi - \frac{\pi}{\sqrt\delta} \le l \le \frac{2\pi}{\sqrt\delta} - \pi.$$

Arguing analogously with the geodesic arc $\alpha_{y^*}|_{[l,l_{y^*}=l+l']}$, we obtain the similar estimate

$$2\pi - \frac{\pi}{\sqrt{\delta}} \leq l' \leq \frac{2\pi}{\sqrt{\delta}} - \pi,$$

concluding that the length l_{y^*} of α_{y^*} satisfies

$$4\pi - \frac{2\pi}{\sqrt{\delta}} \leq l_{y^*} = l + l' \leq \frac{4\pi}{\sqrt{\delta}} - 2\pi. \tag{54}$$

The Jacobi equation for the vector field J along α_{y^*} which is defined in (47) can be written in terms of the scalar function u as

$$u''(t) + K(\alpha_{y^*}(t))u(t) = 0.$$

Writing

$$u(t)' + iu(t) = re^{i\theta}$$

for smooth real functions $r > 0$ and θ, we get

$$\theta' = \cos^2\theta + K(\alpha_{y^*})\sin^2\theta. \tag{55}$$

The initial conditions (48) imply that $r(0) = 1$ and $\theta(0) = 0$. From (55) we have $\delta \leq \theta' \leq 1$. Hence, from the estimate for l_{y^*} given in (54), we find

$$\delta\left(4\pi - \frac{2\pi}{\sqrt{\delta}}\right) \leq \theta(l_{y^*}) \leq \frac{4\pi}{\sqrt{\delta}} - 2\pi. \tag{56}$$

From $\delta > (4+\sqrt{7})/8$ we get

$$\delta\left(4\pi - \frac{2\pi}{\sqrt{\delta}}\right) > \frac{3\pi}{2},$$

and since a fortiori $\delta > 64/81$, we have also

$$\frac{4\pi}{\sqrt{\delta}} - 2\pi < \frac{5\pi}{2}.$$

Therefore, (56) implies that $\cos\theta(l_{y^*})$ is positive. By the identity (50), we conclude that

$$\tilde{y}'(y^*) = u'(l_{y^*}) = r(l_{y^*})\cos\theta(l_{y^*}) > 0,$$

as we wished to prove. □

2.5 Proof of the main theorem

Two more lemmata are needed for the proof of the main theorem of the introduction.

Lemma 2.11. *Assume that the metric g on S^2 is δ-pinched for some $\delta > 1/4$. Then any closed geodesic γ of minimal length on (S^2, g) is a simple curve.*

Proof. If a closed geodesic γ of minimal length is not simple, then it contains at least two distinct geodesic loops. By the lower bound (16) on the injectivity radius, each of these two geodesic loops has length at least

$$\frac{2\pi}{\sqrt{\max K}},$$

and we deduce that

$$L \geq \frac{4\pi}{\sqrt{\max K}}. \tag{57}$$

A celebrated theorem due to Lusternik and Schnirelmann implies the existence of simple closed geodesics on any Riemannian S^2. By Theorem A.12 any simple closed geodesic has length at most

$$\frac{2\pi}{\sqrt{\min K}}.$$

By the pinching assumption,

$$\frac{2\pi}{\sqrt{\min K}} \leq \frac{2\pi}{\sqrt{\delta \max K}} < \frac{4\pi}{\sqrt{\max K}},$$

so by (57) any simple closed geodesic is shorter than L. This contradicts the fact that L is the minimal length of a closed geodesic and proves that γ must be simple. □

Assume the metric g on S^2 to be δ-pinched for some $\delta > 1/4$. Let γ be a closed geodesic on (S^2, g) of minimal length, which by the above lemma is a simple curve. We denote by L its length, so that

$$\ell_{\min}(S^2, g) = L.$$

Let $\varphi : \Sigma_\gamma^+ \to \Sigma_\gamma^+$ be the associated Birkhoff first return map and let $\Phi \in \mathcal{D}_L(S, \omega)$ be the lift of φ with zero flux whose existence is guaranteed by Theorem 2.8. Here is a first consequence of Theorem 2.8:

Lemma 2.12. *Assume that the metric g on S^2 is δ-pinched for some $\delta > 1/4$. Then g is Zoll if and only if $\Phi = \mathrm{id}$.*

Proof. Assume that $\Phi = \mathrm{id}$. Then the action σ of Φ is identically zero, so by Theorem 2.8 (ii) the first return time function τ is identically equal to L. Therefore, all the vectors in the interior of Σ_γ^+ are initial velocities of closed geodesics of length L. Since also the vectors in the boundary of Σ_γ^+ are by construction initial velocities of closed geodesics of length L, we deduce that all the geodesics on (S^2, g) are closed and have length L.

Conversely assume that (S^2, g) is Zoll. Since γ has length L, all the geodesics on (S^2, g) are closed and have length L. Then every v in $\mathrm{int}(\Sigma_\gamma^+)$ is a periodic point of φ, i.e. there is a minimal natural number $k(v)$ such that $\varphi^{k(v)}(v) = v$, and the identity

$$\sum_{j=0}^{k(v)-1} \tau(\varphi^j(v)) = L$$

holds on $\mathrm{int}(\Sigma_\gamma^+)$. Thanks to the continuity of τ and φ and to the positivity of τ, the above identity forces the function k to be constant, $k \equiv k_0 \in \mathbb{N}$. By continuity, the above identity holds also on the boundary of Σ_γ^+, and we have in particular

$$\sum_{j=0}^{k_0-1} \tau(\varphi^j(\dot\gamma(t))) = L \qquad \forall t \in \mathbb{R}/L\mathbb{Z}.$$

By the above identity, there exists $t_0 \in \mathbb{R}/L\mathbb{Z}$ such that

$$\tau(\dot\gamma(t_0)) \leq \frac{L}{k_0},$$

that is, the time to the second conjugate point to $\gamma(t_0)$ along γ is at most L/k_0. Since this time is at least twice the injectivity radius of (S^2, g), we obtain from (16)

$$\frac{L}{k_0} \geq \tau(\dot\gamma(t_0)) \geq 2\,\mathrm{inj}(g) \geq \frac{2\pi}{\sqrt{\max K}}. \tag{58}$$

On the other hand, by Theorem A.12 and by the pinching assumption, the length L of the simple closed geodesic γ satisfies

$$L \leq \frac{2\pi}{\sqrt{\min K}} \leq \frac{2\pi}{\sqrt{\delta \max K}} < \frac{4\pi}{\sqrt{\max K}}. \tag{59}$$

Inequalities (58) and (59) imply that the positive integer k_0 is less than 2, hence $k_0 = 1$ and $\varphi = \mathrm{id}$. Then Φ is a translation by an integer multiple of L and, having zero flux, it must be the identity. □

We can finally prove the theorem which is stated in the introduction: If g is δ-pinched with $\delta > (4+\sqrt{7})/8$, then the bound

$$L^2 = \ell_{\min}(S^2, g)^2 \leq \pi\,\mathrm{Area}(S^2, g) \tag{60}$$

holds, with the equality holding if and only if (S^2, g) is Zoll.

Proof. If (S^2, g) is Zoll, then by the above lemma $\Phi = \mathrm{id}$, so $\mathrm{CAL}(\Phi) = 0$, and Theorem 2.8 (iii) implies that

$$\pi\,\mathrm{Area}(S^2, g) = L^2.$$

This shows that if g is Zoll, then the equality holds in (60).

There remains to show that if (S^2, g) is not Zoll, then the strict inequality holds in (60). Assume by contradiction that

$$L^2 \geq \pi\,\mathrm{Area}(S^2, g).$$

Then by Theorem 2.8 (iii) we have

$$L\ \mathrm{CAL}(\Phi) = \pi\,\mathrm{Area}(S^2, g) - L^2 \leq 0,$$

and $\mathrm{CAL}(\Phi)$ is non-positive. Since (S^2, g) is not Zoll, by Lemma 2.12 the map Φ is not the identity. Therefore, Φ satisfies the hypothesis of Theorem 1.12, which guarantees the existence of a fixed point $(x, y) \in \mathrm{int}(S)$ of Φ with action $\sigma(x, y) < 0$. The geodesic which is determined by the corresponding vector in Σ_γ^+ is closed and, by Theorem 2.8, has length

$$\tau(x,y) = L + \sigma(x,y) < L.$$

This contradicts the fact that L is the minimal length of a closed geodesic. This contradiction implies that when (S^2, g) is not Zoll, then the strict inequality

$$L^2 < \pi \operatorname{Area}(S^2, g)$$

holds. The proof is complete. $\square$

A Toponogov's theorem and its consequences

This appendix is devoted to explaining how to estimate lengths of convex geodesic polygons using a relative version of Toponogov's theorem.

A.1 Geodesic polygons and their properties

For this discussion we fix a Riemannian metric g on S^2. The following definitions are taken from [CE75].

Definition A.1. *Let $X \subset S^2$.*

i) X is strongly convex if for every pair of points p, q in X there is a unique minimal geodesic from p to q, and this geodesic is contained in X.

ii) X is convex if for every p in $\overline{X}$ there exists $r > 0$ such that $B_r(p) \cap X$ is strongly convex.

When $p \in S^2$ and $u, v \in T_pS^2$ are non-colinear vectors, consider the sets

$$\Delta(u,v) = \{su + tv \mid s, t \geq 0\} \tag{61}$$

$$\Delta_r(u,v) = \{w \in \Delta(u,v) \mid |w| < r\}. \tag{62}$$

When $u \in T_pS^2 \setminus \{0\}$ consider also

$$H(u) = \{v \in T_pS^2 \mid g(v,u) \geq 0\} \tag{63}$$

$$H_r(u) = \{w \in H(u) \mid |w| < r\}. \tag{64}$$

A corner of a unit speed broken geodesic $\gamma : \mathbb{R}/L\mathbb{Z} \to S^2$ is a point $\gamma(t)$ such that $\gamma'_+(t) \notin \mathbb{R}^+\gamma'_-(t)$, where $\gamma'_\pm$ denote one-sided derivatives.

Definition A.2. *$D \subset S^2$ is said to be a geodesic polygon if it is the closure of an open disk bounded by a simple closed unit speed broken geodesic $\gamma : \mathbb{R}/L\mathbb{Z} \to S^2$. We call D convex if for every corner $p = \gamma(t)$ of γ we find $0 < r < \operatorname{inj}_p$ small enough such that $D \cap B_r(p) = \exp_p(\Delta_r(-\gamma'_-(t), \gamma'_+(t)))$. The corners of γ are called vertices of D, and a side of D is a smooth geodesic arc contained in ∂D connecting two adjacent vertices.*

Jordan's theorem ensures that every simple closed unit speed broken geodesic is the boundary of exactly two geodesic polygons. At each boundary point which is not a vertex the inner normals to the two polygons are well-defined and opposite to each other.

It is well-known that $B_r(p)$ is strongly convex when r is small enough. By the following lemma the same property holds for $\exp_p(\Delta_r(u,v))$ and $\exp_p(H_r(u))$.

LEMMA A.3. *Choose p in S^2 and let $0 < r < \mathrm{inj}(g)$. If $B_r(p)$ is strongly convex then $\exp_p(\Delta_r(u,v))$ and $\exp_p(H_r(u))$ are strongly convex for all pairs $u, v \in T_pS^2$ of non-colinear vectors.*

Proof. There is no loss of generality to assume that u, v are unit vectors. We argue indirectly. Assume that $y, z \in \exp_p(\Delta_r(u,v))$ are points for which the minimal geodesic γ from y to z (with unit speed) is not contained in $\exp_p(\Delta_r(u,v))$. Let γ_u and γ_v be the geodesic segments $\exp_p(\tau u)$, $\exp_p(\tau v)$ respectively, $\tau \in (-r, r)$. Note that γ is contained in $B_r(p)$ and, consequently, γ must intersect one of the geodesic segments γ_u or γ_v in two points $a \neq b$. Thus we have found two geodesic segments from a to b which are length minimizers in S^2 (one is contained in γ and the other is contained in γ_u or γ_v). This contradicts the fact that $B_r(p)$ is strongly convex. The argument to prove strong convexity of $\exp_p(H_r(u))$ is analogous. □

As an immediate consequence we have the following:

COROLLARY A.4. *A convex geodesic polygon $D \subset S^2$ is convex.*

Let $d(p,q)$ denote the g-distance between points $p, q \in S^2$.

LEMMA A.5. *Let D be a convex geodesic polygon. Then there exists a positive number $\epsilon_1 < \mathrm{inj}(g)$ such that if p, q are in D and satisfy $d(p,q) < \epsilon_1$, then the (unique) minimal geodesic from p to q lies in D.*

Proof. If not we find $p_n, q_n \in D$ such that $d(p_n, q_n) \to 0$ and the minimal geodesic γ_n in S^2 from p_n to q_n intersects $S^2 \setminus D$. Thus, up to selection of a subequence, we may assume that $p_n, q_n \to x \in \partial D$. If x is not a corner of ∂D then we consider the unit vector $n \in T_xS^2$ pointing inside D normal to the boundary and note that, for some $r > 0$ small, $D \cap B_r(x) = \exp_x(H_r(n))$ is strongly convex. Here we used Lemma A.3. This is in contradiction to the fact that $p_n, q_n \in D \cap B_r(x)$ when n is large. Similarly, if x is a corner of ∂D then, in view of the same lemma, we find unit vectors $u, v \in T_xS^2$ and r very small such that $D \cap B_r(x) = \exp_x(\Delta_r(u,v))$ is strongly convex. This again provides a contradiction. □

The next lemma shows that a convex geodesic polygon is 'convex in the large'.

LEMMA A.6. *Let D be a convex geodesic polygon. Then for every p and q in D there is a smooth geodesic arc γ from p to q satisfying*

i) $\gamma \subset D$.

ii) γ *minimizes length among all piecewise smooth curves inside D from p to q.*

Proof. The argument follows a standard scheme. Consider a partition P of $[0,1]$ given by $t_0 = 0 < t_1 < \cdots < t_{N-1} < t_N = 1$, with norm

$$\|P\| = \max_i \{t_{i+1} - t_i\}.$$

Let Λ_P be the set of continuous curves $\alpha : [0,1] \to S^2$ such that each $\alpha|_{[t_i,t_{i+1}]}$ is smooth, $\alpha(0) = p$, $\alpha(1) = q$. On Λ_P we have the usual length and energy functionals

$$L[\alpha] = \int_0^1 |\alpha'(t)| dt, \qquad E[\alpha] = \frac{1}{2}\int_0^1 |\alpha'(t)|^2 dt. \tag{65}$$

Set

$$B_P = \{\alpha \in \Lambda_P \mid \alpha|_{[t_i,t_{i+1}]} \text{ is a geodesic } \forall i\},$$
$$\Lambda_P(D) = \{\alpha \in \Lambda_P \mid \alpha([0,1]) \subset D\}, \quad B_P(D) = B_P \cap \Lambda_P(D).$$

As usual, we use superscritps $\leq a$ to indicate sets of paths satisfying $E \leq a$.

If α is in $\Lambda_P^{\leq a}$ and $\sqrt{\|P\|} \leq \epsilon_1/\sqrt{2a}$, then $d(\alpha(t_i), \alpha(t_{i+1})) \leq \epsilon_1 \ \forall i$, where $\epsilon_1 > 0$ is the number given by Lemma A.5. Thus, for every $\alpha \in \Lambda_P^{\leq a}(D)$ we find $\gamma \in B_P(D)$ such that each $\gamma|_{[t_i,t_{i+1}]}$ is a constant-speed reparametrization of the unique minimal geodesic arc from $\alpha(t_i)$ to $\alpha(t_{i+1})$. Here we have used Lemma A.5 to conclude that $\gamma([0,1]) \subset D$. Clearly $L[\gamma] \leq L[\alpha]$, so minimizing L on $\Lambda_P^{\leq a}(D)$ amounts to minimizing L on $B_P^{\leq a}(D)$. Now pick $a > 0$ and a partition P such that $\Lambda_P^{\leq a}(D) \neq \emptyset$ and $\sqrt{\|P\|} \leq \epsilon_1/\sqrt{2a}$. By the above argument, $B_P^{\leq a}(D) \neq \emptyset$ and, as usual, the map $\gamma \mapsto (\gamma(t_1), \dots, \gamma(t_{N-1}))$ is a bijection between $B_P^{\leq a}(D)$ and a certain closed subset of D^{N-1}. The topology which $B_P^{\leq a}(D)$ inherits from this identification makes L continuous. Thus, by compactness, we find $\gamma_* \in B_P^{\leq a}(D)$ which is an absolute minimizer of L over $\Lambda_P^{\leq a}(D)$.

We claim that γ_* is smooth, i.e., it has no corners. In fact, arguing indirectly, suppose it has a corner, which either lies on $\mathrm{int}(D)$ or on ∂D. In both cases we can use the auxiliary claim below to find a variation of γ_* through paths in $B_P^{\leq a}(D)$ that decreases length; the convexity of D is strongly used. This is a contradiction, and the smoothness of γ_* is established.

Auxiliary Claim. Consider $a < x < b$ and a broken geodesic $\beta : [a,b] \to S^2$, which is smooth and non-constant on $[a,x]$ and on $[x,b]$, satisfying $\beta'_+(x) \notin \mathbb{R}^+\beta'_-(x)$. Let $\alpha : (-\epsilon,\epsilon)\times[a,b] \to S^2$ be a piecewise smooth variation with fixed endpoints of β ($\alpha(0,\cdot) = \beta$) by broken geodesics such that α is smooth on $(-\epsilon,\epsilon)\times[a,x]$ and on $(-\epsilon,\epsilon)\times[x,b]$. If $D_1\alpha(0,x)$ is a non-zero vector in $\Delta(-\beta'_-(x), \beta'_+(x))$, then $\frac{d}{ds}|_{s=0} L[\alpha(s,\cdot)] < 0$. In fact, the first variation formula gives us

$$\frac{d}{ds}\int_a^b |D_2\alpha(s,t)|\,dt\Big|_{s=0} = g_{\beta(x)}\left(D_1\alpha(0,x), \frac{\beta'_-(x)}{\|\beta'_-(x)\|} - \frac{\beta'_+(x)}{\|\beta'_+(x)\|}\right) < 0$$

as desired. □

It remains to be shown that γ_* is an absolute length minimizer among all piecewise smooth curves in D joining p to q. Let α be such a curve, which must belong to $\Lambda_Q^{\leq b}(D)$ for some positive number b and some partition Q. Up to increasing b and refining Q, we may assume that $b \geq a$, $Q \supset P$, and $\sqrt{\|Q\|} \leq \epsilon_1/\sqrt{2b}$. By the previously explained arguments we can find a smooth geodesic $\tilde\gamma$ from p to q in D which is a global minimizer of L over $\Lambda_Q^{\leq b}(D)$. Since $\Lambda_P^{\leq a}(D)$ is contained in $\Lambda_Q^{\leq b}(D)$, we must have $L[\tilde\gamma] \leq L[\gamma_*]$. Noting that $\gamma_*, \tilde\gamma$ are smooth geodesics, we compute $E[\tilde\gamma] = \frac{1}{2}L[\tilde\gamma]^2 \leq \frac{1}{2}L[\gamma_*]^2 = E[\gamma_*]$ and conclude that $\tilde\gamma \in \Lambda_P^{\leq a}(D)$. Thus $L[\gamma_*] = L[\tilde\gamma] \leq L[\alpha]$ as desired. □

Lemma A.7. *If D is a convex geodesic polygon in (S^2, g), p and q are distinct points of ∂D, and d is the distance from p to q relative to D then the following holds: a unit speed geodesic $\gamma : [0,d] \to D$ from p to q minimal relative to D (which exists and is smooth in view of Lemma A.6) is injective, and satisfies either $\gamma((0,d)) \subset \mathrm{int}(D)$ or $\gamma([0,d]) \subset \partial D$. In the former case γ divides D into two convex geodesic polygons*

D', D'' satisfying $D = D' \cup D''$, $\gamma = D' \cap D''$; moreover, a geodesic between two points of D' (D'') which is minimal relative to D is contained in D' (D''). In the latter case there are no vertices of D in $\gamma((0,d))$.

Proof. If there exists t in $(0,d)$ such that $\gamma(t)$ belongs to ∂D, then either $\gamma(t)$ is a vertex or not. But it can not be a vertex since in this case $\gamma'(t)$ would be colinear to one of the tangent vectors of ∂D at $\gamma(t)$, allowing us to find t' close to t such that $\gamma(t')$ is not in D. Not being a vertex, $\gamma(t)$ is a point of tangency with ∂D. By uniqueness of solutions of ODEs, we must have $\gamma([0,d]) \subset \partial D$, hence D has no vertices in $\gamma((0,d))$. By minimality γ has to be injective. If δ', δ'' are the two distinct arcs on ∂D from p to q and $\gamma((0,d)) \cap \partial D = \emptyset$ then $\delta' \cup \gamma$ and $\delta'' \cup \gamma$ bound disks $D', D'' \subset D$ which are clearly geodesic convex polygons. Let $\alpha \subset D$ be a (smooth) geodesic arc connecting distinct points of D' minimal relative to D. If $\alpha \not\subset D'$ then α intersects $\gamma((0,d))$ transversally at (at least) two distinct points $x \neq y$. By minimality, there are subarcs of α and of γ from x to y with the same length. Thus, one can use these transverse intersections in a standard fashion to find a smaller curve in D connecting the end points of α, contradicting its minimality. $\square$

LEMMA A.8. *If the Gaussian curvature of g is everywhere not smaller than $H > 0$ then any two points $p, q \in D$ can be joined by a smooth geodesic arc γ satisfying $\gamma \subset D$, $L[\gamma] \leq \pi/\sqrt{H}$.*

Proof. According to Lemma A.6 we can find a smooth geodesic arc $\gamma : [0,1] \to D$ from p to q which is length minimizing among all piecewise smooth curves from p to q inside D. If $L[\gamma] > \pi/\sqrt{H}$ then for every $\epsilon > 0$ small enough we can find $t_\epsilon \in (\epsilon, 1)$ such that $\gamma(t_\epsilon)$ is conjugated to $\gamma(\epsilon)$ along $\gamma|_{[\epsilon,t_\epsilon]}$. Note that either γ is contained in a single side of D or γ maps $(0,1)$ into $\mathrm{int}(D)$. In latter case we use a Jacobi field J along $\gamma|_{[\epsilon,t_\epsilon]}$ satisfying $J(\epsilon) = 0$, $J(t_\epsilon) = 0$ to construct an interior variation of γ which decreases length, a contradiction. In the former note that, perhaps up to a change of sign, J can be arranged so that it produces variations into D which decrease length, again a contradiction. $\square$

Before moving to Toponogov's theorem and its consequence, we take a moment to study convex geodesic polygons on the 2-sphere equipped with its metric of constant curvature $H > 0$. This space is realized as a spherical shell of radius $H^{-1/2}$ sitting inside the euclidean 3-space, and will be denoted by S_H.

LEMMA A.9. *Let D be a convex geodesic polygon in S_H. Then the following hold.*

i) D coincides with the intersection of the hemispheres determined by its sides and the corresponding inward-pointing normal directions.

ii) The total perimeter of ∂D is not larger than $2\pi/\sqrt{H}$.

iii) If D has at least two sides then all sides of D have length at most $\pi/\sqrt{H}$.

Proof. Assertion iii) is obvious. The argument to be given below to prove i) and ii) is by induction on the number n of sides of D. The cases $n = 1, 2, 3$ are obvious.

Now fix $n > 3$ and assume that i), ii) and iii) hold for cases with $j < n$ sides. Let p, q, r be three consecutive vertices of D, so that minimal geodesic arcs γ_{pq}, γ_{qr} from p to q and from q to r, respectively, can be taken as two consecutive sides of D. Here we used that sides have length at most $\pi/\sqrt{H}$. Let $\gamma_1, \ldots, \gamma_{n-2}$ be the other sides of D

and denote by $H_{pq}, H_{qr}, H_1, \ldots, H_{n-2}$ the corresponding hemispheres determined by these sides and D.

We argue indirectly to show that $D \subset H_{pq} \cap H_{qr}$. If $x \in D \setminus (H_{pq} \cap H_{qr})$, consider a smooth geodesic arc γ from x to q inside D which minimizes length among piecewise smooth paths in D. γ exists by Lemma A.6 and, by the Lemma A.8, $L[\gamma] \leq \pi/\sqrt{H}$. Since x is not antipodal to q we have $L[\gamma] < \pi/\sqrt{H}$ which implies that γ is the unique minimal geodesic from x to q in S_H. Combining $x \notin H_{pq} \cap H_{qr}$ and Definition A.2 one concludes that γ is not contained in D, a contradiction. Repeating this argument for all triples of consecutive vertices we find that

$$D \subset H_{pq} \cap H_{qr} \cap H_1 \cap \cdots \cap H_{n-2}. \tag{66}$$

Now let $\gamma_{pr} \subset D$ be the smooth geodesic arc from p to r which is minimal relatively to D. This arc exists by Lemma A.6. Moreover, $\gamma_{pr} \setminus \{p, r\} \subset \operatorname{int}(D)$ since otherwise, by the previous lemma, $\gamma_{pr} \subset \partial D$ contradicting the fact that $n > 3$. Note that γ_{pr} divides D into $D = D' \cup T$, where D' is a convex geodesic polygon with sides $\gamma_{pr}, \gamma_1, \ldots, \gamma_{n-2}$, and T is the convex geodesic triangle bounded by $\gamma_{pq}, \gamma_{qr}, \gamma_{pr}$. Finally, let H_{pr} be the hemisphere determined by γ_{pr} and D', and let H'_{pq} is the closure of $S_H \setminus H_{pr}$. By the induction step $D' = H_{pr} \cap H_1 \cap \cdots \cap H_{n-2}$, and $T = H_{pq} \cap H_{qr} \cap H'_{pr}$. Thus

$$\begin{aligned}
&H_{pq} \cap H_{qr} \cap H_1 \cap \cdots \cap H_{n-2} \\
&= H_{pq} \cap H_{qr} \cap H_1 \cap \cdots \cap H_{n-2} \cap S_H \\
&= H_{pq} \cap H_{qr} \cap H_1 \cap \cdots \cap H_{n-2} \cap (H_{pr} \cup H'_{pr}) \\
&\subset (H_{pr} \cap H_1 \cap \cdots \cap H_{n-2}) \cup (H_{pq} \cap H_{qr} \cap H'_{pr}) \\
&= D' \cup T = D.
\end{aligned} \tag{67}$$

Hence (66) and (67) prove that i) holds for all convex geodesic polygons with at most n sides.

To prove ii) we again assume $n > 3$ and consider a, b, c, d four consecutive vertices of D, the consecutive sides $\gamma_{ab}, \gamma_{bc}, \gamma_{cd}$ connecting them, and let $\gamma_1, \ldots, \gamma_{n-3}$ be the other sides of D. Let H_{bc} be the hemisphere containing D whose equator contains γ_{bc}, and let H'_{bc} be the closure of $S_H \setminus H_{bc}$. If we continue γ_{ab} along b and γ_{cd} along c till they first meet at a point $e \in \operatorname{int}(H'_{bc})$. If γ_{be}, γ_{ec} are the minimal arcs connecting b to e and e to c, respectively, and T is the convex triangle with sides $\gamma_{be}, \gamma_{ec}, \gamma_{bc}$, then we claim that $F = D \cup T$ is a convex geodesic polygon with $n - 1$ sides. To see this the reader will notice that the closed curve $\alpha = \gamma_{ab} \cup \gamma_{be} \cup \gamma_{ec} \cup \gamma_{cd} \cup \gamma_1 \cup \cdots \cup \gamma_{n-3}$ is simple since $T \subset H'_{bc}$ and $D \subset H_{bc}$ (D satisfies i)), and $\alpha = \partial F$. By the induction step α has length $< 2\pi/\sqrt{H}$ and, since γ_{bc} is minimal, the length of ∂D is smaller than that of α. $\square$

A.2 The Relative Toponogov's Theorem

Toponogov's triangle comparison theorem is one of the most important tools in global Riemannian geometry. In the case of convex surfaces, it had been previously proven by Aleksandrov in [Ale48]. Here we need a relative version for triangles in convex geodesic polygons sitting inside positively curved two-spheres.

We fix a metric g on S^2, a convex geodesic polygon $D \subset S^2$, and follow [CE75] closely. However, we need to work with distances relative to D. For instance given points of D, the distance between them relative to D is defined to be the infimum of lengths of piecewise smooth paths in D connecting these points. Lemma A.5 tells us that the relative distance is realized by a smooth geodesic arc contained in D. We say that a (smooth) geodesic arc between two points of D is minimal relative to D if it realizes the distance relative to D.

A geodesic triangle in D is a triple of non-constant geodesic arcs (c_1, c_2, c_3) parametrized by arc-length, $c_i : [0, l_i] \to S^2$ (l_i is the length of c_i), satisfying $c_i([0, l_i]) \subset D$, $c_i(l_i) = c_{i+1}(0)$ and the triangle inequalities $l_i \leq l_{i+1} + l_{i+2}$ (indices modulo 3). These arcs may or may not self-intersect and intersect each other. The angle $\alpha_i \in [0, \pi]$ is defined as the angle between $-c'_{i+1}(l_{i+1})$ and $c'_{i+2}(0)$ (indices modulo 3).

Theorem A.10 (Relative Toponogov's Theorem). *Let g be a Riemannian metric on S^2 with Gaussian curvature pointwise bounded from below by a constant $H > 0$, and let $D \subset S^2$ be a convex geodesic polygon. If (c_1, c_2, c_3) is a geodesic triangle in D such that c_1, c_3 are minimal relative to D and $l_2 \leq \pi/\sqrt{H}$, then for every $0 < \epsilon < H$ there exists a so-called comparison triangle $(\bar{c}_1, \bar{c}_2, \bar{c}_3)$ in $S_{H-\epsilon}$ with angles $\bar{\alpha}_1, \bar{\alpha}_2, \bar{\alpha}_3$ such that $L[c_i] = L[\bar{c}_i]$ and $\bar{\alpha}_i \leq \alpha_i$, where α_i are the angles of (c_1, c_2, c_3).*

In [Kli82, page 297] Klingenberg observes that the relative version of Toponogov's theorem holds, and that this observation is originally due to Alexandrov [Ale48]. A proof of the above theorem would be too long to be included here, but the reader familiar with the arguments from [CE75] will notice two facts:

- The proof from [CE75] for the case of complete Riemannian manifolds essentially consists of breaking the given triangle into many 'thin triangles' (these are given precise definitions in [CE75, chapter 2]), and the analysis of these thin triangles is done by estimating lengths of arcs which are C^0-close to them. Hence all estimates of the perimeters of these thin triangles are obtained relative to an arbitrarily small neighborhood of the given convex geodesic polygon.
- Distances relative to the convex geodesic polygon are only at most a little larger than distances relative to a very small neighborhood of the convex geodesic polygon. This is easy to prove since we work in two dimensions.

Putting these remarks together the relative version of Toponogov's theorem can be proved using the arguments from [CE75].

Remark A.11. *A geodesic triangle in $S_{H-\epsilon}$ with sides of length at most $\pi/\sqrt{H}$, either is contained in a great circle, or its sides bound a convex geodesic polygon.*

A.3 The perimeter of a convex geodesic polygon

Theorem A.12. *Let (S^2, g) be a Riemannian two-sphere such that the Gaussian curvature is everywhere bounded from below by $H > 0$. If D is a convex geodesic polygon in (S^2, g) then the perimeter of ∂D is at most $2\pi/\sqrt{H}$. The same estimate holds for the perimeter of a two-gon consisting of two non-intersecting simple closed geodesic loops based at a common point.*

This is proved in [Kli82, page 297] for the case ∂D is a closed geodesic (no vertices). We reproduce the argument here, observing that it also works for the general convex geodesic polygon.

Proof of Theorem A.12. Let $d > 0$ be the perimeter of ∂D. We can parametrize ∂D as the image of a closed simple curve $c : \mathbb{R}/d\mathbb{Z} \to S^2$ which is a broken unit speed geodesic. For each $n \geq 1$ and $k \geq 0$ we denote by $\gamma_{k,2^n}$ a (smooth) geodesic arc from $c(kd2^{-n})$ to $c((k+1)d2^{-n})$ in D which minimizes length relative to D. We make these choices 2^n-periodic in k, $\gamma_{k+2^n,2^n} = \gamma_{k,2^n}$, and also choose $\overline{\gamma_{0,2} = \gamma_{1,2}}$. We can assume that $L[\gamma_{0,2}] < d/2$ since, otherwise, $d/2 \leq L[\gamma_{0,2}] \leq \pi/\sqrt{H}$ (Lemma A.8) and the proof would be complete. In particular, $\gamma_{0,2}$ is not contained in ∂D, and Lemma A.7 implies that $\gamma_{0,2}$ touches ∂D only at its endpoints $c(0), c(d/2)$.

Notice that if the distance from $c(kd2^{-n})$ to $c((k+1)d2^{-n})$ relative to D is $d2^{-n}$, then Lemma A.7 implies that $c|_{[kd2^{-n},(k+1)d2^{-n}]}$ is a smooth geodesic arc. Therefore, we are allowed to make the following important choice:

(C) If the distance from $c(kd2^{-n})$ to $c((k+1)d2^{-n})$ relative to D is $d2^{-n}$, then we choose $\gamma_{k,2^n} = c|_{[kd2^{-n},(k+1)d2^{-n}]}$.

The above choice forces $\gamma_{l,2^{n+m}}$ to be $c|_{[ld2^{-n-m},(l+1)d2^{-n-m}]}$ for all $k2^m \leq l < (k+1)2^m$, whenever $\gamma_{k,2^n} = c|_{[kd2^{-n}(k+1)d2^{-n}]}$.

For $n \geq 2$ set D_n to be the subregion of D bounded by the simple closed broken geodesic $\partial D_n = \cup\{\gamma_{k,2^n} \mid 0 \leq k < 2^n\}$. It follows readily from Lemma A.7 that this is a convex geodesic polygon. Moreover, sides of D_n fall into two classes: either a side is not contained in ∂D and coincides precisely with $\gamma_{k,2^n}$ for some k, or it lies in ∂D is a union of adjacent $\gamma_{k,2^n} \cup \gamma_{k+1,2^n} \cup \cdots \cup \gamma_{k+m,2^n} \subset \partial D$ for some k and some m. By construction

i) $D_n \subset D_{n+1}$ and $L[\partial D_n] \to d$ as $n \to \infty$.

ii) The vertices of D_n form a subset of $\{c(kd2^{-n}) \mid 0 \leq k < 2^n\}$.

Fix $0 < \epsilon < H$. We would like to construct a sequence of convex geodesic polygons $E_n \subset E_{n+1}$ in $S_{H-\epsilon}$ such that $L[\partial E_n] = L[\partial D_n]$.

Consider geodesic triangles $T_{k,2^n} = (\gamma_{k,2^n}, \gamma_{2k,2^{n+1}}, \gamma_{2k+1,2^{n+1}})$ in the sense of §A.2. The triangle inequalities hold, since all sides are minimal relative to D.

According to Theorem A.10, associated to $T_{0,2}, T_{1,2}$ there are comparison triangles $\bar{T}_{0,2} = (\bar{\gamma}_{0,2}, \bar{\gamma}_{0,4}, \bar{\gamma}_{1,4})$, $\bar{T}_{1,2} = (\bar{\gamma}_{1,2}, \bar{\gamma}_{2,4}, \bar{\gamma}_{3,4})$ in $S_{H-\epsilon}$ with sides of same length as the corresponding sides in $T_{0,2}, T_{1,2}$. The angles of $\bar{T}_{0,2}, \bar{T}_{1,2}$ are not larger than the corresponding angles on $T_{0,2}, T_{1,2}$. Up to reflection and a rigid motion, we can assume $\bar{\gamma}_{0,2}$ coincides with $\bar{\gamma}_{1,2}$ (along with vertices corresponding to endpoints of $\gamma_{0,2} = \gamma_{1,2}$) on a given great circle e, and $\bar{T}_{0,2}, \bar{T}_{1,2}$ lie on opposing hemispheres determined by e. Of course, $\bar{T}_{0,2}$ and/or $\bar{T}_{1,2}$ could lie on e, but this forces $L[\gamma_{0,2}]$ to be $d/2$, a case we already treated. Again the angle comparison can be used to deduce that $E_2 := \bar{T}_{0,2} \cup \bar{T}_{1,2}$ is a convex geodesic polygon in $S_{H-\epsilon}$ with the same perimeter as D_2 ($\partial E_2 = \cup_{k=0}^{3}\bar{\gamma}_{k,4}$).

To construct E_3, note that each side of D_2 not contained in ∂D is of the form $\gamma_{k,4}$ for some fixed $0 \leq k < 4$. Moreover, $\bar{\gamma}_{k,4}$ is a side of E_2 by construction and angle comparison. By Lemma A.7 $\gamma_{k,4}$ divides D into two convex geodesic polygons, only one of which, denoted by $D_{k,4}$, contains $c([kd/4, (k+1)d/4])$ in its boundary. By the same lemma, $T_{k,4}$ is contained in $D_{k,4}$ (and determines a convex geodesic polygon). By the relative Toponogov theorem, there exists a comparison triangle $\bar{T}_{k,4}$ which we can assume is of the form $(\bar{\gamma}_{k,4}, \bar{\gamma}_{2k,8}, \bar{\gamma}_{2k+1,8})$, i.e. one of its sides matches precisely the side $\bar{\gamma}_{k,4}$ of E_2 together with corresponding vertices of $\bar{\gamma}_{k,4}$. Moreover, possibly after reflection, we can assume E_2 and $\bar{T}_{k,4}$ lie on the opposing hemispheres determined by the great circle containing $\bar{\gamma}_{k,4}$. This last step strongly uses Lemma A.9

and Remark A.11. Again by the angle comparison, $E_2 \cup \bar{T}_{k,4}$ is a convex geodesic polygon in $S_{H-\epsilon}$ with the same perimeter as the convex geodesic polygon $D_2 \cup T_{k,4}$. Repeating this procedure for another side of D_2 not in ∂D, which is of the form $\gamma_{k',4}$ for some $k' \neq k$, with $E_2 \cup \bar{T}_{k,4}$ in the place of E_2, we obtain a larger geodesic convex polygon $E_2 \cup \bar{T}_{k,4} \cup \bar{T}_{k',4}$ in $S_{H-\epsilon}$ with the same perimeter as the geodesic convex polygon $D_2 \cup T_{k,4} \cup T_{k',4}$. After exhausting all the sides of D_2 not in ∂D we complete the construction of E_3.

The construction of E_n from D_{n-1}, E_{n-1} follows the same algorithm, since sides of D_{n-1} not in ∂D must be of the form $\gamma_{k,2^{n-1}}$ for some $0 \leq k < 2^{n-1}$. In this case, there will be a corresponding side $\bar{\gamma}_{k,2^{n-1}}$ of E_{n-1} with the same length as $\gamma_{k,2^{n-1}}$ along which we fit the comparison triangle $\bar{T}_{k,2^{n-1}}$ obtained by applying the relative Toponogov theorem to $T_{k,2^{n-1}}$. Doing this step by step at each side of D_{n-1} not in ∂D we obtain E_n.

By Lemma A.9 we know that

$$L[\partial D_n] = L[\partial E_n] \leq 2\pi/\sqrt{H-\epsilon}, \qquad \forall n.$$

Together with (i) above, we deduce that $L[\partial D] \leq 2\pi/\sqrt{H-\epsilon}$. Letting $\epsilon \downarrow 0$ we get the desired estimate.

To get the estimate for the two-gon as in the statement note that its perimeter can clearly be approximated by the perimeter of convex geodesic polygons. $\square$

References

[Ale48] A. D. Aleksandrov, *Vnutrennyaya Geometriya Vypuklyh Poverhnosteĭ*, [translated as: *Intrinsic geometry of convex surfaces*, in *A. D. Aleksandrov selected works, Part II*, Chapman & Hall/CRC, 2006], OGIZ, Moskow-Leningrad, 1948.

[APB14] J. C. Álvarez Paiva and F. Balacheff, *Contact geometry and isosystolic inequalities*, Geom. Funct. Anal. **24** (2014), 648–669.

[Bal06] F. Balacheff, *Sur la systole de la sphère au voisinage de la métrique standard*, Geom. Dedicata **121** (2006), 61–71.

[Bal10] F. Balacheff, *A local optimal diastolic inequality on the two-sphere*, J. Topol. Anal. **2** (2010), 109–121.

[Ban93] Victor Bangert, *On the existence of closed geodesics on two-spheres*, Internat. J. Math. **4** (1993), 1–10.

[Bir27] G. D. Birkhoff, *Dynamical systems*, Amer. Math. Soc. Colloquium Publ., American Mathematical Society, Providence, R.I., 1927.

[BM13] K. Burns and V. S. Matveev, *Open problems and questions about geodesics*, `arXiv:1308.5417 [math.DG]`, 2013.

[Cal70] E. Calabi, *On the group of automorphisms of a symplectic manifold*, Problems in analysis (Lectures at the Sympos. in honor of Salomon Bochner) (Princeton, N.J.), Princeton Univ. Press, 1970, pp. 1–26.

[CE75] J. Cheeger and D. G. Ebin, *Comparison theorems in Riemannian geometry*, North Holland Publishing Co., Amsterdam-Oxford, 1975.

[Cro88] C. B. Croke, *Area and length of the shortest closed geodesic*, J. Differential Geom. **18** (1988), 1–21.

[Fat80] A. Fathi, *Transformations et homéomorphismes préservant la mesure. Systèmes dynamiques minimaux*, Ph.D. thesis, Orsay, 1980.

[GG95] J.-M. Gambaudo and E. Ghys, *Enlacements asymptotiques*, Topology **36** (1995), 1355–1379.

[Gro83] M. Gromov, *Filling Riemannian manifolds*, J. Differential Geom. **18** (1983), 1–147.

[Gui76] V. Guillemin, *The Radon transform on Zoll surfaces*, Adv. Math. **22** (1976), 85–119.

[Kli59] W. Klingenberg, *Contributions to Riemannian geometry in the large*, Ann. of Math. **69** (1959), 654–666.

[Kli82] W. Klingenberg, *Riemannian geometry*, Walter de Gruyter & Co., Berlin, 1982.

[MS98] D. McDuff and D. Salamon, *Introduction to symplectic topology*, second ed., Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1998.

[NR02] A. Nabutovski and R. Rotman, *The length of the shortest closed geodesic on a 2-dimensional sphere*, Int. Math. Res. Not. (IMRN) (2002), 1211–1222.

[Rot06] R. Rotman, *The length of a shortest closed geodesic and the area of a 2-dimensional sphere*, Proc. Amer. Math. Soc. **134** (2006), 3041–3047.

[Sab04] S. Sabourau, *Filling radius and short closed geodesics of the 2-sphere*, Bull. Soc. Math. France **132** (2004), 105–136.

[Sab10] S. Sabourau, *Local extremality of the Calabi-Croke sphere for the length of the shortest closed geodesic*, J. London Math. Soc. **82** (2010), 549–562.

[Sch14] J. P. Schröder, *Ergodic components and topological entropy in geodesic flows of surfaces*, `arXiv:1407.6259 [math.DS]`, 2014.

(Alberto Abbondandolo) Ruhr Universität Bochum, Fakultät für Mathematik, Gebäude NA 4/33, D–44801 Bochum, Germany

E-mail address: `alberto.abbondandolo@rub.de`

(Barney Bramham) Ruhr Universität Bochum, Fakultät für Mathematik, Gebäude NA 5/32, D–44801 Bochum, Germany

E-mail address: `barney.bramham@rub.de`

(Umberto L. Hryniewicz) Universidade, Federal do Rio de Janeiro – Departamento de Matemática Aplicada, Av. Athos da Silveira Ramos 149, Rio de Janeiro RJ, Brazil 21941-909

E-mail address: `umberto@labma.ufrj.br`

(Pedro A. S. Salomão) Universidade de São Paulo, Instituto de Matemática e Estatística – Departamento de Matemática, Rua do Matão, 1010 - Cidada Universitãria - São Paulo SP, Brazil 05508-090

E-mail address: `psalomao@ime.usp.br`