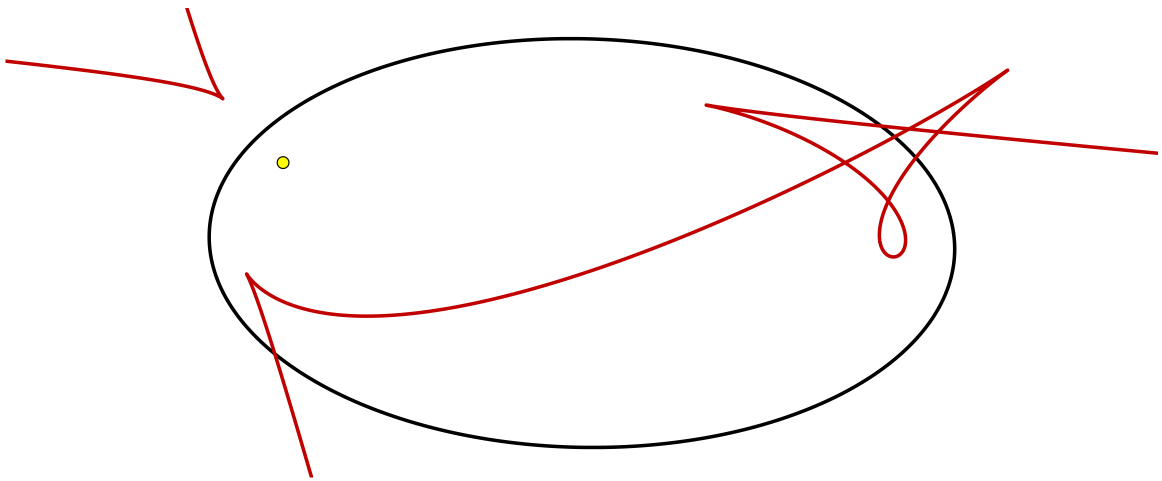
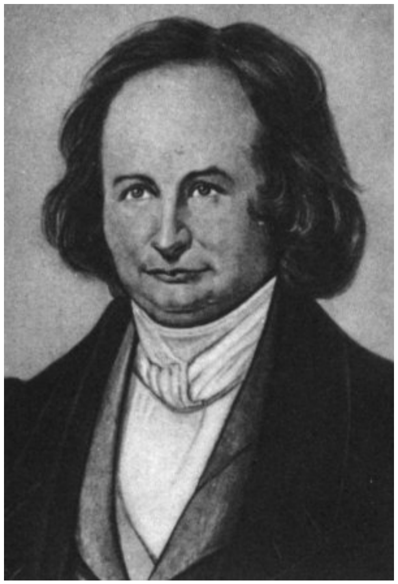
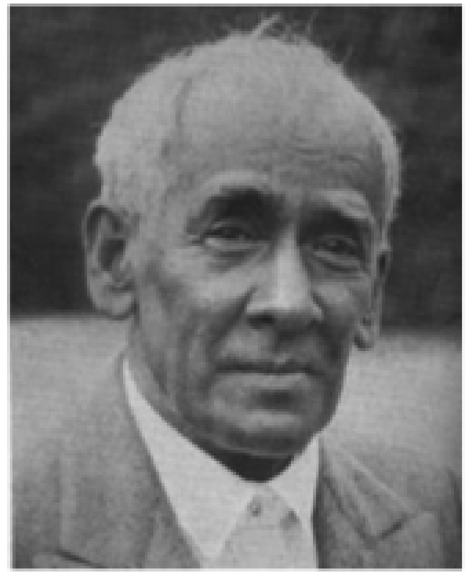
4-point theorems, a nostalgic trip

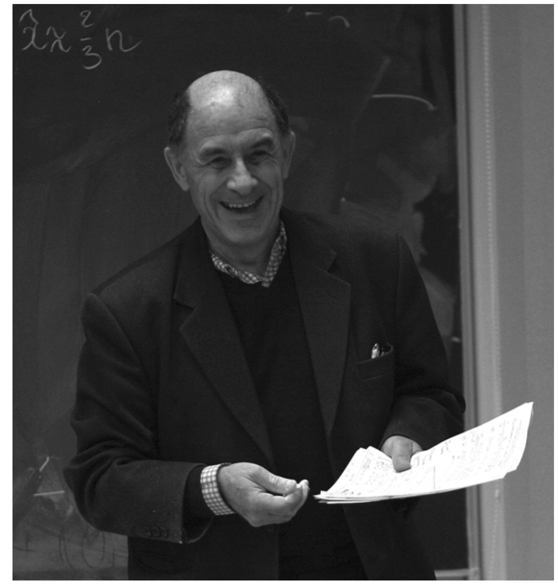
New trends in Geometry, Combinatorics, and Mathematical Physics

Vieille Perrotine, Oleron, October 2024



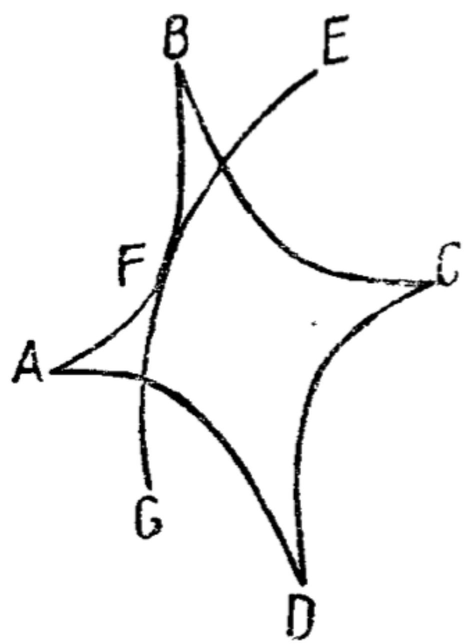






Part 1: cusps of caustics by reflection

The Last Geometric Statement of Jacobi ("Lectures on Dynamics"): *the conjugate locus of a (non-umbilic) point of a triaxial ellipsoid has exactly four cusps.*



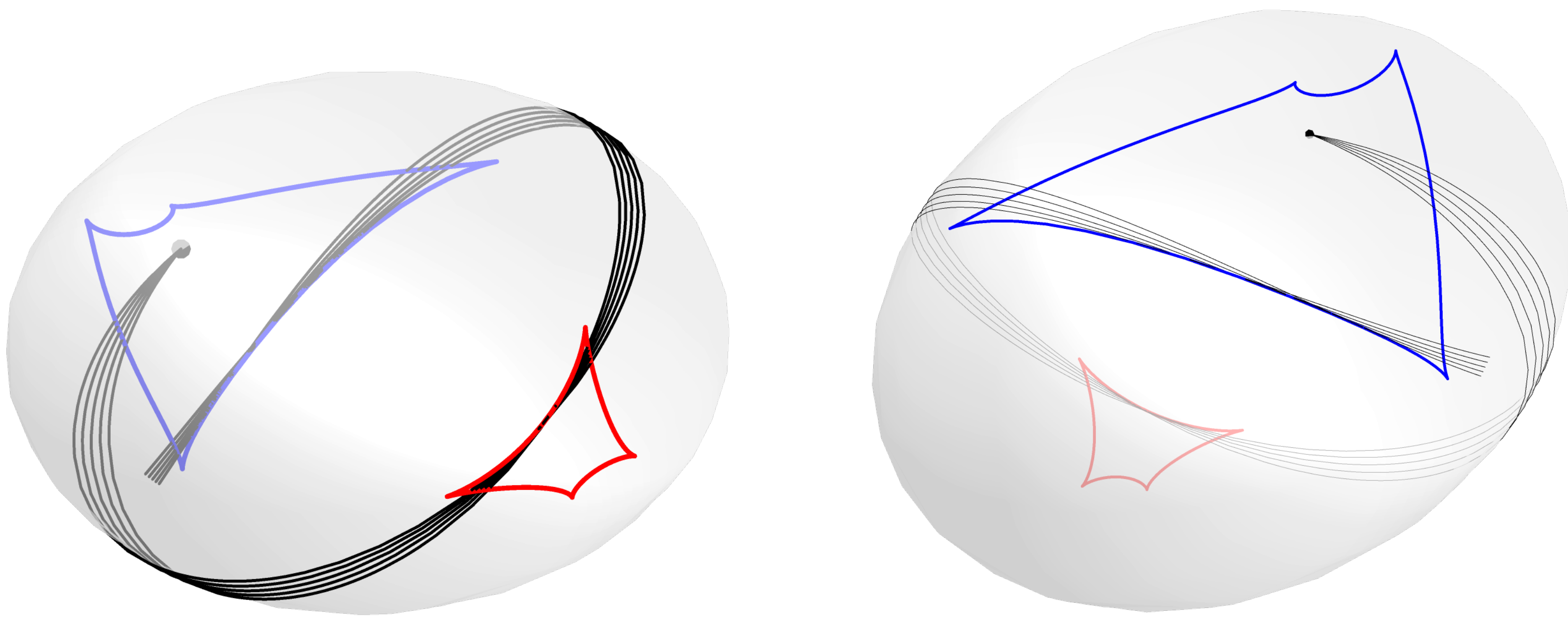
To quote Marcel Berger (*Riemannian Geometry During the Second Half of the Twentieth Century*, AMS, 2002):

> ... But this latter assumption depends on the scandalously unproved Jacobi "statement": the conjugate locus of a non-umbilical point of an ellipsoid has exactly four cusps.

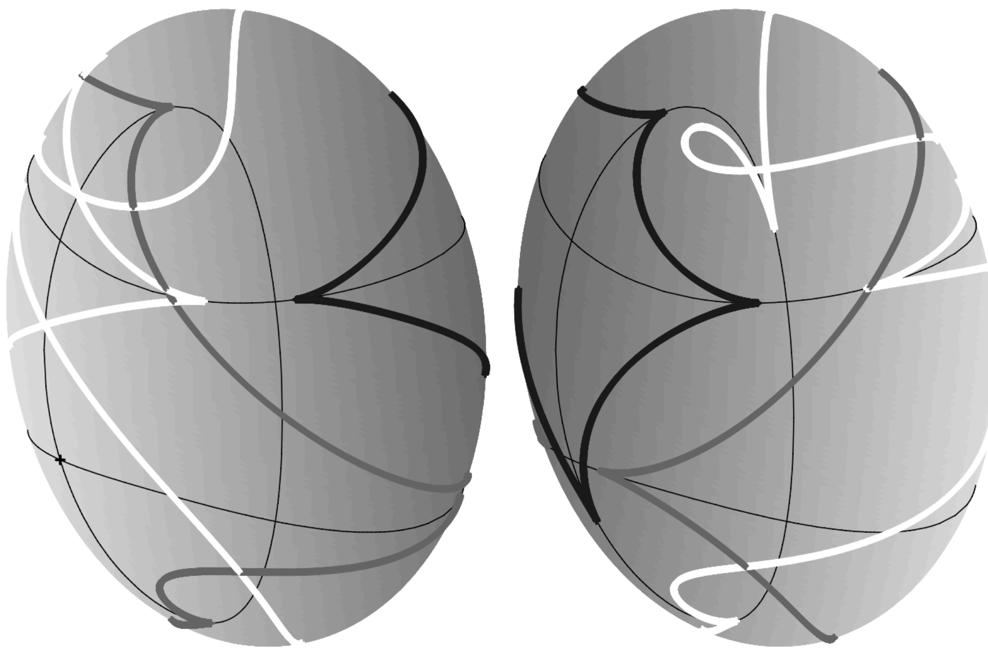
Jacobi's statement was proved only in this century:

J. Itoh, K. Kiyohara. *The cut loci and the conjugate loci on ellipsoids*, Manuscripta Math. 114 (2004), 247–264.

Likewise, the n*th caustic* is the locus of nth intersections of infinitesimally close geodesic emanating from O. (This can be defined in terms of zeros of the Jacobi fields along the geodesics.)



Experiments with nth caustics: conjecturally, they all have four cusps:



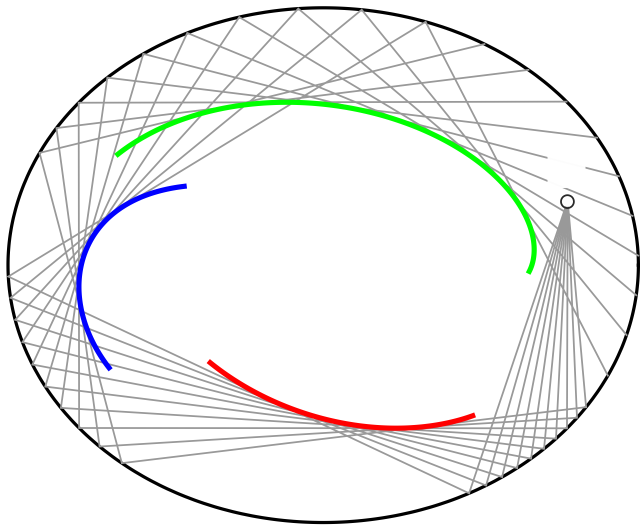
The first, second, and third caustics (black, grey, and white): R. Sinclair. *On the Last Geometric Statement of Jacobi*, Experimental Math. 12 (2003), 477–485.

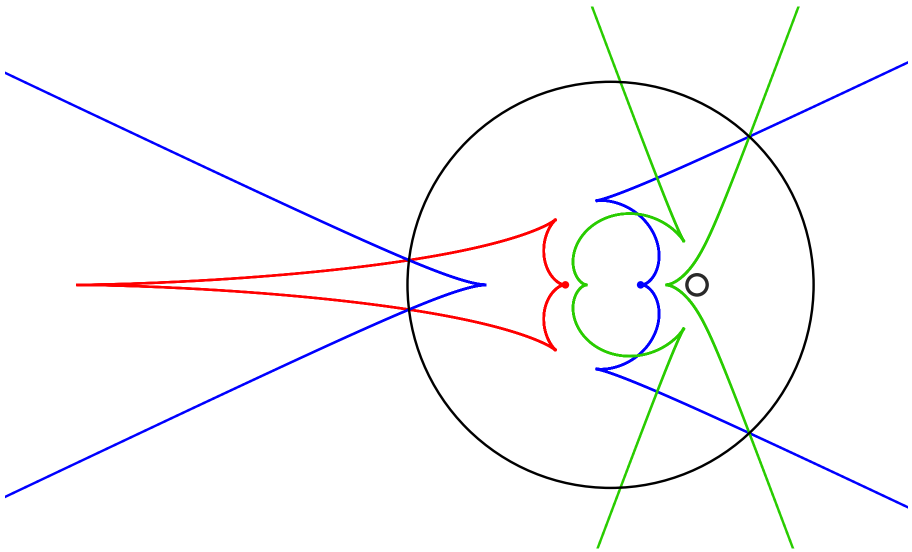
Also one has a version of the 4-vertex theorem: *the conjugate locus of a generic point on a convex surface has at least four cusps*.

W. Blaschke attributed this “vierspitzensatz” to C. Carathéodory. For a recent proof, see
T. Waters. *The conjugate locus on convex surfaces*, Geom. Dedicata 200 (2019), 241–254.

V. Arnold, and later V. Vassiliev, proved that this results extends to the nth caustic, provided the surface is sufficiently close to the round sphere (the needed closeness increases with n).

Billiard version: caustics by reflection





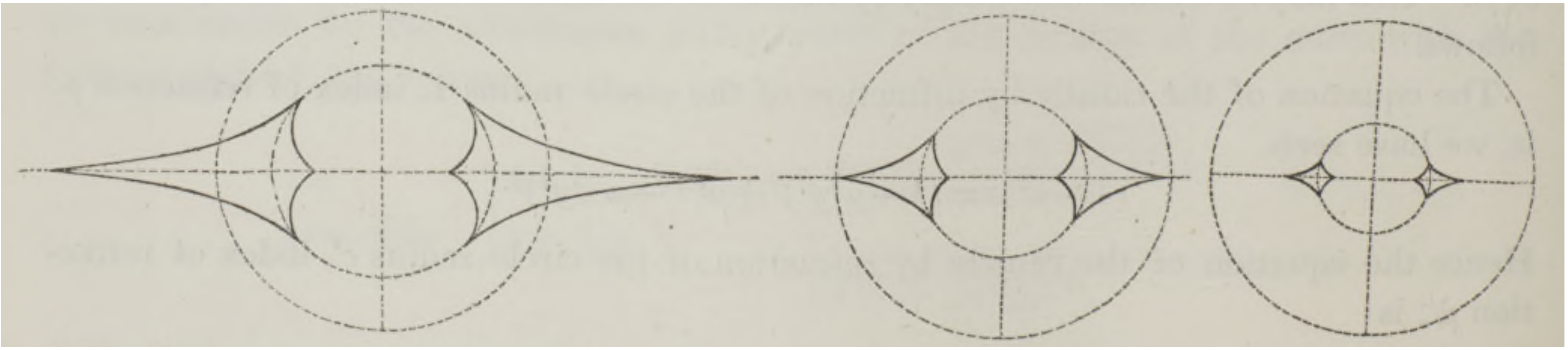
The caustics may escape to infinity, but they are closed in $\mathbf{RP}^2$ (the envelopes are projectively dual to the curves in the space of lines).

The case of the first caustic by reflection, the *catacaustic*, is well studied. See, e.g.,

J. Bruce, P. Giblin, C. Gibson. *Caustics through the looking glass.* Math. Intelligencer **6** (1984), no. 1, 47–58, or

A. Cayley. *A memoir upon caustics.* Philos. Trans. Royal Soc. London 147 (1857), 273–312,

where also the first caustics by refraction were considered.

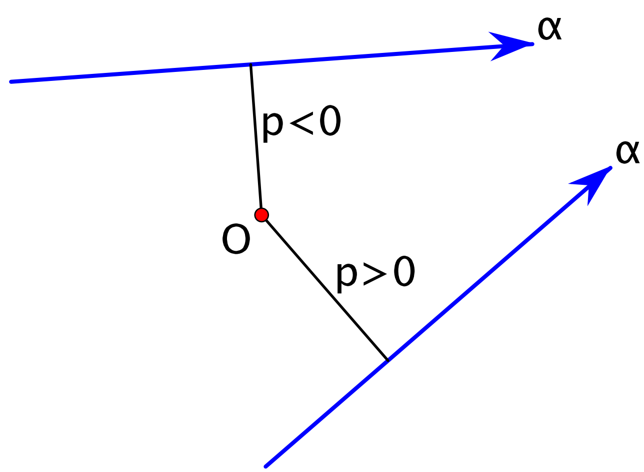


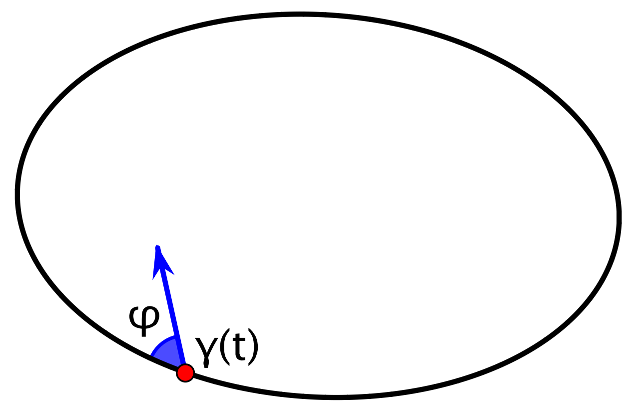
Theorem 1 [G. Bor, ST 2023]: *For every* $n \geq 1$, *the* n*th caustic by reflection* Γ_n *in an oval has at least four cusps.*

The same result holds in the spherical and hyperbolic geometries, and for more general Finsler billiards with respect to projective, but not necessarily symmetric, Finsler metrics (the topic of Hilbert's 4th Problem).

Sketch of a proof

The space of oriented lines $\mathcal{L}$ with its area form $\omega = dp \wedge d\alpha$.





Also: the inward unit tangent vectors with the foot point on the arc length parameterized curve $\gamma(t)$, making angle φ with it.

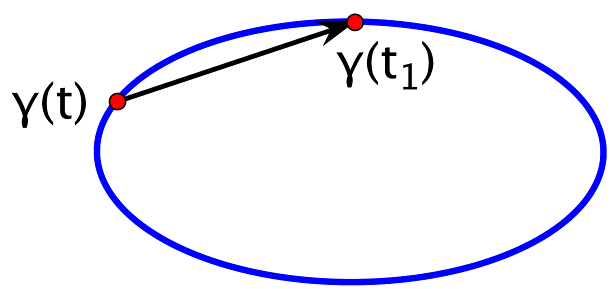
Then $\omega = \sin\varphi \; dt \wedge d\varphi = d(\cos\varphi \; dt)$, and the billiard map

$$T : (t, \varphi) \mapsto (t_1, \varphi_1)$$

is *exact symplectic*:

$$\cos\varphi_1 \; dt_1 - \cos\varphi \; dt = dL,$$

where $L = |\gamma(t), \gamma(t_1)|$ is the generating function.



Furthermore, $pd\alpha$ is cohomologous to $\cos\varphi \; dt$, hence

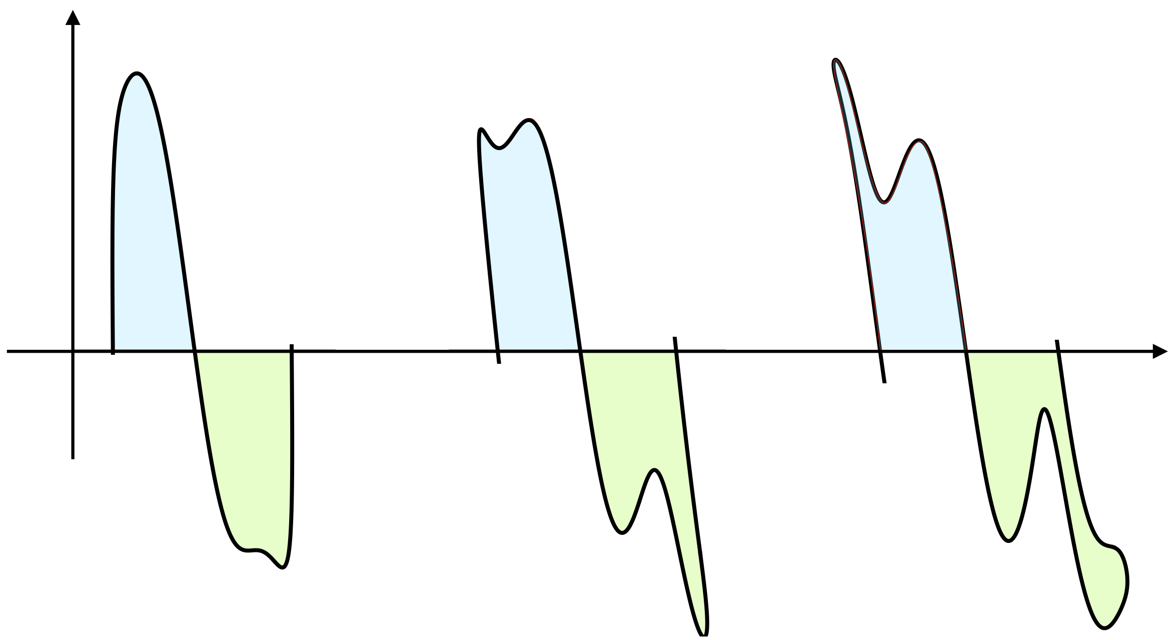
$$T^*(pd\alpha) - pd\alpha = dF.$$

The pencils of rays through fixed points are curves in the space of lines $\mathcal{L}$. In the spirit of projective duality, one could call them "lines". In the (α, p)-coordinates, they are sine curves

$$p = a \cos \alpha + b \sin \alpha.$$

The cusps of the caustic are the second order tangencies with these "lines", *inflections*, of the curve C_n in the space of rays $\mathcal{L}$, comprising the rays that have undergone n reflections. The curve C_n is projectively dual to the caustic Γ_n.

Let C_0 be the circle in the phase cylinder representing the initial beam, the zero section of the cylinder, and let $C_n = T^n(C_0)$. Since T is exact symplectic, C_n bounds zero area.



We are interested in the number of inflections of C_n.

Claim 1: *C_n is homotopic to C_0 in the class of smooth embedded closed curves that bound zero area.*

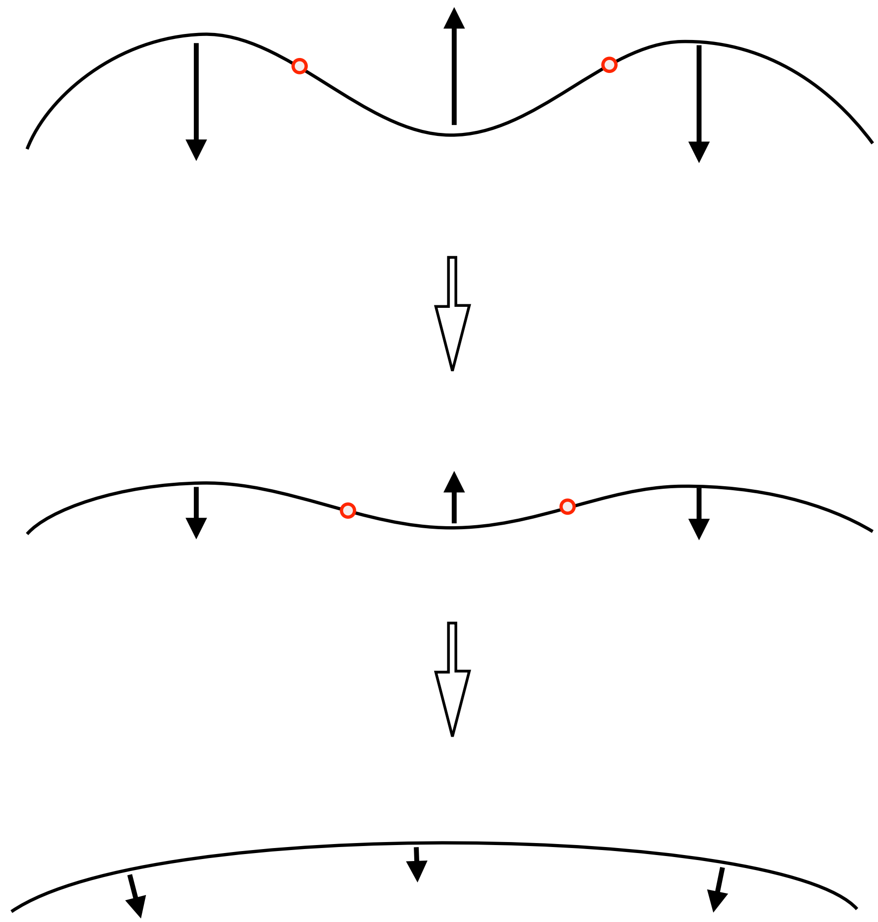
The *curve shortening flow* takes C_n to a horizontal curve on the phase cylinder, and the area is preserved by the flow:

$$\frac{d(\int p d\alpha)}{dt} = \int k(s) ds = 0,$$

the second equality for the winding number holds for every simple closed arc length parameterized curve $C(s)$ that goes around the cylinder.

Claim 2: *The number of inflections can only decrease under the curve shortening flow.*

This is a manifestation of the maximum principle. See S. Angenent. *Inflection points, extatic points and curve shortening*.



Claim 3: *When C_n becomes a graph $p = F(\alpha)$, the inflections correspond to the solutions of*

$$G(\alpha) := F''(\alpha) + F(\alpha) = 0.$$

This is because the first harmonics comprise the kernel of the operator $d^2 + 1$.

Since F has zero average, the Fourier expansion of $G(\alpha)$ starts with the second harmonics (or higher).

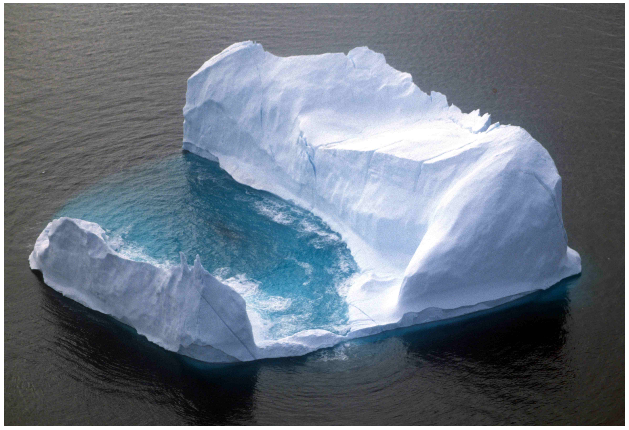
The result now follows from the Sturm-Hurwitz theorem: *the number of roots of a 2π-periodic function is not less than the number of roots of its first harmonic*.

This theorem has a proof by the heat equation (G. Polya, 1933):

Let $G(\alpha) = G(\alpha, 0)$ be the initial distribution of heat on the circle. Consider the propagation of heat:

$$\frac{\partial G(\alpha,t)}{\partial t} = \frac{\partial^2 G(\alpha,t)}{\partial \alpha^2}.$$

The number of sign changes of $G(\alpha, t)$ (as a function of α) does not increase with t: an iceberg can melt down in a warm sea, but cannot appear out of nowhere (the *maximum principle* in PDE).



One solves the heat equation:

$$G(\alpha, t) = \sum_{k \geq n} e^{-k^2 t} \left(a_k \cos k\alpha + b_k \sin k\alpha\right).$$

The higher harmonics tend to zero faster than the first non-trivial one. As $t \to \infty$, the (renormalized) function tends to its first non-trivial harmonic, that has exactly $2n$ zeroes. Hence $G(\alpha)$ has no less than $2n$ zeroes.

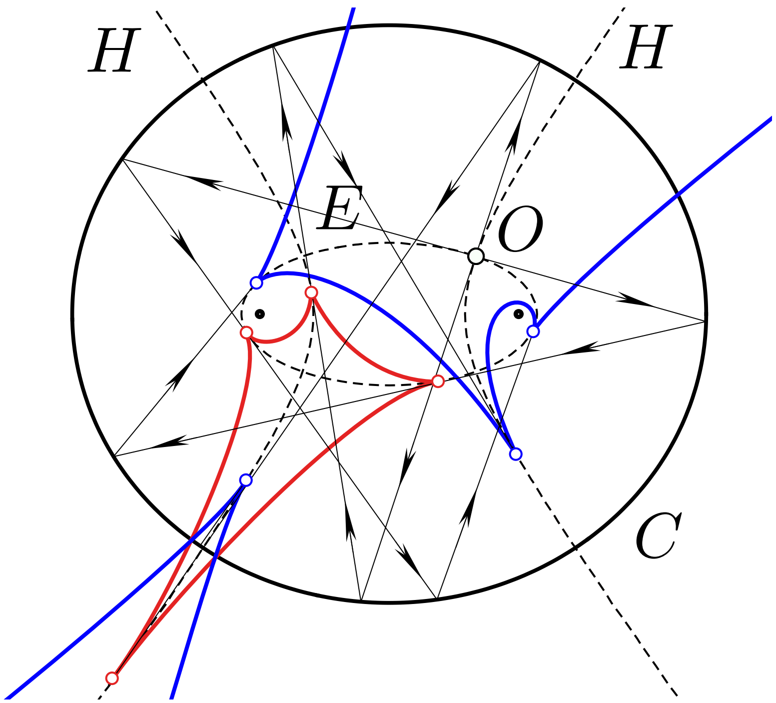
The case of ellipses

Conjecture 1: *If the curve is an ellipse, then all caustics by reflection generically have exactly four cusps (if the source of light is not a focus).*

Conjecture 2: *If the curve is not an ellipse then, for some choice of the source of light and some $n \geq 1$, the nth caustic by reflection has more than four cusps.*

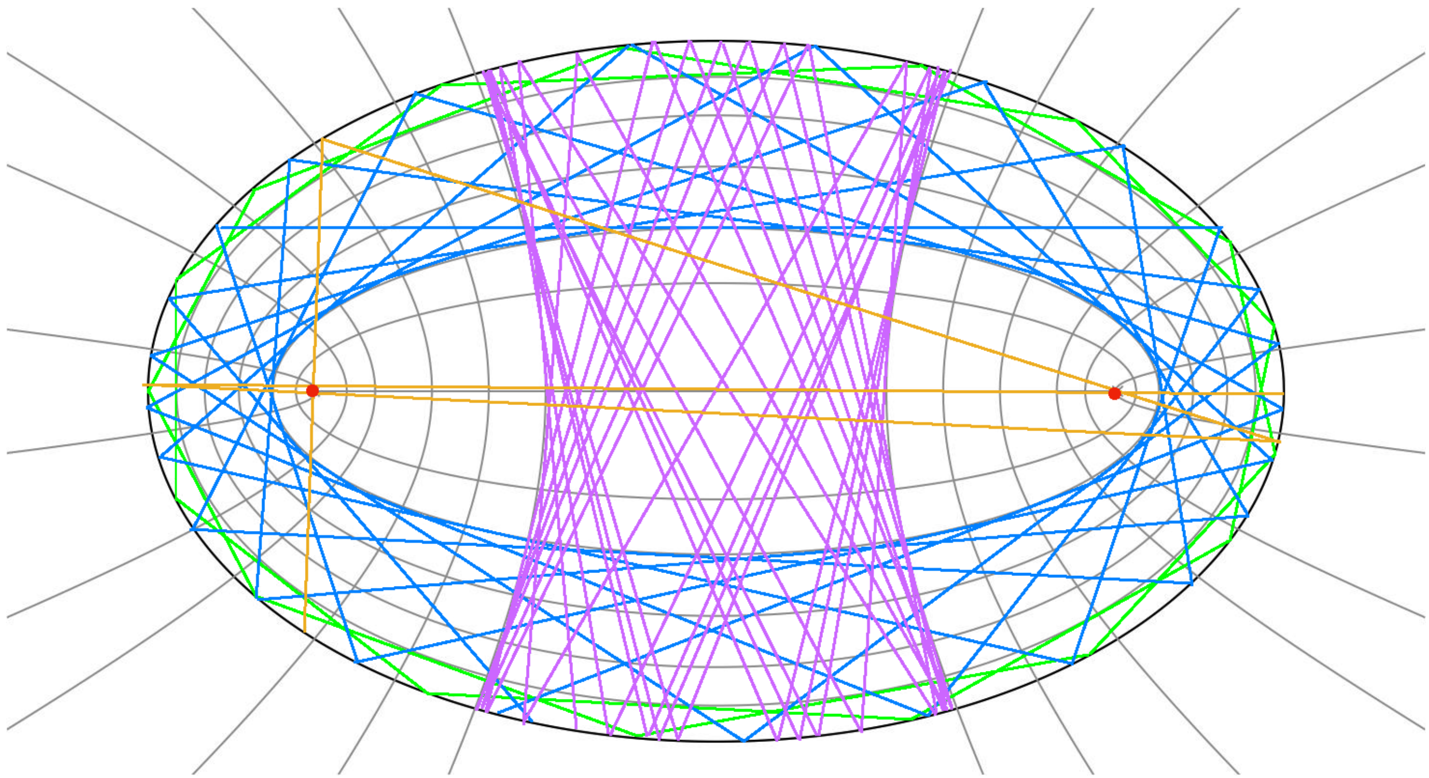
Theorem 2 [G. Bor-M. Spivakovsky-ST 2024]: *Conjecture 1 holds for circles.*

Theorem 3 [G. Bor-M. Spivakovsky-ST 2024]: *Consider the four rays tangent to the two confocal conics through point O. After n reflections, their tangency points to these conics are cusps of the nth caustic by reflection.*

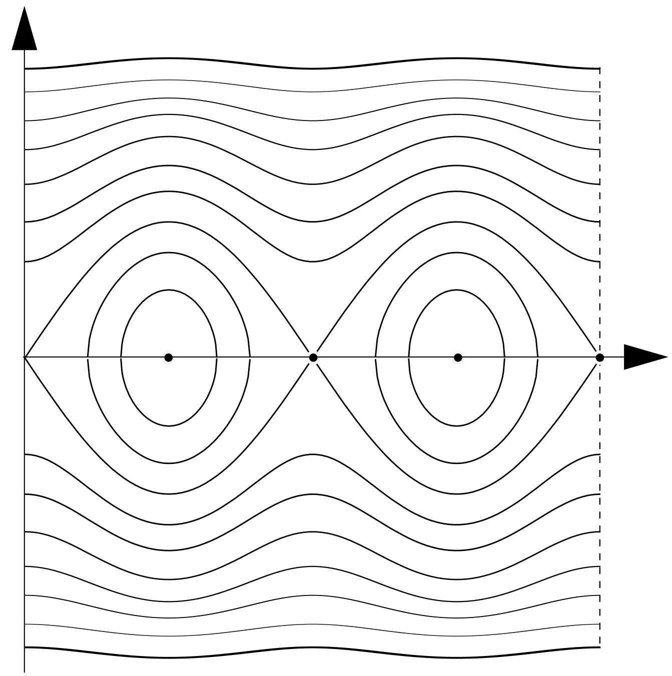


Sketch of a proof

The billiard ball map T is integrable:



Phase portrait:

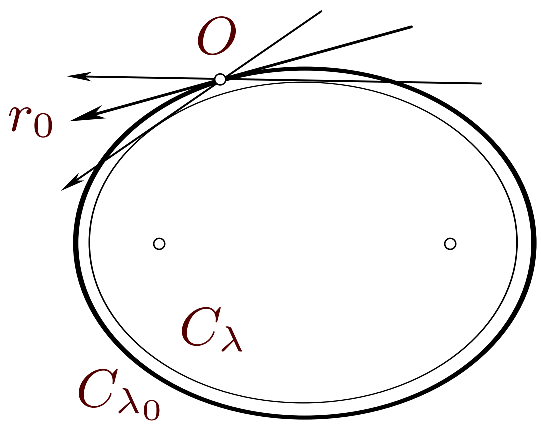


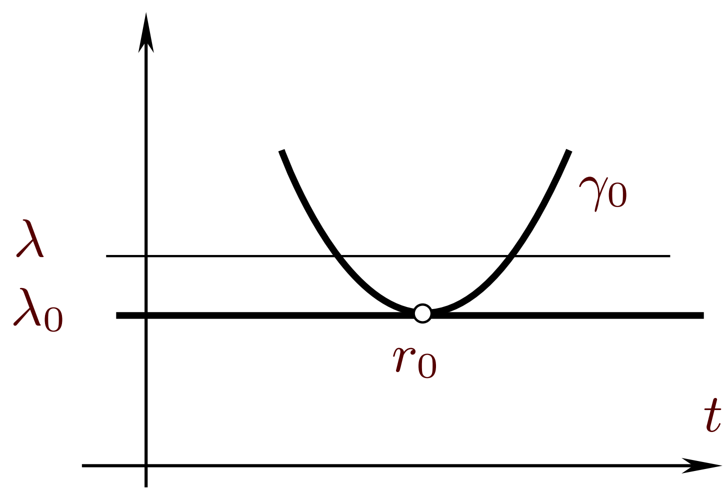
The billiard ball map T, restricted to invariant curves, is a parallel translation therein (a particular case of the Arnold-Liouville theorem).

Let λ enumerate confocal conics

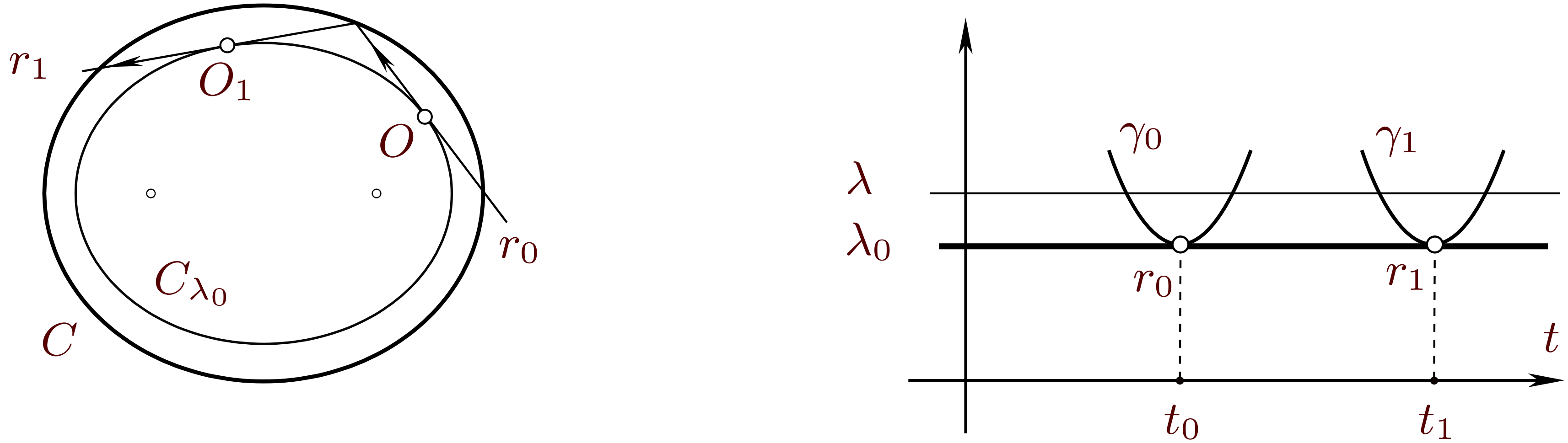
$$\frac{x^2}{a^2-\lambda}+\frac{y^2}{b^2-\lambda}=1,$$

and t be the "Arnold-Liouville" coordinate on these conics. Then (t,λ) are local coordinates in the phase space, and γ_0 is the "line" through point O.





Claim: $r_1 := T^n(r_0)$ *is an inflection point of the curve* $T^n(\gamma_0)$*, i.e., the 2-jets of* γ_1 *and* $T^n(\gamma_0)$ *coincide.*



One has: $\gamma_0 \underset{r_0}{\sim} \{(t_0 + \varepsilon, \lambda_0 + a\varepsilon^2)\}, \quad \gamma_1 \underset{r_1}{\sim} \{(t_1 + \delta, \lambda_0 + a\delta^2)\}$ because the difference of the t-coordinates of the intersections of these two curves with a horizontal line are equal.

Next, $T^n(t, \lambda) = (t + c(\lambda), \lambda)$, hence

$$\begin{aligned} T(\gamma_0) &\underset{r_1}{\sim} \{(t_0 + \varepsilon + c(\lambda_0 + a\varepsilon^2), \lambda_0 + a\varepsilon^2)\} \\ &\underset{r_1}{\sim} \{(t_0 + \varepsilon + c(\lambda_0) + ac'(\lambda_0)\varepsilon^2, \lambda_0 + a\varepsilon^2)\} \\ &= \{(t_1 + \varepsilon + ac'(\lambda_0)\varepsilon^2, \lambda_0 + a\varepsilon^2)\}. \end{aligned}$$

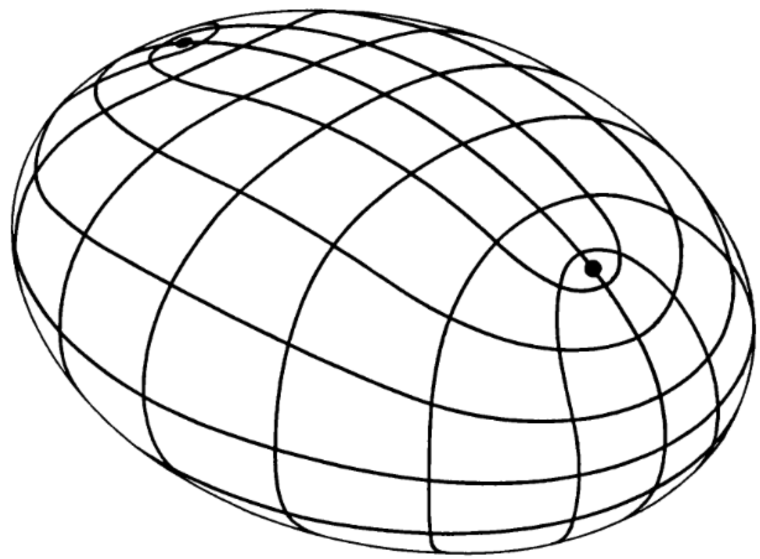
Set $\delta = \varepsilon + ac'(\lambda_0)\varepsilon^2$. Then

$$\begin{aligned} \gamma_1 &\underset{r_1}{\sim} \{(t_1 + \delta, \lambda_0 + a\delta^2)\} \\ &\underset{r_1}{\sim} \{(t_1 + \varepsilon + ac'(\lambda_0)\varepsilon^2, \lambda_0 + a(\varepsilon + ac'(\lambda_0)\varepsilon^2)^2)\} \\ &\underset{r_1}{\sim} \{(t_1 + \varepsilon + ac'(\lambda_0)\varepsilon^2, \lambda_0 + a\varepsilon^2)\} \underset{r_1}{\sim} T(\gamma_0), \end{aligned}$$

as needed.

Generalization: Liouville billiards

Liouville metric: $(f(x)+g(y))(dx^2+dy^2)$; the coordinate lines form a *Liouville net*. Example:



Liouville billiards are completely integrable, and a version of the above proof applies.

Part 2: a 4-point theorem

Theorem 4 [W. Graustein 1936]: *The average curvature of a plane oval is attained at least at four points.*

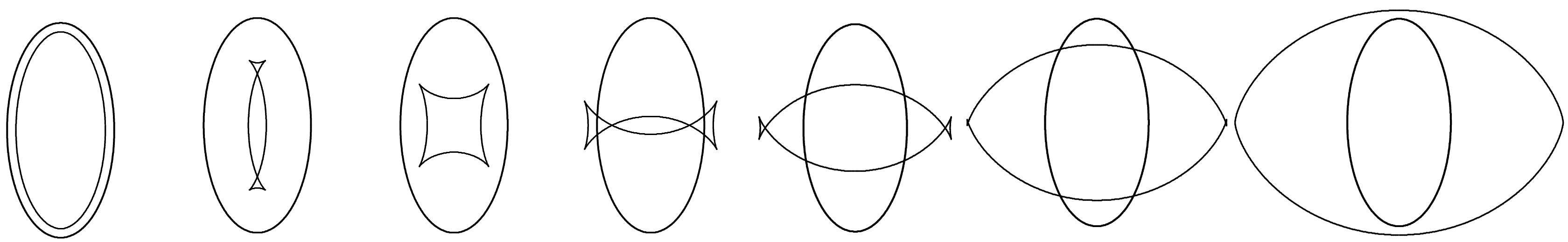
This, of course, implies the 4-vertex theorem.

Theorem 5 [ST 2024]: *The same holds for a closed strictly convex spherical curve and for a closed* horocyclically convex *curve in the hyperbolic plane.*

Recall that a curve of constant curvature in H^2 is a circle if $k > 1$, a horocycle if $k = 1$, and an equidistant (from a straight line) curve if $k < 1$.

Sketch of a proof

Let $\gamma(s) = \gamma_0(s)$ be the initial oval and $\gamma_t(s)$ be the equidistant family of curves. That is, γ_0 is the source of light, and γ_t is the time-t wave front.



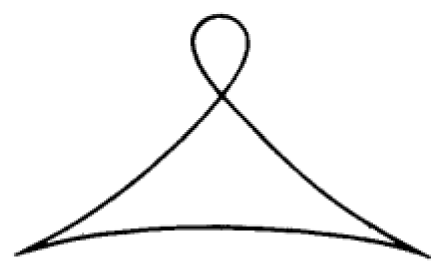
In the Euclidean geometry,

$$R_t(s) = R(s)+t,\ k_t(s) = \frac{k(s)}{1+tk(s)},\ L_t = L+2\pi t,\ A_t = A+Lt+\pi t^2$$

(Steiner's formulas), and if $k(s) = \bar{k}$, then $k_t(s) = \bar{k}_t$ for all t.

Let $t = -\frac{1}{\bar{k}} = -\frac{L}{2\pi}$. Then $L_t = 0$, and $k(s) = \bar{k} = \frac{2\pi}{L}$ iff $k_t(s) = \infty$, that is, $\gamma_t(s)$ is a cusp.

Assume that γ_t has only two cusps (which exist "for free"):



Then $L_t \neq 0$, a contradiction.

In the spherical geometry:

$$R_t(s) = R(s) + t, \quad k_t(s) = \cot R_t(s) = \cot(R(s) + t),$$

$$L_t = L\cos t + (2\pi - A)\sin t, \quad A_t = 2\pi + L\sin t - (2\pi - A)\cos t,$$

and if $k(s) = \bar{k}$, then $k_t(s) = \bar{k}_t$ for all t.

Let $t = \tan^{-1}\left(\frac{2\pi - A}{L}\right)$. Then $A_t = 2\pi$, and $\bar{k}_t = \frac{2\pi - A_t}{L_t} = 0$ by Gauss-Bonnet. We need to show that γ_t has four inflections.

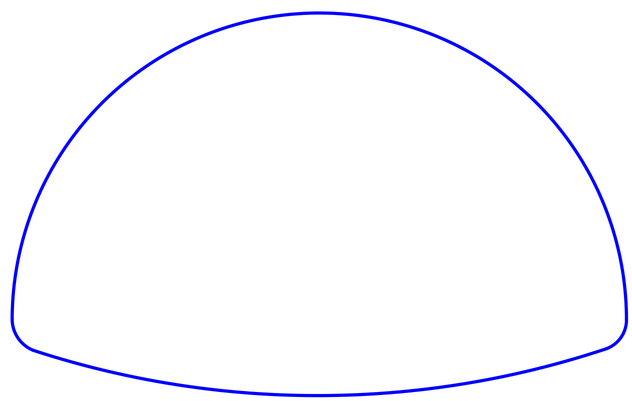
This is due to Arnold's *tennis ball theorem*:

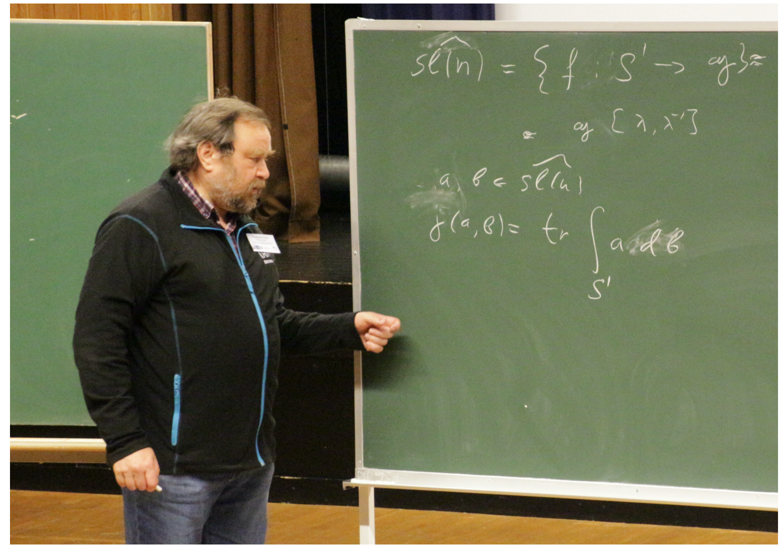
In the hyperbolic geometry:

$$R_t(s) = R(s) + t, \quad k_t(s) = \coth R_t(s) = \coth(R(s) + t),$$
$$L_t = L\cosh t + (2\pi + A)\sinh t, \quad A_t = -2\pi + L\sinh t + (2\pi + A)\cosh t.$$

Let $t = -\coth^{-1}\left(\frac{2\pi+A}{L}\right)$, then $L_t = 0$, and the argument is similar to the Euclidean case.

And a *counter-example*, a "fattened" half-circle ($R > 1.386$):



Valya and Volodya,

You two, combined, have reached the Biblical 120.

Wish each of you to reach it individually!