OPEN ACCESS
IOP Publishing | London Mathematical Society
Nonlinearity
Nonlinearity **37** (2024) 025006 (21pp)
https://doi.org/10.1088/1361-6544/ad1496

Cusps in heavy billiards

Boris Hasselblatt[1], Ki Yeun Kim[1] and Mark Levi[2,3,*]

[1] Department of Mathematics, Tufts University, Medford, MA 02144, United States of America
[2] Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, United States of America

E-mail: mxl48@psu.edu

Received 7 March 2023; revised 27 October 2023
Accepted for publication 12 December 2023
Published 4 January 2024

Recommended by Dr Richard Montgomery

CrossMark

Abstract

We consider billiards with cusps and with gravity pulling the particle into the cusp. We discover an adiabatic invariant in this context; it turns out that the invariant is in form almost identical to the Clairaut integral (angular momentum) for surfaces of revolution. We also approximate the bouncing motion of a particle near a cusp by smooth motion governed by a differential equation—which turns out to be identical to the differential equation governing geodesic motion on a surface of revolution. We also show that even in the presence of gravity pulling into a cusp of a billiard table, only the direct-hit orbit reaches the tip of the cusp. Finally, we provide an estimate of the maximal depth to which a particle penetrates the cusp before being ejected from it.

Keywords: billiards, adiabatic invariant, cusp
Mathematics Subject Classification numbers: 37, 70

1. Introduction

In a *gravity-free billiard system* a particle moves freely in a planar domain and is reflected in the boundary so that the angle of incidence equals the angle of reflection. Convex, polygonal, and dispersing billiards give rise to rather different dynamical phenomena, and we focus on a specific question pertaining to the latter class: can a billiard orbit reach a (finite dispersing)

[3] ML was partially supported by NSF Grant DMS-9704554.
[*] Author to whom any correspondence should be addressed.

cusp in the defining region of the billiard in any way other way than along a straight line into it? The answer is negative: the tip of a cusp can only be reached in a straight shot; any collision with the boundary means ejection after finitely many collisions. This problem was presented by King [13] for the infinite cusp bounded by $y = \pm 1/x$. In this note we consider billiards with gravity, and ask whether gravity can pull the particle into the cusp so that it limits at the tip of the cusp.

We show that the cusp behaves as a repeller so strong that every motion that is not the straight shot (figure 1, right) is ejected from the cusp before any future visits, if any. This ejection happens no matter how sharp the cusp is—even if the order of contact of the two sides is infinite, and no matter how strong the gravity is. By contrast, there are examples of a billiard trajectory near a concave wall undergoing infinitely many collisions in finite time, see [9].

2. Results

In this note we consider billiards with gravity as in figure 1. The billiard region includes a cusp-shaped domain

$$\mathcal{C} := \{(x,y) \in \mathbb{R} \times [0,1] \mid -g_-(y) \leqslant x \leqslant g_+(y)\}$$

with $g = g_\pm \colon [0,1] \to [0,\infty)$ satisfying

$$g(0) = g'(0) = 0, \quad g''(y) \geqslant 0. \tag{1}$$

Gravity produces acceleration $a = \text{const.}$ pointing into the cusp, in the direction tangent at the cusp, as shown in figure 1.

Our main results are:

(1) an observation on the existence of an adiabatic invariant[4] associated with a cusp (section 2.1);
(2) a close connection between billiard in the cusp and geodesics on a surface of revolution: the adiabatic invariant is an almost identical ‘twin’ of the Clairaut integral, and a smooth approximation to the billiard is identical to the equation for the geodesics (section 2.2);
(3) an estimate of the penetration depth into a cusp using the adiabatic invariant (section 2.3);
(4) the ejection property of cusps with gravity (section 2.4, theorem 4).

In the rest of this section we summarise the above items 1–4. Details and proofs are in the following sections.

2.1. Adiabatic invariant of the cusp

Consider a particle bouncing near the tip of the cusp as in figure 2. As long as the velocity is not ‘too vertical’, collisions happen in rapid succession. This will allow us to approximate the

[4] The idea of adiabatic invariants dates back at least as far as 1905, when, before the discovery of quantum mechanics, there were attempts to explain why the ratio of energy to frequency of an atom’s radiation is always (Planck’s) constant. An explanation offered by Einstein went like this: an electron in an atom (then thought of as a miniature Solar System held together by electric forces) is orbiting around the nucleus so fast that the ambient electromagnetic forces seem to change slowly. Loosely speaking, the atom is similar to a pendulum whose string slowly changes length. In general, the energy and the frequency in a pendulum are unrelated. It was discovered by Einstein that if the length of a pendulum changes slowly, the frequency becomes tied to the energy; in fact, the ratio of the two becomes almost constant. This ratio is an example of what became known as an *adiabatic invariant*.

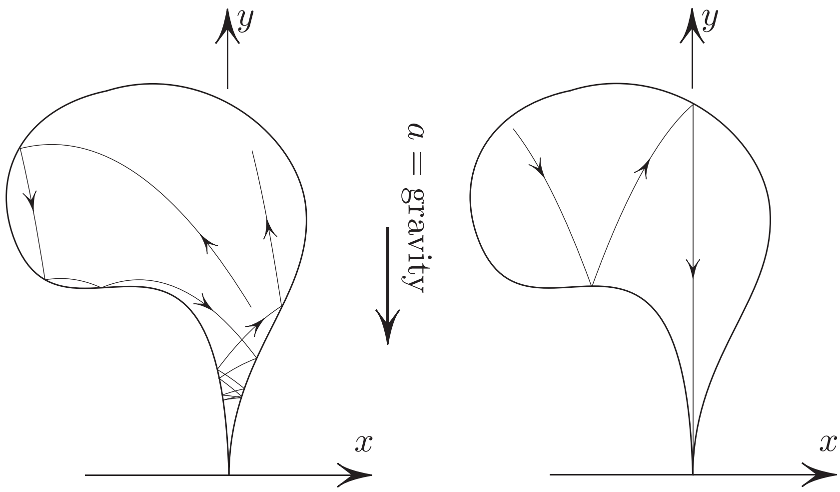

Figure 1. Left: a typical trajectory in a heavy billiard. Right: an exceptional trajectory that falls into the cusp.

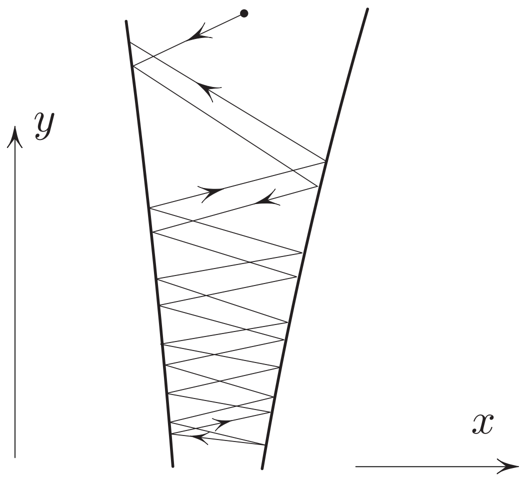

Figure 2. A magnified view of a trajectory bouncing in a small neighbourhood of the cusp. Parabola's segments appear straight in this magnification.

discrete process with a continuous one (the details are in section 3). Each collision results in some increase in the y-component of the particle's velocity. This velocity increase averaged over the short inter-collision time amounts to a steady upward force. In section 3 we show that this force is given to the leading order of accuracy by

$$\ddot{y} = v^2 \frac{g'(y)}{g(y)} - a, \quad \text{where} \;\; v^2 = 2(E - ay) - \dot{y}^2 \tag{2}$$

where E is the energy (kinetic plus potential) of the particle and where g defines the cusp: its walls are the graphs of $x = \pm g(y)$. Solutions of this differential equation approximate vertical motions of a billiard particle in a sufficiently small neighbourhood of the tip of the cusp. Note that $v = v(y, \dot{y})$ in (2) approximates the x-velocity $v_{\rm hor}$ of the actual billiard particle. Indeed, for the bouncing particle with energy E we have

$$\frac{v_{\rm hor}^2}{2} + \frac{v_{\rm vert}^2}{2} + ay = E, \tag{3}$$

so that

$$v_{\rm hor}^2 = 2\,(E - ay) - v_{\rm vert}^2.$$

Therefore, if $v_{\rm vert} \approx \dot{y}$ then $v^2 \approx v_{\rm hor}^2$.

A conserved quantity of (2). A direct computation[5] shows that (2) possesses a conserved quantity[6], namely

$$I = g\,(y)\,v\,(y, \dot{y}) = g\,(y)\,\sqrt{2\,(E - ay) - \dot{y}^2}. \tag{4}$$

While I is an exact integral of (2), it is an approximate integral of the billiard motion. This allows us to estimate the depth to which a particle can penetrate the cusp: at the closest approach to the cusp, $v_{\rm H} \approx v = \sqrt{2(E - ay)}$, the speed, so $I(0) = g(y_0)(v_{\rm H})_0 \approx g(y_{\min})v = g(y_{\min})\sqrt{2(E - ay_{\min})}$; this implicitly determines $y_{\min}$. We return to this later.

2.1.2. A geometrical interpretation of I. is illustrated by figure 3: I approximates $g(y)v_{\rm hor}$, the action (phase area of a frozen closed orbit) in the phase plane $\{x\dot{x}\}$ of horizontal motion. This coincides with the adiabatic invariant of the Fermi–Ulam 'ping-pong' [8, 22]: a particle on the line bouncing between two slowly (compared to the particle's speed) moving walls. In our billiard problem the 'moving walls' are due to the particle's descent or ascent. By analogy with (4), the product of speed and the distance between the walls is an adiabatic invariant of Fermi–Ulam's 'ping-pong'. Geometrically, this is the area in the phase plane enclosed by a trajectory of the 'frozen' system, i.e. stationary walls. In the case of cusp, 'frozen' would amount to parallel walls; and near the tip of the cusp the walls are nearly parallel.

For a billiard particle, the deeper it descends into the cusp, i.e. the more it is 'compressed' horizontally, the greater is $v_{\rm hor}$, according to (4). This is closely related to the adiabatic process in ideal gas: without heat exchange, decrease in volume raises temperature.

We do not address the problem of rigorous estimates on the change of the adiabatic invariant in the neighbourhood of the cusp.

2.2. Adiabatic invariant and Clairaut's integral

We describe now a connection between (2) and geodesics on a surface of revolution. It will turn out that the adiabatic invariant of the billiard has the same expression as Clairaut's integral ([7, p 257], [20, proposition 9.3.2]), i.e. the angular momentum of the particle on the surface of revolution, and that the differential equation for the 'smoothed' motion of a billiard is the same as the equation (2) describing the longitudinal motion of a geodesic on a surface of revolution.

[5] $2v\dot{v} = \frac{d}{dt}v^2 = \frac{d}{dt}[2(E - ay) - \dot{y}^2] = -2\dot{y}[\ddot{y} + a] \overset{(2)}{=\!=} -2v^2\dot{y}\frac{g'}{g}$ implies $\frac{dI}{dt} = vg'\dot{y} + \dot{v}g = 0$.

[6] The existence of a conserved quantity may seem surprising at first glance since the presence of $\dot{y}$ in the right-hand side of (2) may suggest dissipation.

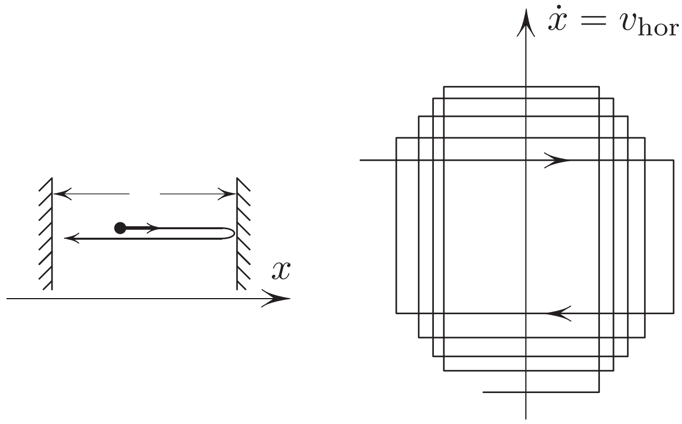

Figure 3. Adiabatic invariant of the billiard in the cusp is the product of the width and the horizontal velocity.

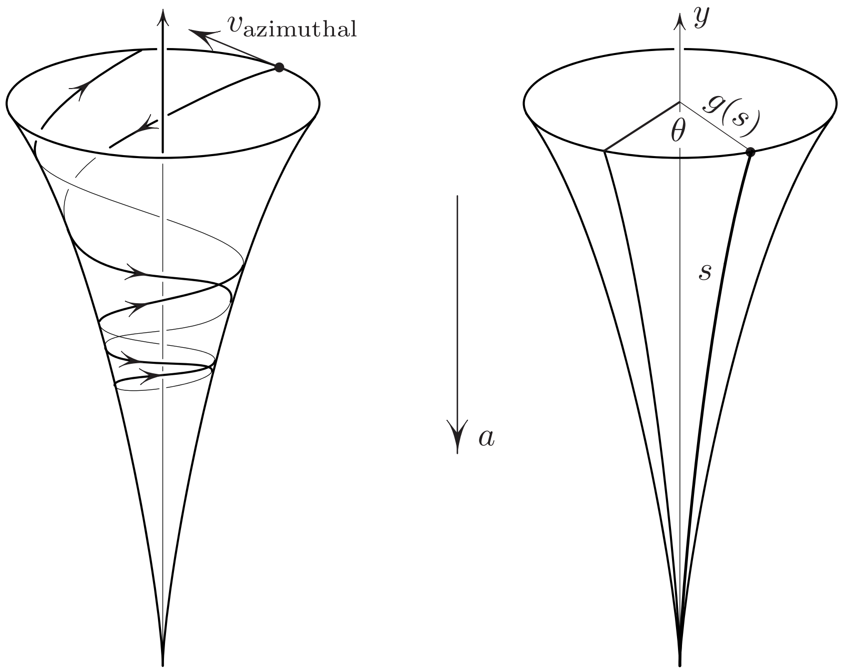

Figure 4. The averaged acceleration in (2) due to collisions is identical to the s-acceleration $\ddot{s}$ of a particle on a surface of revolution. And the adiabatic invariant (4) of the billiard in a cusp is exactly of the form (7) of the angular momentum, i.e. of Clairaut's integral.

Along with the cusped billiard let us consider the classical problem: a point mass on a surface of revolution with gravity, figure 4. The surface of revolution is given by the same function g as the cusp, except that we use the arc length s distance from the cusp along a meridian. Since $s = y + o(y)$ (due to $g'(0) = 0$), this surface is closely approximated by the surface of revolution of $x = g(y)$ near the tip of the cusp. The Lagrangian for the point mass is

$$L\left(s, \dot{s}, \theta, \dot{\theta}\right) = \frac{1}{2}\left(\left(g(s)\dot{\theta}\right)^2 + \dot{s}\right) - ay,$$

where $y = y(s) = \int_0^s \sqrt{1 - g'(\sigma)}\mathrm{d}\sigma = s + o(s)$. The Euler–Lagrange equations for s and θ are

$$\begin{cases} \ddot{s} - gg'\dot{\theta}^2 + a = o(s), \\ \frac{\mathrm{d}}{\mathrm{d}t}\left(g^2\dot{\theta}\right) = 0, \end{cases} \tag{5}$$

the latter equation expressing the conservation of the angular momentum a.k.a. Clairaut's integral $M = g^2\dot{\theta}$. We are interested in all motions with an (arbitrarily) fixed energy E:

$$\frac{1}{2}\left(\left(g(s)\dot{\theta}\right)^2 + \dot{s}^2\right) + ay = E.$$

Solving this for $\dot{\theta}$ and substituting in the first equation of (5) yields the differential equation for the motions with energy E:

$$\ddot{s} = \left(2E - \dot{s}^2 - as\right)\frac{g'}{g} - a, \tag{6}$$

where we replaced $y = s + o(s)$ with s. This equation is identical to the billiard equation (2)!

The connection between the billiard and the particle on a surface of revolution extends further. Namely, the seemingly mysterious conserved billiard quantity (4) is identical to the angular momentum M. Indeed,

$$M = g^2\dot{\theta} = g\left(g\dot{\theta}\right) = gv_{\text{azimuthal}} = \text{const.}; \tag{7}$$

here $v_{\text{azimuthal}}$ (shown in figure 4) has the same expression as v in (4): indeed,

$$\frac{1}{2}\left(v_{\text{azimuthal}}^2 + \dot{s}^2\right) + ay = E$$

gives

$$v_{\text{azimuthal}}^2 = 2(E - ay) - \dot{s}^2,$$

same as in (2) (modulo replacing y with $s = y + o(y)$).

The bouncing particle's motion is therefore approximated by projecting the smooth motion on the particle on the surface onto a plane through the axis of revolution.

Summarising, the billiard's approximate equation (2) is the same as the s-equation (6) for the geodesic; and the expression of the conserved quantity (4) of the billiard equation is identical to Clairaut's integral M.

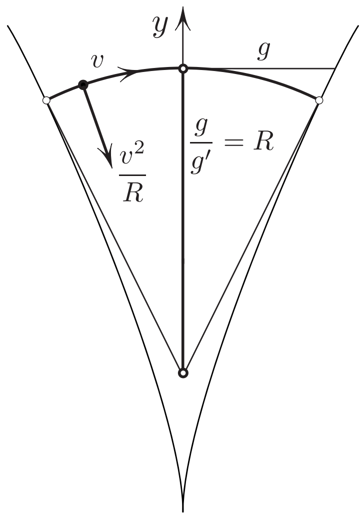

Figure 5. The main term $v^2\frac{g'}{g} = \frac{v^2}{R}$ $(R = g'/g)$ is the centripetal force acting on the point unit mass constrained to the circle and moving with speed v.

Remark 1. The repelling term $v^2\frac{g'}{g}$ in (2) has the following physical interpretation. Consider an auxiliary problem: a particle constrained to the circular arc normal figure 5. The particle slides back and forth along the arc, bouncing off the two walls. Neglecting gravity, the particle moves with constant speed v which we take to be the same as in (2). The particle applies the outward centrifugal force v^2/R to the arc. Since the arc is perpendicular to the walls, no additional force normal to the circle is generated by the particle in collisions. Now since the circle's radius $R = g/g'$ we have:

$$\frac{v^2}{R} = v^2\frac{g'}{g},$$

showing that the first term in (2) equals the centrifugal force of figure 5! This term therefore equals the force of the imagined constraint. This observation suggests a way to guess the form of (2) by a thought experiment without calculations, as described in the following remark.

Remark 2 (a heuristic derivation of (2)). The thought experiment of constraining the particle suggests that removing the constraint (thus reverting to the original problem) should amount to applying the force opposite to the constraint, i.e. the force $v^2\frac{g'}{g}$ away from the tip of the cusp. This is exactly what (2) states!

One of the gaps in this heuristic argument is the use of the implicit assumption that the velocity is nearly tangent to the circle. We do not fill this gap in the interest of brevity.

Remark 3 (an observation on Clairaut's problem). The equation (6) governing the geodesic motion can be rewritten in a remarkably simple form:

$$\ddot{s} = k_g v_{\mathrm{azimuthal}}^2, \tag{8}$$

where k_g is the *geodesic curvature of the meridian* through the particle's location.

For the proof we only need to show that

$$\frac{g'(s)}{g(s)} = k_g.$$

To that end, consider the cone tangent to the surface at a meridian $s = \text{const.}$, figure 4. The length R of the generator of this cone, i.e. the distance from the vertex to the meridian, is given by $R = \frac{g}{\sin\alpha}$ where α is the angle between the axis and a generator. But $\sin\alpha = g'$, so that $R = \frac{g}{g'}$. It remains to cut the cone along a generator and unroll it to the plane, thus turning the meridian into an arc of a circle of radius R and thus of curvature $R^{-1} = \frac{g'}{g}$. And since unrolling is an isometry, $k_g = R^{-1} = \frac{g'}{g}$ is indeed the geodesic curvature of the meridian.

It remains to observe that (8) could have been guessed without the argument of the preceding paragraph by imitating the heuristic argument of remark 2, namely by observing that restricting the particle to the meridian will generate the reaction force $k_g v^2_{\text{azimuthal}}$.

2.3. *Penetration depth*

We consider first billiards without gravity; in a small vicinity of the cusp gravity's role is small since parabolic segments are short and thus appear almost straight. In the absence of gravity we can pick any energy value, say $E = \frac{1}{2}$, so that the particle's speed $v = 1$. The adiabatic invariant (4) takes form

$$I = g\cos\alpha,$$

where α is the angle between the trajectory and the horizontal. Thus I has a remarkably simple geometrical interpretation (figure 6): it is *the projection of the width g onto the trajectory segment that cuts through this width*. Figure 6 illustrates a geometrical way to estimate maximal penetration depth: projecting the width g_0 onto the trajectory segment AB gives the width of the cusp at the maximal penetration depth: $g_0\cos\alpha \approx g_{\min}\cos 0$ since $\alpha \approx 0$ at the deepest penetration, and we have

$$g_{\min} \approx g_0\cos\alpha.$$

This approximation gets worse as α gets too close to $\pi/2$ and it gets better the deeper A is in the cusp. We do not address error estimates in these approximations.

Actually the same estimate holds for billiards with gravity since parabolic segments are nearly straight deep in the cusp; in this case α can be taken as the angle between the horizontal and the tangent to the rebounding trajectory at the point of impact.

2.4. *Ejection theorem*

Returning to billiards with gravity we prove that the cusp acts as an infinitely strong repeller, as anticipated in the heuristic discussion above. In particular, only a straight shot can reach the cusp. Any other trajectory may possibly approach the cusp repeatedly and arbitrarily closely, but such approaches are separated by departures from the cusp. In particular, there are no trajectories whose height $y(t) \to 0$ as t approaches a finite or infinite value. All this is so no matter how strong the gravity is and no matter how high the order of contact.

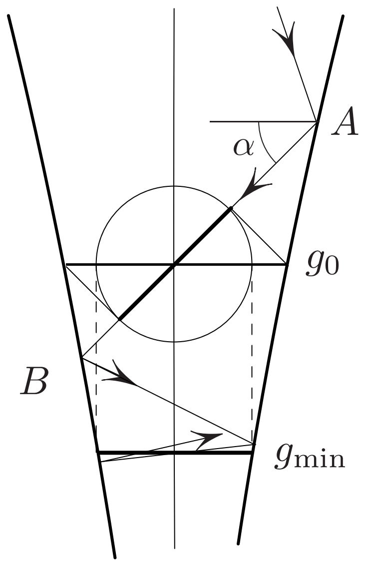

Figure 6. '$g_{\min} \approx \mathrm{proj}_{AB} g_0$': maximal width of penetration is well approximated (when deep in the cusp) by the projection of the horizontal section (e.g. g_0) onto the segment (e.g. AB) of the trajectory. In fact, *projection of any horizontal section onto the trajectory segment crossing this section is an adiabatic invariant*, i.e. is near-constant along the trajectory near the cusp provided the trajectory segments are 'not too vertical'.

2.4.1. Assumptions and notations. Recall that the billiard region includes the cusp–shaped domain

$$\mathcal{C} := \{(x,y) \in \mathbb{R} \times [0,1] \mid -g_-(y) \leqslant x \leqslant g_+(y)\}$$

We assume that $g = g_\pm \colon [0,1] \to [0,\infty)$ satisfy

$$g(0) = g'(0) = 0, \quad g''(y) \geqslant 0 \tag{9}$$

and

$$\lim_{y\to 0^+} \frac{g'}{g''} = 0; \tag{10}$$

a geometrical meaning of the last condition is explained in remark 6. In addition, the billiard is subject to constant gravity $a > 0$ in the direction of the outward (downward in figure 1) tangent vector of the cusp (we say 'the cusp is vertical'). Finally, when speaking of energy, we take potential energy at $(0,0)$ to be zero, and the mass of the particle $m = 1$, so that the total energy of a particle with speed v is $E = v^2/2 + ay$.

Theorem 4 (ejection theorem). *Under the above assumptions on g, for any fixed energy $E > 0$ there exists the height $D = D(E) \in (0,1]$ (figure 7) such any trajectory with energy E entering $\{y < D\}$ not as straight shot into the cusp tip leaves $\{y < D\}$ after a finite time, with finitely many collisions during that time, figure 7(a). Moreover, the sequence of collision heights y_k first monotonically decreases and then increases:*

$$D > y_1 > y_2 > \ldots > y_{n-1} > y_n \leqslant y_{n+1} < \ldots < y_N < D; \tag{11}$$

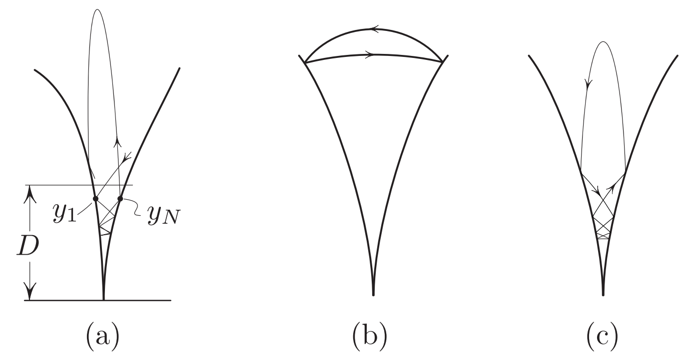

Figure 7. If a trajectory with energy E ever visits the region $\{y < D(E)\}$ and not as a direct shot into the cusp, it must leave the region after finitely many uninterrupted collisions (this does not preclude a later return). The heights of this finite sequence of collisions obey the monotonicity property (11).

here y_1 is the first collision after the trajectory descends below $y = D$ and y_N is the last collision height before the trajectory rises above $y = D$; the trajectory remains in $\{y < D\}$ between adjacent collisions.

Corollary 5. *Under the assumption of theorem 4, no trajectory other than the straight shot converges to the tip $(0,0)$ in finite or infinite time.*

Indeed, if the trajectory is not a straight shot, then convergence $y(t) \searrow 0$ is forbidden by (11).

It should be mentioned that some trajectories with energy E may never visit $\{y < D(E)\}$, figure 7(b); for such trajectories the theorem is vacuous.

Remark 6. Condition (10) is equivalent to the statement that the *centre of the osculating circle to the graph of $x = g(y)$ approaches the x-axis* (i.e. the line through the tip of the cusp and normal to the common tangent of the two walls) as $y \to 0$.

Remark 7. Despite the heuristic discussion above it may seem intuitively plausible that gravity, if strong enough, may overcome collisions in cusps with sufficiently high order of contact, causing the particle approach the cusp in a series of infinitely many collisions. This, however, does not happen: the above conditions on g allow cusps with arbitrarily high order of contact—for instance, $g(y) = e^{-1/y}$, as well as $g(y) = cy^r + O(y^{r+1})$ for any $r \geqslant 2$. Moreover, if g satisfies the hypotheses of theorem 4, then so does $h = e^{-1/g}$.

Remark 8. Theorem 4 does not exclude the possibility of a trajectory with infinitely many collisions in a vicinity of a cusp; however, these collisions must be broken into finite groups separated by exits from $\{y < D(E)\}$. In fact, by Poincaré's recurrence theorem there exist trajectories that visit any neighbourhood of the tip infinitely many times, but these visits are punctuated by departures. For such trajectories the number of collisions during a visit has no upper bound since the further into the cusp the trajectory reaches, the more collisions it will undergo before exiting.

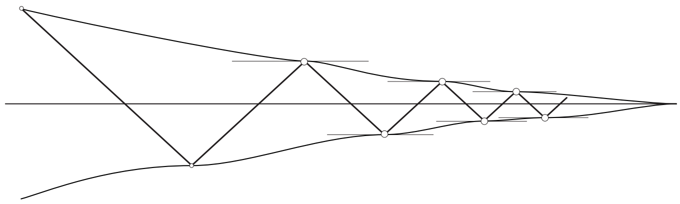

Figure 8. A trajectory with infinitely many collisions approaching the cusp for a gravity-free billiard.

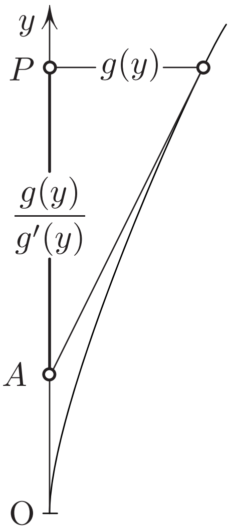

Figure 9. Proof of (12).

Remark 9. (1) Without convexity of g (or something to replace it), the conclusion of theorem 4 can fail even without gravity, as figure 8) illustrates. Parenthetically, For billiards with infinite cusps and with zero gravity there are conditions other than convexity that allow only the trivial escape to infinity [14].

(2) Convexity of g implies

$$yg'(y) > g(y) \quad \text{for all} \quad y \in [0,1]\,. \tag{12}$$

Indeed, referring to figure 9, $y = PO > PA = g(y)/g'(y)$, hence (12). In particular, $g/g' \xrightarrow[y\to 0^+]{} 0$.

Note that in other respects, gravity, no matter how weak, produces qualitative change such as the presence of periodic orbits in the cusp or the possibility of upward motion turning downwards between collisions.

Remark 10 (symmetry). For convenience we assume $g_+ = g_-$; the arguments do not depend on this assumption but the exposition is slightly simplified.

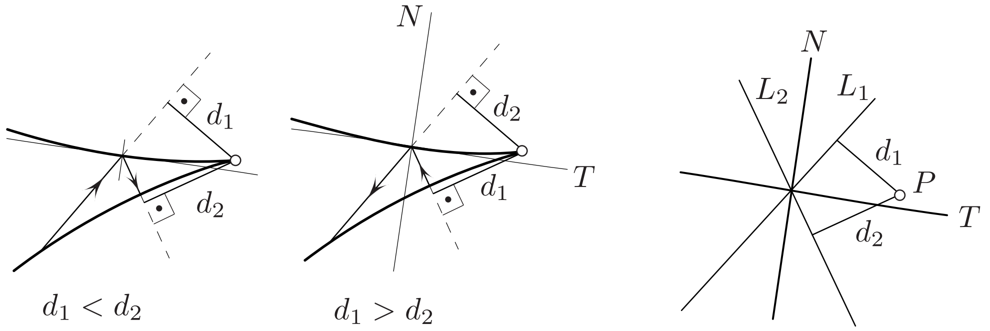

Figure 10. After an impact arriving from the outward (from the cusp) side of the normal (left) the distance increases: $d_2 > d_1$; the opposite happens when impact is from the inward of the normal.

2.5. History

As we learned from a referee, this subject is at least 60 years old. This is the age of a proof that only a straight shot (there called a trivial trajectory) goes all the way into an eventually convex infinite cusp [17]. More than three decades later, the same issue was raised in a paper by King [13], who, like us, was unaware of that early paper. For the infinite cusp bounded by $y = \pm 1/x$, David and Jacob Feldman had posed the problem whether a particle can reach infinity in a nontrivial way. Benjamin Weiss (unpublished, to our knowledge) determined that this is essentially impossible, i.e. that the set of such billiard trajectories is a null set, and indeed empty for a dispersing cusp [13, p 13]. (The argument Weiss gave for this does not use convexity. Initially unaware of Weiss' insight, King gave a direct 'bare-hands' proof that only a straight shot goes to infinity if the bounding function is convex. We note that this billiard with $y = 1$ as the third wall is ergodic [15] (likewise with a finite cusp [21]) and that cusps have also been of interest with respect to correlation decay [1–3, 5, 6, 10–12, 18, 19, 24]. Indeed, the main purpose of King's article was to bring measure theory to bear on this and related issues. Once finite invariant measures enter, ergodicity becomes a natural issue, and for a range of infinite cusps, Lenci established hyperbolicity and ergodicity of the resulting billiard map [14, 16].

Considerations of finite cusps seem to be even more a matter of folklore. David Bernstein first told us how to prove that a particle colliding with the boundary will be ejected after finitely many collisions, figure 10, [4]. This proof is based on a simple observation: Let T and N be two perpendicular lines (figure 10(left)) with two lines L_1 and L_2 symmetric to each other with respect to N and T, and let P be a point in one of the quadrants determined by T and N. Of the two skew lines the one passing through the quadrant of P (namely, L_1 in the figure) is closer to P than the other.

In figure 10(left and middle) tangent and normal lines play the roles of T and N in the above remark, the lines L_1 and L_2 are the lines of the incoming and reflected rays. If the impact is from the left of the normal (figure 10(left)), i.e. if the line of trajectory passes through the quadrant of the cusp P, then $d_1 < d_2$. Otherwise, $d_1 > d_2$, figure 10(middle). Thus the sequence d_k increases while the trajectory approaches the cusp—more precisely, while it impacts from the left of normal) and decreases on departure from the cusp.

We should also mention the work by Wojtkowski for geodesic flows of surfaces, which in particular implies a corresponding result [23, theorem 3]: only straight geodesics reach a nonpositively curved cusp (more specifically, there is a natural parametrisation of the cusp by

$(s,t) \in S^1 \times [0,1)$ such that only the geodesics $s = \text{const}$ reach the cusp). This is better known for rotationally symmetric cusps, where the Clairaut integral is the obstruction to reaching the cusp.

In summary, there are three extant arguments that establish the impossibility of reaching the cusp except by a direct hit: the simple Bernstein argument, the enhanced ergodic argument by Weiss, and King's bare-handed argument.

3. Continuum approximation for adiabatic invariant

In this section we derive the differential equation (2) whose solutions approximate the y-coordinate of billiard particles in the vicinity of the cusp. We derive (2) by 'smoothing' the reflection law which states that the incidence and reflection angles are equal. In this discussion we first take the gravity $a = 0$: in a small neighbourhood of the cusp, short parabolic segments look nearly straight.

Lemma 11 (reflection law). *In the notations of figure 11 and with gravity $a = 0$ the law of reflection at the nth collision is given by*

$$\cos\theta_{n-1} - \cos\theta_n = (\sin\theta_{n-1} + \sin\theta_n)\, g_n'. \tag{13}$$

Proof. If $g_n' = \tan\alpha_n$ so that $\mathrm{e}^{\pm\mathrm{i}\alpha_n} = \cos\alpha_n(1 \pm i g_n')$, then 'angle of incidence = angle of reflection' means that $\theta_{n-1} + \alpha_n = \theta_n - \alpha_n$, and thus

$$\mathrm{e}^{\mathrm{i}\theta_{n-1}}\mathrm{e}^{\mathrm{i}\alpha_n} = \mathrm{e}^{\mathrm{i}(\theta_{n-1}+\alpha_n)} = \mathrm{e}^{\mathrm{i}(\theta_n-\alpha_n)} = \mathrm{e}^{\mathrm{i}\theta_n}\mathrm{e}^{-\mathrm{i}\alpha_n},$$

or $\mathrm{e}^{\mathrm{i}\theta_{n-1}}(1 + i g_n') = \mathrm{e}^{\mathrm{i}\theta_n}(1 - i g_n')$. Taking the real part results in (13). $\square$

We now derive the differential equation (2). In a small vicinity of the cusp collisions happen in rapid succession, provided horizontal velocity v_{hor} is not too small. At nth collision vertical velocity v_{vert} increases instantaneously by

$$\Delta v_{\text{vert}} = v\cos\theta_{n-1} - v\cos\theta_n,$$

where v is the constant speed of the billiard particle. The time between this collision and the next one is

$$\Delta t = \frac{g_n + g_{n+1}}{v_{\text{hor}}}.$$

The increase Δv_{vert} averaged over this time gives an average acceleration:

$$\frac{\Delta v_{\text{vert}}}{\Delta t} = v\frac{\cos\theta_{n-1} - \cos\theta_n}{(g_n + g_{n+1})/v_{\text{H}}} \overset{(13)}{=\!=} v v_{\text{hor}} \frac{\left(\overbrace{\sin\theta_{n-1}}^{\approx\sin\theta_n} + \overbrace{\sin\theta_n}^{=v_{\text{hor}}/v}\right) g_n'}{g_n + \underbrace{g_{n+1}}_{\approx g_n}} \approx v_{\text{hor}}^2 \frac{g_n'}{g_n}. \tag{14}$$

In conclusion, the impulsive acceleration due to collisions is approximated by the continuously acting acceleration $v_{\text{hor}}^2 \frac{g_n'}{g_n}$.

So far this was for billiards without gravity. Now the presence of gravity $a > 0$ has two effects: (i) it adds the downward acceleration a, explaining the last term in (2), and (ii) horizontal velocity now depends on height according to the energy conservation (3); this completes the derivation of (2).

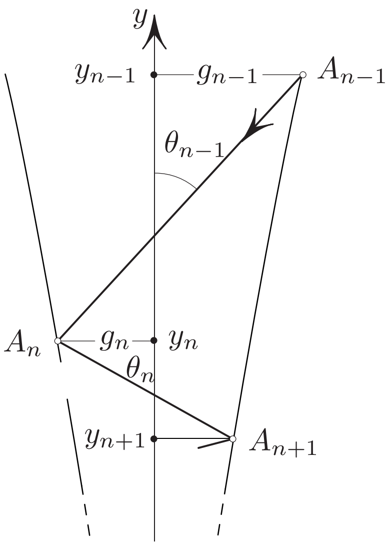

Figure 11. Derivation of the smooth approximation to the billiard motion).

4. The ejection theorem

This section is dedicated to the proof of the ejection theorem (theorem 4). Some technical arguments are deferred.

4.1. Assumptions on cusp depth

Fix $E > 0$ and let $D(E)$ satisfy the following conditions.

(1) $\{y \leqslant D\}$ is a Jacobi region, i.e. any parabolic segment with energy E lying in $\{y \leqslant D\}$ has no pairs of conjugate points.
(2) D satisfies

$$\sup_{[0,D]} g' < 1, \tag{15}$$

$$D + \sup_{[0,D]} \frac{g'}{g''}\left(1 + g'^2\right) < \frac{E}{a}, \tag{16}$$

and

$$D + \sup_{[0,D]} g < \frac{E}{a}. \tag{17}$$

We note that for any $E > 0$ the last two inequalities are satisfied for sufficiently small D thanks to the assumptions (9) and (10) of the theorem. Incidentally, E/a is the maximal height to which a particle with energy E can rise. Therefore for the weightless billiards for which $a = 0$ (16) and (17) pose no restriction on D and we can simply take $D = 1$.

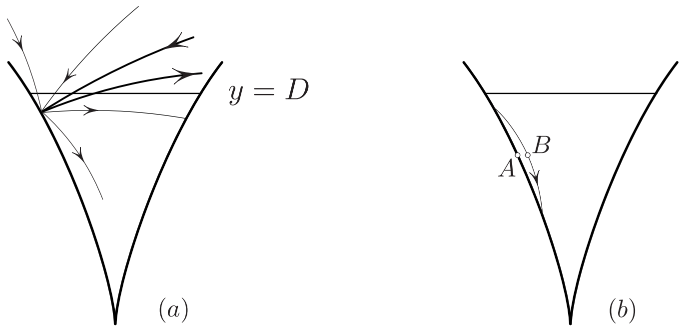

Figure 12. (a) Incoming trajectories may leave $\{y \leqslant D\}$ after one collision); (b) consecutive collisions with the same wall cannot happen if curvatures satisfy $k_B < k_A$ (unlike what is shown). Here tangents at A and B are parallel.

4.2. Proof of the ejection theorem (theorem 4)

Fixing the energy E (recall the potential energy at the tip of the cusp is taken as zero), we will choose $D = D(E)$ sufficiently small as specified in the previous subsection, and consider any non-straight-shot trajectory with energy E crossing $y = D$ downwards. If the trajectory rises above $y = D$ after one collision (as the bold trajectory in figure 12(a)), there is nothing to prove and we consider the non-trivial case of at least two collisions after descending below $y = D$, figure 12(a). Our goal is to show that the sequence of collisions (while the trajectory stays in $y < D$) is finite and that collision heights first decrease and then increase, ending with the trajectory (not necessarily the collisions!) rising above $y = D$. As a prerequisite we will establish that collisions happen with alternate walls, to exclude a possibility sketched in figure 12(b).

Lemma 12 (alternating collisions). *Fix $E > 0$. For $D = D(E)$ satisfying (16) any parabolic trajectory segment with energy E lying entirely in $\{y \leqslant D\}$ has its ends on the opposite sides of the cusp.*

Proof. Assume first that $\dot{y} \neq 0$ for the parabolic segment; to be specific, let $\dot{y} < 0$, figure 12(b). The parabolic segment then is the graph of $x = h(y)$, with two consecutive collisions at $y = y_0$ and $y = y_1$ with the same wall $x = g(y)$, so that then the horizontal distance $f = g - h$ vanishes at $y = y_0,\ y_1$ and is positive (and smooth, because of no collisions) in between. At $y = y_{\max}$ the graphs of g, h have parallel tangents and their curvatures then must satisfy

$$k_g - k_h \leqslant 0. \tag{18}$$

But in fact the opposite holds as we now show—namely that if the tangents to the trajectory and the boundary are parallel:

$$h'(y_{\max}) = g'(y_{\max}),$$

then

$$k_g - k_h > 0 \tag{19}$$

at $y = y_{\max}$. Contradiction with (18) will then show that the scenario of figure 12(b) is impossible. We have $k_h = a_\perp / v^2$ where $a_\perp$ is the component of the acceleration normal to the velocity; and at the level $y = y_{\max}$ where the tangents to the boundary and the trajectory are parallel $a_\perp = a\frac{g'}{\sqrt{1+g'^2}}$; here $g' = g'(y_{\max})$, etc. We have

$$k_h = \frac{a}{v^2}\frac{g'}{\sqrt{1+g'^2}}, \quad \text{while} \quad k_g = \frac{g''}{\left(1+g'^2\right)^{3/2}}$$

so that

$$\frac{k_h}{k_g} = \frac{a}{v^2}\frac{g'}{g''}\left(1+g'^2\right) < \frac{a}{2(E-aD)}\frac{g'}{g''}\left(1+g'^2\right) \overset{(16)}{<} 1;$$

this proves (19).

To complete the proof of lemma 12 it remains to address the possibility of y being non-monotone, i.e. of $\dot{y} = 0$ at some point $(x_{\max}, y_{\max})$ of the parabolic trajectory. The slope of the graph of this trajectory is

$$-a\frac{x - x_{\max}}{v^2},$$

where v is the horizontal velocity, which, by the conservation of energy, satisfies $v^2 \geqslant 2(E - aD)$. And since $|x - x_{\max}| < 2g(D)$ the slope of the parabolic arc inside the cusp does not exceed

$$\frac{ag(D)}{E - aD}$$

in absolute value. On the other hand, the absolute value of the slope of the boundaries of the cusp is bounded by $1/g'(D)$ from below. But two curves whose slopes are never equal may intersect at most once. Therefore it remains to make sure that

$$\frac{ag(D)}{E - aD} < \frac{1}{g'(D)}.$$

But this holds due to (15) and (17).

This completes the proof of lemma 12.

□

Returning to the trajectory whose collisions in $y < D$ we now know alternate, we consider the **maximal** n such that the collision heights are monotone decreasing[7]:

$$D > y_1 \geqslant y_2 \geqslant \ldots \geqslant y_n. \tag{20}$$

To prove the theorem we must show (i) that n is finite, and (ii) that all subsequent collisions, prior to the trajectory rising above $y = D$, are monotone increasing and finite in number.

To prove that n is finite, assume the contrary: y_k is an infinite sequence, monotone decreasing, thus leaving two alternatives:

$$\lim_{k\to\infty} y_k = 0 \tag{21}$$

or

[7] It will be clear from the proof that all inequalities but the last one will be strict.

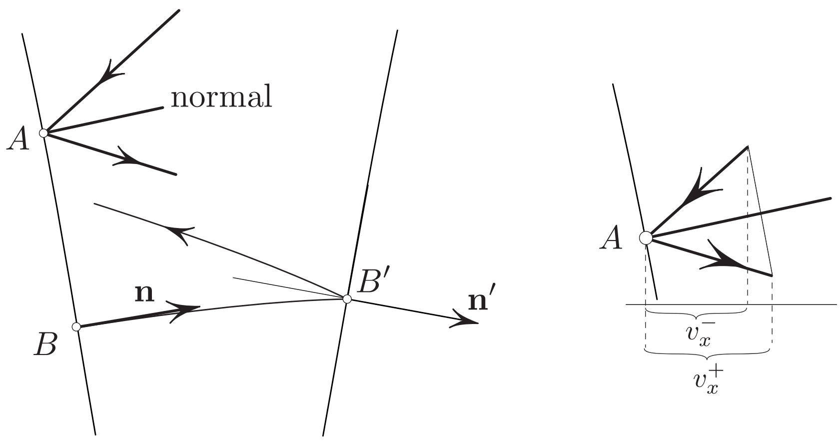

Figure 13. For a trajectory leaving B along the normal the next collision B' is above B.

$$\lim_{k\to\infty} y_k > 0. \tag{22}$$

Lemma 13. *No trajectory with energy E can have an infinite monotone decreasing set of collisions heights while staying under $y = D$, i.e. neither (21) nor (22) may hold with y_k decreasing.*

The proof is given in the next subsection.
Next we let

$$y_n \leqslant y_{n+1} \leqslant y_{n+2} \leqslant \ldots \leqslant y_N \tag{23}$$

be the *maximal* monotone increasing sequence of subsequent collisions with the trajectory staying under $y = D$ between all collisions. The ejection theorem is now a consequence of the following lemmas.

Lemma 14. *For any trajectory with energy E, subsequent collision heights satisfy*

$$y_{k+1} \geqslant y_k \text{ implies } y_{k+2} > y_{k+1}, \tag{24}$$

for as long as the trajectory stays in $y < D$ between collisions.

Lemma 15. *For a trajectory with energy $E > 0$, any monotone increasing sequence of collisions while the trajectory remains in $y < D$ terminates with the trajectory exiting $y < D$—that is, N in (23) is necessarily finite.*

Proof. Otherwise, the suprema of the collision points on the walls are a period-2 orbit, which is impossible by choice of D; see figure 13. □

Save for proving lemmas 13 and 14, this concludes the proof of the ejection theorem (theorem 4).

4.3. Proof of lemma 13

First, we show that (21) cannot hold. Assume the contrary: $y_k \searrow 0$ as $k \to \infty$.

First we observe that the downward monotonicity of y_k implies that every impact happens from above the normal as at A in figure 13. Indeed, otherwise the reflected velocity will be on or above the normal $\mathbf{n}$ as at B. But this implies $y_{k+1} > y_k$ as we now show, contradicting monotonicity. To that end we will first show that the trajectory leaving B along a normal impacts at $y_{k+1} > y_k$; and by the assumption on no conjugate points, increasing the departure slope at B will cause an increase in y_{k+1}, thus proving that the impacts indeed are as at point A in figure 13. We consider therefore a trajectory leaving B along the normal, as in figure 13 and prove that the monotonicity is violated in this case. Indeed, the trajectory has curvature $\leqslant a/v^2 < a/(2(E-aD)) = \kappa$, while the length of the arc BB' is $\leqslant 3g(y_k)$ (if D is further decreased, if necessary). The angle between the tangents to the arc at its ends B, B' is therefore

$$\angle\left(T_{B'}, T_B\right) \leqslant 3\kappa g\left(y_k\right). \tag{25}$$

On the other hand,

$$\angle\left(\mathbf{n},\, \mathbf{n}'\right) > g'\left(y_k\right). \tag{26}$$

if D is decreased further if necessary. Roughly speaking, the bend of the trajectory is negligible compared to the angle between the normals at B and B'. But

$$\angle\left(T_{B'}, T_B\right) < \angle\left(\mathbf{n},\, \mathbf{n}'\right) \tag{27}$$

by (12), provided $y < 2(E-aD)/3a$, which in turn holds if $D < 2(E-aD)/3a$, i.e. if $D < \frac{2}{5}\frac{E}{a}$, which we add to our assumptions on D.

Having shown that all impacts at y_k are from above the normal (as at point A in figure 13) we now show that the cumulative upward kick in velocity from infinitely many collisions is infinite, even accounting for the downward gravity, contradicting $y_k \downarrow 0$. A key observation is that the horizontal velocity increases after each impact. The reason is illustrated by the collision at A in Figure 13, right: for the pre-and post-impact velocities satisfy $|\mathbf{v}_-| = |\mathbf{v}_+|$, while $|\mathrm{slope}(\mathbf{v}_-)| < |\mathrm{slope}(\mathbf{v}_+)|$. Putting it differently, if $\alpha \in (0, \pi/2)$ is the angle between the normal and the incoming and reflected velocity, and if $\beta \in (0, \pi/2)$ is the angle between the normal and the horizontal, we have, with v denoting speed just before and just after impact:

$$v_x^+ = v\cos\left(\alpha - \beta\right) > v\cos\left(\alpha + \beta\right) = v_x^-,$$

as claimed. We thus showed that for the descending trajectory

$$|\dot{x}| > v_1, \tag{28}$$

for all time after the first collision, where $v_1 > 0$ is the horizontal speed after the first collision at y_1. We now use this to show that each collision adds upward kick in velocity estimated thus:

$$\dot{y}\left(t_k^+\right) - \dot{y}\left(t_k^-\right) \geqslant Cg'\left(x_k\right), \tag{29}$$

where $C > 0$ depends only on the initial condition but not on k. Indeed, according to the reflection law,

$$\dot{y}\left(t_k^+\right) - \dot{y}\left(t_k^-\right) = 2v\cos\alpha n_y, \tag{30}$$

where $v = \sqrt{2(E-ay_k)}$ is the speed prior and after the impact; where α is the angle between the normal and the horizontal and where n_y is the y-component of the inward unit normal to the boundary: $n_y = g'/\sqrt{1+(g')^2}$ (evaluated at $y-y_k$). Thanks to (28) we have $\cos\alpha \geqslant c > 0$ for some c independent of k; substituting the above estimates into (30) results in (29). On the other hand, between two collisions at $t = t_k$ and $t = t_{k+1}$ vertical velocity acquires a negative addition:

$$\dot{y}\left(t_{k+1}^{-}\right) - \dot{y}\left(t_k^{+}\right) = -a\left(t_{k+1} - t_k\right).$$

Fortunately this downward speed-up is dominated by the instantaneous increase (29) upwards. Indeed, thanks to (28),

$$t_{k+1} - t_k < \frac{2g\left(x_k\right)}{v_1},$$

so that

$$\dot{y}\left(t_k^{+}\right) - \dot{y}\left(t_{k+1}^{-}\right) < \frac{2ag\left(x_k\right)}{v_1}. \tag{31}$$

Subtracting (31) from (29) we have

$$\dot{y}\left(t_{k+1}^{-}\right) - \dot{y}\left(t_k^{-}\right) \geqslant Cg'\left(x_k\right) - \frac{2ag\left(x_k\right)}{v_1} = Cg'\left(x_k\right)\left(1 - 2\frac{ag\left(x_k\right)}{Cv_1}\right) > \frac{C}{2}g'\left(x_k\right), \tag{32}$$

for k sufficiently large. The last key point is the lower bound

$$g'\left(x_k\right) > C_1\frac{y_k - y_{k+1}}{y_k}, \tag{33}$$

where $C_1 > 0$ is independent of k for sufficiently large k. Indeed, due to (28) the absolute value of the slope of the parabolic arc is bounded by a constant c_1, and therefore

$$y_k - y_{k+1} \leqslant c_2 g\left(x_k\right) \tag{34}$$

for some $c_2 > 0$ independent of k. Together with $g(x_k) < y_k g'(x_k)$ this implies (33).

It remains to show that the sum of added upward velocity increases is infinite. We have

$$\dot{y}\left(t_{N+1}^{-}\right) - \dot{y}\left(t_1^{-}\right) = \sum_{k=1}^{N}\dot{y}\left(t_{k+1}^{-}\right) - \dot{y}\left(t_k^{-}\right) \overset{(32)}{\geqslant} \frac{C}{2}\sum_{k=1}^{N} g\left(x_k\right) \overset{(32)}{\geqslant} \frac{CC_1}{2}\sum_{k=1}^{N}\frac{y_k - y_{k+1}}{y_k}.$$

But

$$\sum_{k=1}^{N}\frac{y_k - y_{k+1}}{y_k} \geqslant \sum_{k=1}^{N}\int_{y_{k+1}}^{y_k}\frac{dy}{y} = \ln y_0 - \ln \underbrace{y_{N+1}}_{\to 0^+} \xrightarrow[N\to\infty]{} \infty.$$

We thus showed that $\dot{y}(t_n) \to \infty$ as $n \to \infty$, which contradicts the monotone descent into the cusp. We thus eliminated the possibility (21).

To complete the proof of lemma 13 it remains to eliminate the possibility (22). To that end we observe that $y_\infty = \lim y_k > 0$ is the endpoint of a period 2 orbit bouncing between the two walls, in the region $\{y < D\}$; but this is impossible. Indeed, such a trajectory would impact the boundary at the right angle, which is impossible as we showed earlier when referring to figure 13. This completes the proof of lemma 13.

Proof of lemma 14. Consider first the extreme case: $y_k = y_{k+1}$, figure 14. The estimate (27) in the proof of lemma 13 shows that the impact at y_{k+1} is below the normal. Fixing $y_{k+1} < D$ arbitrarily, we now decrease y_k, i.e. shoot at y_{k+1} (that has been fixed) from the new point y_k';

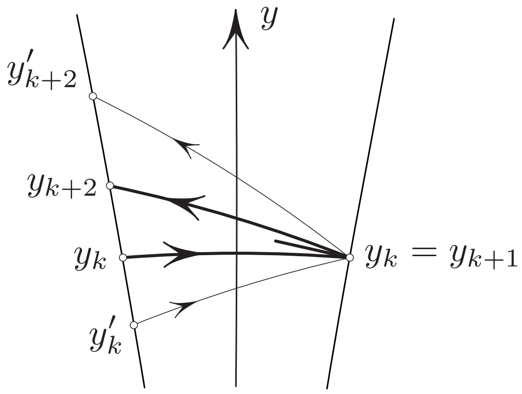

Figure 14. Proof of lemma 14. Points are labelled by their y-coordinates.

since all the trajectories are in the Jacobi region, the next impact y'_{k+2} will be above y'_{k+2}. This completes the proof of the lemma.

Proof of lemma 15. Assume the contrary. This leaves two possibilities: the sequence is infinite, monotone increasing—but that implies that the limit is a period 2 periodic orbit, which is impossible as proven in the proof of lemma 13. The remaining alternative is the failure of monotonicity. But this contradicts lemma 14 and completes the proof of lemma 15.

This also completes the proof of the ejection theorem 4.

References

[1] Bálint P, Chernov N and Dolgopyat D 2011 Limit theorems for dispersing billiards with cusps *Commun. Math. Phys.* **308** 479–510
[2] Bálint P, Chernov N and Dolgopyat D 2017 Convergence of moments for dispersing billiards with cusps *Dynamical Systems, Ergodic Theory and Probability: In Memory of Kolya Chernov* (*Contemporary Mathematics* vol 698) (American Mathematical Society) pp 35–67
[3] Bálint P and Melbourne I 2008 Decay of correlations and invariance principles for dispersing billiards with cusps and related planar billiard flows *J. Stat. Phys.* **133** 435–47
[4] Chernov N and Markarian R 2006 *Chaotic Billiards* (*Mathematical Surveys and Monographs* vol 127) (American Mathematical Society)
[5] Chernov N and Markarian R 2007 Dispersing billiards with cusps: slow decay of correlations *Commun. Math. Phys.* **270** 727–58
[6] Chernov N and Zhang H K 2008 Improved estimates for correlations in billiards *Commun. Math. Phys.* **277** 305–21
[7] do Carmo M P 1976 *Differential Geometry of Curves and Surfaces* (Prentice-Hall, Inc.) (Portuguese transl.)
[8] Fermi E 1949 On the origin of the cosmic radiation *Phys. Rev.* **75** 1169–74
[9] Halpern B 1977 Strange billiard tables *Trans. Am. Math. Soc.* **232** 297–305
[10] Jung P, Melbourne I, Pène F, Varandas P and Zhang H-K 2021 Necessary and sufficient condition for $\mathcal{M}_2$-convergence to a Lévy process for billiards with cusps at flat points *Stoch. Dyn.* **21** 2150024
[11] Jung P, Pène F and Zhang H-K 2020 Convergence to α-stable Lévy motion for chaotic billiards with several cusps at flat points *Nonlinearity* **33** 807–39
[12] Jung P and Zhang H-K 2018 Stable laws for chaotic billiards with cusps at flat points *Ann. Henri Poincare* **19** 3815–53
[13] King J L 1995 Billiards inside a cusp *Math. Intell.* **17** 8–16
[14] Lenci M 1996 Escape orbits for non-compact flat billiards *Chaos* **6** 428–31
[15] Lenci M 2002 Semi-dispersing billiards with an infinite cusp. I *Commun. Math. Phys.* **230** 133–80

[16] Lenci M 2003 Semidispersing billiards with an infinite cusp. II *Chaos* **13** 105–11
[17] Leontovič A M 1962 On the existence of unbounded oscillating trajectories in a billiard problem *Dokl. Akad. Nauk SSSR* **145** 523–6
[18] Machta J 1983 Power law decay of correlations in a billiard problem *J. Stat. Phys.* **32** 555–64
[19] Melbourne I and Varandas P 2020 Convergence to a Lévy process in the Skorohod $\mathcal{M}_1$ and $\mathcal{M}_2$ topologies for nonuniformly hyperbolic systems, including billiards with cusps *Commun. Math. Phys.* **375** 653–78
[20] Pressley A 2010 *Elementary Differential Geometry* (*Springer Undergraduate Mathematics Series*) 2nd edn (Springer)
[21] Řeháček J 1995 On the ergodicity of dispersing billiards *Random Comput. Dyn.* **3** 35–55
[22] Ulam S 1950 On some statistical properties of dynamical systems *Berkeley Symp. on Mathematical Statistics and Probability* vol 3 pp 315–20
[23] Wojtkowski M P 1982 Geodesics on open surfaces containing horns *Stud. Math.* **73** 115–51
[24] Zhang H-K 2017 Decay of correlations for billiards with flat points II: cusps effect *Dynamical Systems, Ergodic Theory and Probability: In Memory of Kolya Chernov* (*Contemporary Mathematics* vol 698) (American Mathematical Society) pp 287–316