GAFA Geometric And Functional Analysis

A ZOLL COUNTEREXAMPLE TO A GEODESIC LENGTH CONJECTURE

Florent Balacheff, Christopher Croke and
Mikhail G. Katz

Abstract. We construct a counterexample to a conjectured inequality $L \leq 2D$, relating the diameter D and the least length L of a nontrivial closed geodesic, for a Riemannian metric on the 2-sphere. The construction relies on Guillemin's theorem concerning the existence of Zoll surfaces integrating an arbitrary infinitesimal odd deformation of the round metric. Thus the round metric is not optimal for the ratio L/D.

1 Zoll Surfaces and Guillemin Deformation

Given a Riemannian metric on the 2-sphere, we consider its diameter D and the length L of its shortest nontrivial closed geodesic. The first inequality relating the two invariants was obtained by the second-mentioned author [C], who proved the bound $L \leq 9D$. The constant in the inequality was successively improved by M. Maeda [M], A. Nabutovsky and R. Rotman [NR1], and S. Sabourau [S]. The best known bound is $L \leq 4D$. Nabutovsky and Rotman conjectured the inequality $L \leq 2D$ [NR1, Intro.], meaning that the round metric of S^2 is optimal for the relationship between these two invariants. We give a few examples of surfaces satisfying the case of equality $L = 2D$:

1. a surface of revolution in $\mathbb{R}^3$ obtained from an ellipse with major axis on the x-axis;

2. a circular "pillow", obtained by doubling the flat unit disk;
3. a more general pillow obtained by doubling the region enclosed by a closed curve of constant width in the plane;
4. rotationally invariant Zoll surfaces.

The existence of such diverse examples may have led one to expect that none of these metrics is optimal for the ratio L/D.

It turns out that a counterexample to the inequality $L \leq 2D$ may be found among Zoll surfaces, namely surfaces all of whose geodesics are closed, and whose prime geodesics all have equal length 2π ([Z]). More precisely, while the rotationally symmetric Zoll surfaces do satisfy (the boundary case of equality of) the conjectured inequality, there exist other families of Zoll surfaces such that $L > 2D$. Such surfaces can be obtained as smooth variations of the round metric.

Let (S^2, g_0) be the 2-sphere endowed with the round metric. Denote by $a : S^2 \to S^2$ its antipodal map. Let $C_{\text{odd}}^\infty(S^2, \mathbb{R})$ be the space of smooth odd functions on S^2, i.e. smooth real valued functions f satisfying $f \circ a = -f$. The following existence theorem for Zoll surfaces is due to V. Guillemin [G].

Theorem 1.1 (Guillemin).   *For every* $f \in C_{\text{odd}}^\infty(S^2, \mathbb{R})$, *there exists a smooth one-parameter family* $g_t = \Psi_t^f g_0$ *of smooth Zoll metrics such that* $\Psi_0^f = 1$, *the conformal factor* Ψ_t^f *satisfies* $(d\Psi_t^f/dt)|_{t=0} = f$, *and all prime periodic geodesics of* (S^2, g_t) *have length* 2π.

Note that this result is a converse to P. Funk's theorem [F], to the effect that a smooth variation $g_t = \Phi_t g_0$ of the round metric by smooth Zoll metrics necessarily satisfies $(d\Phi_t/dt)|_{t=0} \in C_{\text{odd}}^\infty(S^2, \mathbb{R})$. A survey of Zoll surfaces appeared in [B, Chap. 4], see also [LM].

We exploit such Guillemin deformations to show that the round metric is not even a local maximum of the ratio L/D among Zoll surfaces. The precise statement of our result relies on the notion of a Y-like set.

DEFINITION 1.2.   A subset of the unit circle is called Y-*like* if it contains a triple of vectors $\{u, v, w\}$ such that there exist positive real numbers $a > 0$, $b > 0$, $c > 0$ satisfying $au + bv + cw = 0$. A subset of the unit tangent bundle US^2 of S^2 will be called Y-*like* if its intersection with the unit tangent vectors at p is Y-*like* for every $p \in S^2$.

Note that a subset of the unit circle is Y-like if and only if every open semicircle contains an element of the set.

We will denote by ds_0 the element of length for the round metric g_0 on the sphere. The notion of an amply negative function is motivated in Remark 2.4 below.

DEFINITION 1.3. An odd function f is called *amply negative* if the set of unit tangent directions to great half-circles τ, satisfying $\int_\tau f ds_0 < 0$, is a Y-like subset of US^2.

Theorem 1.4. *If f is an amply negative function then the smooth variation $\{g_t\} = \Psi_t^f g_0$ of the round metric g_0 by smooth Zoll metrics satisfies $L(g_t) > 2D(g_t)$ for sufficiently small $t > 0$.*

Combined with the existence of amply negative functions proved in section 5, our theorem yields the desired counterexample.

These metrics also provide a counterexample to another conjecture of Nabutovsky and Rotman [NR2, Conj. 1, p. 13]. Their conjecture would imply that for every point p of a closed Riemannian manifold (M, g), there is a nontrivial geodesic loop at p of length at most $2D(g)$. Here a geodesic loop is a geodesic segment with identical endpoints. This conjecture is easily seen to be true for non-simply connected manifolds, by exploiting non-contractible loops, cf. [K]. In our examples, the shortest geodesic loop at every p has length 2π, while the diameter is strictly smaller than π.

Sections 2 and 3 contain a proof of Theorem 1.4 modulo on the existence of amply negative functions. The existence of the latter is verified in sections 4 and 5.

2   Amply Negative Odd Functions

Our goal is to find amply negative functions $f \in C_{\mathrm{odd}}^\infty(S^2, \mathbb{R})$, such that the corresponding Guillemin deformation g_t of the standard round metric g_0 satisfies $D(g_t) < \pi$ for t small enough (while all geodesics remain closed of length 2π). By the compactness of the unit tangent circle bundle US^2, we obtain the following lemma.

LEMMA 2.1. *For every amply negative function f there is a constant $\nu(f) > 0$ with the following property. For every $(p, v) \in US^2$, there is a great half-circle τ issuing from $p \in S^2$, forming an acute angle with v, and satisfying $\int_\tau f ds_0 < -\nu(f)$.*

Denote by L_0 the length functional with respect to the round metric g_0. Given a geodesic segment γ of length $L_0(\gamma) < \pi$, we denote by P_γ the 1-parameter family of piecewise geodesic paths with the following two properties:

-   The path joins the endpoints of γ,

- The path consists of a pair of imbedded geodesic segments of equal length.

Elements of P_γ are parametrized by the non-smooth midpoint of the piecewise geodesic path, which traces out the equidistant great circle of the two endpoints. We let $SP_\gamma \subset P_\gamma$ be the closed subfamily consisting of the shorter paths, namely

$$SP_\gamma = \left\{ \tau \in P_\gamma \mid L_0(\tau) \leq \pi \right\}.$$

If γ is a great semi-circle, define P_γ to be the circular family of great half-circles joining the endpoints of γ, and the subfamily SP_γ to be the family of paths forming either an acute or a right angle with γ at the endpoints. The following lemma is obvious but crucial.

LEMMA 2.2. The family SP_γ for a geodesic segment γ with $L_0(\gamma) = \pi$ is the limit of the families SP_{γ_i} for subarcs γ_i of γ of length tending to π. In fact, if γ_i is any sequence of minimizing geodesic segments converging to γ, then SP_{γ_i} converges to SP_γ.

Our main technical tool in the next section will be the following result.

LEMMA 2.3. If f is amply negative then there is an $\epsilon > 0$ so that for all geodesic segments γ with $\pi - \epsilon \leq L_0(\gamma) \leq \pi$, there is a path $\tau \in SP_\gamma$ with $\int_\tau f ds_0 < -\nu(f)$.

Proof. If no such ϵ exists, then there is a sequence $\{\gamma_i\}$ with $L_0(\gamma_i) < \pi$ and $L_0(\gamma_i) \to \pi$ such that all $\tau \in SP_{\gamma_i}$ satisfy $\int_\tau f ds_0 \geq -\nu(f)$. There is a convergent subsequence such that $\gamma_i'(0) \to \gamma'(0)$ with $L_0(\gamma) = \pi$. By Lemma 2.2, the family SP_{γ_i} converges to SP_γ. By the continuity of f, for every $\tau \in SP_\gamma$ we have $\int_\tau f ds_0 \geq -\nu(f)$, contradicting the assumption that the function f is amply negative. $\square$

REMARK 2.4. Given a piecewise geodesic τ over which the integral of f is negative, we will show in the next section that the length of τ decreases under the Guillemin deformation. If, in addition, the curve τ has length at most π with respect to the metric g_0, then the length with respect to the metric g_t will be shorter than π. That is why we need to work with piecewise geodesics specifically in SP_γ. In order to make the continuity argument above work, one needs to find in each SP_γ, a curve τ over which f integrates negatively. This leads to the amply negative condition we introduced.

3 Diameter of Guillemin Deformation

Let Ψ^f be the conformal factor of the Guillemin deformation, as in Theorem 1.1 above. Thus, the metric $g_t = \Psi_t^f g_0$ is Zoll, while $\Psi_0^f = 1$ and $(d\Psi_t^f/dt)|_{t=0} = f$. Consider the arclength parametrisation $\tau(s)$ of a path $\tau \subset S^2$ for the round metric g_0.

LEMMA 3.1. *The energy $E_t(\tau)$ of a path $\tau \subset S^2$ for the metric g_t satisfies*

$$\left.\frac{dE_t}{dt}\right|_{t=0} = \int_\tau f \circ \tau \, ds_0 \,.$$

Proof. We have

$$\begin{aligned}
\frac{d}{dt} E_t(\tau) &= \frac{d}{dt} \int_0^{L_0(\tau)} g_t\big(\tau'(s), \tau'(s)\big) ds \\
&= \frac{d}{dt} \int_\tau \Psi_t^f \circ \tau \, ds_0 \\
&= \int_\tau \left(\frac{d}{dt} \Psi_t^f\right) \circ \tau \, ds_0 \\
&= \int_\tau f \circ \tau \, ds_0
\end{aligned}$$

at $t = 0$.   □

PROPOSITION 3.2. *If f is amply negative, then the associated Guillemin deformation $g_t = \Psi_t^f g_0$ as in Theorem 1.1 satisfies $D(g_t) < \pi$ for all sufficiently small $t > 0$.*

Proof. Denote by L_t and d_t the length and the distance with respect to the metric g_t. Let $\epsilon > 0$ be chosen as in Lemma 2.3, and let $\mathcal{A}_\epsilon \subset S^2 \times S^2$ be the set of nearly antipodal pairs, defined by setting

$$\mathcal{A}_\epsilon = \big\{ (p,q) \in S^2 \times S^2 \mid d_0(p,q) \geq \pi - \epsilon \big\} \,.$$

By continuity, there is a $\delta > 0$ such that whenever $0 < t < \delta$, we have

$$d_t(p,q) < \pi \quad \text{for all} \quad (p,q) \notin \mathcal{A}_\epsilon \,. \tag{3.1}$$

Now let $(p,q) \in \mathcal{A}_\epsilon$, and γ a minimizing geodesic joining them. Let

$$N(\gamma) = \left\{ \tau \in SP_\gamma \ \Big| \ \int_\tau f ds_0 \leq -\nu(f) \right\},$$

and let $N = \{\tau \in N(\gamma) \mid \pi - \epsilon \leq L_0(\gamma) \leq \pi\}$. By Lemma 2.3, whenever

$$\pi - \epsilon \leq L_0(\gamma) \leq \pi \,,$$

the set $N(\gamma)$ is non-empty. Furthermore, the sets N and $N(\gamma)$ are compact. Now for small $t > 0$, define a continuous function $F : N \times \mathbb{R} \to \mathbb{R}$ by setting $F(\tau, t) = dE_t(\tau)/dt$. By Lemma 3.1 and the definition of N, we have $F(\tau, 0) \leq -\nu(f)$. Hence by the compactness of N and the continuity of F there is a real $\delta' > 0$ so that for all $0 \leq t \leq \delta'$ and all $\tau \in N$, we have $F(\tau, t) < -\frac{1}{2}\nu(f)$. Therefore the energy given by the expression

$$\int_0^{L_0(\tau)} g_t\big(\tau'(s), \tau'(s)\big) ds_0$$

is strictly decreasing in t. Hence for $0 < t \leq \delta'$, it is strictly smaller than the quantity

$$\int_0^{L_0(\tau)} g_0\big(\tau'(s), \tau'(s)\big) ds_0 = L_0(\tau)\,.$$

In particular, we obtain for $0 < t \leq \delta'$,

$$L_t(\tau) = \int_0^{L_0(\tau)} \sqrt{g_t\big(\tau'(s), \tau'(s)\big)} ds_0$$
$$\leq L_0(\tau)^{\frac{1}{2}} \bigg(\int_0^{L_0(\tau)} g_t\big(\tau'(s), \tau'(s)\big) ds_0 \bigg)^{1/2}$$
$$< L_0(\tau)\,.$$

Thus for each pair $(p, q) \in \mathcal{A}_\epsilon$ and every $0 < t \leq \delta'$, there is a path τ from p to q with $L_t(\tau) < L_0(\tau) \leq \pi$. Hence $\mathrm{dist}_t(p, q) < \pi$. Combined with (3.1), this yields the diameter bound $D(g_t) < \pi$ whenever $0 < t < \min\{\delta, \delta'\}$, proving the proposition as well as Theorem 1.4. $\qquad\square$

4   Fine Sets and Their Properties

Recall that an open hemisphere is an open ball of radius $\pi/2$ centered at any point of the unit sphere. The construction of amply negative functions in section 5 exploits *fine* sets, in the following sense.

DEFINITION 4.1. A spherical pointset X is called *fine* if the following three conditions are satisfied:

1. No triple of X is collinear;
2. No triple of great circles pp', where $p, p' \in X$, is concurrent other than at points of X (as well as their antipodal points);
3. Every open hemisphere contains at least 3 of the points of X.

Note that the non-collinearity implies, in particular, that X contains no pair of antipodal points. Meanwhile, condition (3) implies that at every

point of the sphere, there is a Y-like set of tangent directions leading to points of X.

To see that fine sets exist, start with the set of 4 vertices of the regular inscribed tetrahedron. This gives a set with at least one point in every open hemisphere. We replace each point of the tetrahedron by a generic triple of nearby points. The non-collinearity and non-concurrency follow from genericity, and property (3) follows by construction.

DEFINITION 4.2. Given a fine pointset X, choose $\epsilon(X) > 0$ such that

1. The closed $\epsilon(X)$ balls centered at the points of $X \cup -X$ are disjoint;
2. There are at least 3 points of X in $B(p, \pi/2 - \epsilon(X))$ for every $p \in S^2$.

We note that property (3) of fineness along with standard compactness arguments shows that such a positive $\epsilon(X)$ exists.

LEMMA 4.3. *Let X be a fine set and choose $\epsilon(X)$ as above. Let Σ the set of unit vectors in US^2 tangent to geodesic segments τ of length π satisfying the following two conditions:*

- $\tau(0, \pi) \cap -X = \emptyset$;
- $\tau(\epsilon(X), \pi - \epsilon(X)) \cap X \neq \emptyset$.

Then Σ is Y-like.

Proof. Fix a unit vector v at $p \in S^2$ and let γ be the corresponding geodesic segment of length π. We need to find a $w \in \Sigma$ at p making an acute angle with v. Let H be the (closed) hemisphere obtained as the union of the $\tau \in SP_\gamma$. Then by assumption there are at least three points of X (call them p_1, p_2, and p_3) in the interior of H and at a distance greater than $\epsilon(X)$ from the boundary of H hence the endpoints of γ. Hence there are at least 3 geodesic segment τ_1, τ_2, and τ_3 in the interior of SP_γ passing through the points p_i, $i = 1, 2, 3$. If two of these paths coincide (say $\tau_1 = \tau_2$) then τ_1 passes through p_1 and p_2 so the initial point p of γ is not in X and τ_1 avoids $-X$ (by condition 1 of being fine). If all of these paths are pairwise distinct and also pass through points of $-X$ (say $-p_4, -p_5$, and $-p_6$ respectively) then the initial point p of γ would lie on the 3 great circles $p_1 p_4$, $p_2 p_5$, $p_3 p_6$ which contradicts either condition 2 (if $p \notin X \cup -X$) or condition 1 (if $p \in X \cup -X$). Thus we see that there is a τ in the interior of SP_γ containing an element of X at least $\epsilon(X)$ from the endpoints, and no element of $-X$ in its interior. The tangent vector of τ is the w we seek. □

5   Existence of Amply Negative Functions

The goal of this section is to prove the following proposition.

PROPOSITION 5.1.   *There exist amply negative functions.*

We will construct such functions by defining odd functions that approximate the sum of $\pm\delta$ (Dirac delta) functions centered at points of $-X$ and X for a fine set X. For our approximate δ functions we take for each $p \in S^2$ the smooth function δ_p^ϵ with support included in the ball $B(p, \epsilon)$ with

$$\delta_p^\epsilon(q) = \exp(1/\epsilon) \cdot \exp\left(\frac{1}{d(p, q) - \epsilon}\right)$$

for $q \in B(p, \epsilon)$.

We will use the following (nearly obvious) lemma.

LEMMA 5.2. *If γ is a diameter of $B(p, \epsilon)$ (i.e. geodesic through the center of length 2ϵ) and τ is any geodesic segment in $B(p, \epsilon)$ then $\int_\tau \delta_p^\epsilon \leq \int_\gamma \delta_p^\epsilon$ with equality holding if and only if τ is also a diameter.*

Proof. To see this (since $\delta_p^\epsilon \geq 0$) we can assume (by extending τ if needed) that τ runs from a boundary point to a boundary point and has length $2l < 2\epsilon$. Since for $t \leq l$ we have $d(p, \tau(t)) \geq \epsilon - t = d(p, \gamma(t))$ we have

$$\int_\tau \delta_p^\epsilon = 2\int_0^l \delta_p^\epsilon(\tau(t))dt \leq 2\int_0^l \delta_p^\epsilon(\gamma(t)) < 2\int_0^\epsilon \delta_p^\epsilon(\gamma(t)) = \int_\gamma \delta_p^\epsilon. \qquad \square$$

We are now ready to define our functions.

DEFINITION 5.3. For $\epsilon(X) > \epsilon > 0$ set

$$f_X^\epsilon = \sum_{p_i \in X} \left(\delta_{-p_i}^\epsilon - \delta_{p_i}^\epsilon\right).$$

Note that f_X^ϵ is a smooth odd function. We will now prove that for sufficiently small $\epsilon > 0$ the function f_X^ϵ is amply negative.

LEMMA 5.4.   *For every $v \in US^2$ there is an $\epsilon(v)$ with $\epsilon(X) > \epsilon(v) > 0$ and an open neighborhood $U(v)$ of v in US^2 (note that the base point also varies) such that for all $w \in U(v)$ there is a geodesic segment τ of length π whose initial tangent vector makes an acute angle with w (hence it starts at the base point of w) and*

$$\int_\tau f_X^\epsilon < 0$$

for all $\epsilon(v) > \epsilon > 0$.

Note that the $\epsilon(v) > 0$ we find in the proof below will tend to 0 as the base point of v tends to X (while not being in X). This turns out not to be a problem since by the compactness of US^2 a finite number $U(v_i)$ cover US^2 and hence we can take any ϵ less than the smallest of the $\epsilon(v_i)$ and have that f_X^ϵ is amply negative. This therefore proves Proposition 5.1.

Proof of Lemma 5.4.   To prove the lemma, we consider two cases. First assume that the base point of v is not in $X \cup -X$. Let τ (whose existence is promised in Lemma 4.3) be a geodesic segment of length π making an acute angle with v that misses $-X$ and passes through at least one $p \in X$ that has distance greater than $\epsilon(X)$ from its endpoints. Thus we can choose $\epsilon(v)$ so small that τ misses $B(q, 2\epsilon(v))$ for all $q \in -X$. Now for w in a small enough neighborhood U of v, let $\bar\tau$ be the geodesic segment of length π through the base point of w and p. For small enough U, $\bar\tau$ will still miss all the $B(q, \epsilon(v))$ for $q \in -X$ while $\bar\tau$ will make an acute angle with w. Thus, for $\epsilon(v) \geq \epsilon > 0$, we have $f_X^\epsilon \leq 0$ along $\bar\tau$ and is negative near p so we see $\int_{\bar\tau} f_X^\epsilon < 0$.

In the second case the basepoint p_0 of v is in $X \cup -X$. We will assume $p_0 \in -X$ since the other case is the same (by reversing orientation of all geodesics). Note that any τ making an acute angle with v which intersects X in its interior cannot also intersect $-X$ in its interior by property 1 of a fine set. So there are two (in fact three) geodesic segments τ_1 and τ_2 in the interior of SP_γ that pass through a p_1 and a p_2 respectively and no element of $-X$ in its interior and such that p_1 and p_2 have distance greater than $\epsilon(X)$ from p_0 and $-p_0$. Now we choose $\epsilon(v) > 0$ so small that for all $q \in -X$ and $q \neq p_0$, $B(q, 2\epsilon(v))$ miss both τ_1 and τ_2. Again for w in a small neighborhood U of v let $\bar\tau_1$ (resp. $\bar\tau_2$) be the geodesic segment of length π starting at the basepoint of w and passing through p_1 (resp. p_2).

If U is small enough we can assume both that τ_i make acute angles with w and that for all $q \in -X$ with $q \neq p_0$, $\bar\tau_1$ and $\bar\tau_2$ miss $B(q, \epsilon(v))$. Along $\bar\tau_1$ (resp. $\bar\tau_2$) we have, for $\epsilon(v) \geq \epsilon > 0$, $f_X^\epsilon \leq 0$ except on

$$\bar\tau_1 \cap B(p_0, \epsilon)$$

(respectively, on $\bar\tau_2 \cap B(p_0, \epsilon)$). Both $\bar\tau_1 \cap B(p_0, \epsilon)$ and $\bar\tau_2 \cap B(p_0, \epsilon)$ cannot be diameters since that would put p_0, p_1 and p_2 on the same great circle (namely the one through p_0 and the base point of w). So assume $\bar\tau_1 \cap B(p_0, \epsilon)$ is not a diameter. Then since we know on the other hand that $\bar\tau_1 \cap B(p_1, \epsilon)$ is a diameter, Lemma 5.4 tells us that $\int_{\bar\tau_1} f_X^\epsilon < 0$.   □

References

[B]   A. BESSE, Manifolds all of Whose Geodesics are Closed, Ergebisse Gren-
        zgeb. Math. 93, Springer, Berlin, 1978.

[C]   C. CROKE, Area and the length of the shortest closed geodesic, J. Differ-
        ential Geom. 27:1 (1988), 1–21.

[F]   P. FUNK, Über Fläschen mit lauter geschlossenen geodätischen Linien,
        Math. Ann. 74 (1913), 278–300.

[G]   V. GUILLEMIN, The Radon transform on Zoll surfaces, Advances in Math.
        22 (1976), 85–119.

[K]   M. KATZ, Systolic Geometry and Topology, with an appendix by J.P.
        SOLOMON, Mathematical Surveys and Monographs 137, American Math-
        ematical Society, Providence, RI, 2007.

[LM]   C. LEBRUN, L. MASON, Zoll manifolds and complex surfaces, J. Differen-
        tial Geom. 61:3 (2002), 453–535.

[M]   M. MAEDA, The length of a closed geodesic on a compact surface, Kyushu
        J. Math. 48:1 (1994), 9–18.

[NR1]   A. NABUTOVSKY, R. ROTMAN, The length of the shortest closed geodesic
        on a 2-dimensional sphere, Int. Math. Res. Not. 23 (2002), 1211–1222.

[NR2]   A. NABUTOVSKY, R. ROTMAN, Lengths of geodesics between two points
        on a Riemannian manifold, Electron. Res. Announc. Amer. Math. Soc. 13
        (2007), 13–20.

[S]   S. SABOURAU, Filling radius and short closed geodesics of the 2-sphere,
        Bull. Soc. Math. France 132:1 (2004), 105–136.

[Z]   O. ZOLL, Über Flächen mit Scharen geschlossener geodätischer Linien (in
        German), Math. Ann. 57:1 (1903), 108–133.

FLORENT BALACHEFF, Section de Mathématiques, Université de Genève, Suisse
                                    florent.balacheff@math.unige.ch

CHRISTOPHER CROKE, Department of Mathematics, University of Pennsylvania,
Philadelphia, PA 19104-6395 USA   ccroke@math.upenn.edu

MIKHAIL G. KATZ, DEPARTMENT OF MATHEMATICS, BAR ILAN UNIVERSITY,
RAMAT GAN 52900 ISRAEL   katzmik@math.biu.ac.il