THE BILLIARD INSIDE AN ELLIPSE DEFORMED BY THE CURVATURE FLOW

JOSUÉ DAMASCENO, MARIO J. DIAS CARNEIRO, AND RAFAEL RAMÍREZ-ROS

(Communicated by ???)

ABSTRACT. The billiard dynamics inside an ellipse is integrable. It has zero topological entropy, four separatrices in the phase space, and a continuous family of convex caustics: the confocal ellipses. We prove that the curvature flow destroys the integrability, increases the topological entropy, splits the separatrices in a transverse way, and breaks all resonant convex caustics.

1. INTRODUCTION

One can shorten a smooth plane curve by moving it in the direction of its normal vector at a speed given by its curvature. This evolution generates a flow (called *curvature flow* or *curve shortening flow*) in the space of smooth plane curves that coincides with the negative L^2-gradient flow of the length of the curve. That is, the curve is shrinking as fast as it can using only local information.

M. Gage and R. Hamilton [10] described the long time behavior of smooth convex plane curves under the curvature flow. They proved that convex curves stay convex and shrink to a point as they become more circular. This convergence to a "limit" circle takes place in the C^∞-norm after a suitable normalization. M. Grayson proved that any embedded planar curve becomes convex before it shrinks to a point [11].

The length, the enclosed area, the total absolute curvature, the isoperimetric ratio (for convex curves), the number of inflection points, and other geometric quantities never increase along the curvature flow [6]. On the contrary, we present an example of how the curvature flow can increase the topological entropy of the billiard dynamics inside convex curves. The *topological entropy* of a dynamical system is a nonnegative extended real number that is a measure of the complexity of the system [15]. To be precise, the topological entropy represents the exponential growth rate of the number of distinguishable orbits as the system evolves. Therefore, increasing entropy means a more complex billiard dynamics, which is a bit surprising since the curvature flow rounds any convex smooth curve and circles are the curves with the simplest billiard dynamics.

Birkhoff [2] introduced the problem of *convex billiard tables* almost 90 years ago as a way to describe the motion of a free particle inside a closed convex smooth

curve. The particle is reflected at the boundary according to the law "angle of incidence equals angle of reflection".

If the boundary is an ellipse, then the billiard dynamics is *integrable* [5, 13, 19]. In particular, billiards inside ellipses have zero topological entropy. The motion along the major axis of the ellipse corresponds to a hyperbolic two-periodic orbit whose unstable and stable invariant curves coincide, forming four *separatrices*. The points on these separatrices correspond to the billiard trajectories passing through the foci of the ellipse. The interior of an ellipse is foliated with a continuous family of convex caustics: its confocal ellipses. A *caustic* is a curve inside the billiard table with the property that a billiard trajectory, once tangent to it, stays tangent after every reflection. Caustics with Diophantine rotation numbers persist under small smooth perturbations of the boundary [16], but *resonant caustics* —the ones whose tangent trajectories are closed polygons, so that their rotation numbers are rational— are fragile structures that generically break up [18, 17].

All these dynamical and geometric manifestations of the integrability of billiards inside ellipses disappear when the ellipse is slightly deformed by the curvature flow.

Theorem 1.1. *The curvature flow breaks all resonant convex caustics, splits the separatrices in a transverse way, increases the topological entropy, and destroys the integrability of the billiard inside an ellipse.*

The proof of this theorem has two steps. First, we introduce the subharmonic and homoclinic Melnikov potentials associated to the perturbation of the ellipse under the curvature flow following the theory developed in [8, 9, 18, 17]. In order to study these Melnikov potentials, we need several explicit formulas for the unperturbed billiard dynamics that can be found in [5, 8]. Second, we check that none of these Melnikov potentials is constant, which implies that the separatrices split and all resonant convex caustics break up. The loss of integrability follows directly from a theorem of Cushman [7], whereas the increase of the topological entropy follows from a theorem of Burns and Weiss [3].

We also find all the critical points of the Melnikov potentials, so we can locate all primary homoclinic points and all Birkhoff periodic trajectories, at least for small enough perturbations. Finally, we relate the homoclinic Melnikov potential to the limit of the subharmonic Melnikov potential when the resonant caustic tends to the separatrices. This is, up to our knowledge, the first time that such relation is explicitly shown up in a discrete system. Similar relations in continuous systems (that is, for ODEs) have been known from the eighties, see [12, §4.6].

Our perturbed ellipses are static, we do not deal with time-dependent billiards.

This paper is strongly inspired by Dan Jane's example [14] of a Riemannian surface for which the Ricci flow increases the topological entropy of the geodesic flow. His example is also based in a Melnikov computation, although the final step of his argument requires the numerical evaluation of some Melnikov function. On the contrary, our result is purely analytic, since we characterize our Melnikov potentials in a quite explicit way using the theory of elliptic functions.

We complete this introduction with a note on the organization of the article. In Section 2 we review some known results concerning billiards inside ellipses. The first order deformation of the ellipse under the curvature flow is given in Section 3. We review the Melnikov theory for billiard maps inside perturbed ellipses in Section 4. We check that the Melnikov potentials are not constant by analyzing their complex singularities in Section 5. Finally, we prove Theorem 1.1 in Section 6.

2. The billiard inside an ellipse

We consider the billiard dynamics inside the unperturbed ellipse

$$(2.1) \qquad Q_0 = \left\{ (x,y) \in \mathbb{R}^2 : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \right\}, \qquad 0 < b < a.$$

Let $c = \sqrt{a^2 - b^2}$ be the semi-focal distance of Q_0, so the foci of Q_0 are the points $(\pm c, 0)$. We recall a geometric property of ellipses [19]. Let

$$C_\lambda = \left\{ (x,y) \in \mathbb{R}^2 : \frac{x^2}{a^2 - \lambda^2} + \frac{y^2}{b^2 - \lambda^2} = 1 \right\}, \qquad \lambda \notin \{a, b\},$$

be the family of *confocal conics* to the ellipse Q_0. It is clear that C_λ is an ellipse for $0 < \lambda < b$ and a hyperbola for $b < \lambda < a$. No real conic exists for $\lambda > a$.

The fundamental property of the billiard inside Q_0 is that any segment (or its prolongation) of a billiard trajectory is tangent to C_λ for some fixed caustic parameter $\lambda > 0$. The notion of tangency in the degenerate case $\lambda = b$ is the following. A trajectory is tangent to C_b if it passes alternatively through the foci.

We refer to [1, 20] for a general background on Jacobian elliptic functions. Let us recall some definitions. Given a quantity $k \in (0,1)$, called the *modulus*, the *elliptic integral of the first kind* and the *complete elliptic integral of the first kind* are

$$F(\varphi) = F(\varphi, k) = \int_0^\varphi (1 - k^2 \sin^2 \phi)^{-1/2} d\phi, \qquad K = K(k) = F(\pi/2, k).$$

We also write $K' = K'(k) = K(\sqrt{1 - k^2})$. The *elliptic sine* and the *elliptic cosine* are defined by the relations

$$(2.2) \quad \operatorname{sn} t = \operatorname{sn}(t, k) = \sin \varphi, \qquad \operatorname{cn} t = \operatorname{cn}(t, k) = \cos \varphi, \qquad t = F(\varphi) = F(\varphi, k).$$

If the angular variable φ changes by 2π, then the angular variable t changes by $4K$. Thus, any 2π-periodic function in φ, becomes a $4K$-periodic function in t. We will usually denote the functions in t by putting a tilde above the name of the function in φ. For instance, the $4K$-periodic parameterization of the ellipse

$$(2.3) \qquad \tilde{q}_0 : \mathbb{R} \to Q_0, \qquad \tilde{q}_0(t) = (a \operatorname{sn} t, b \operatorname{cn} t),$$

is obtained from the 2π-periodic parameterization

$$(2.4) \qquad q_0 : \mathbb{R} \to Q_0, \qquad q_0(\varphi) = (a \sin \varphi, b \cos \varphi).$$

Clearly, $\tilde{q}_0(t) = q_0(\varphi)$. The billiard dynamics associated to the convex caustic C_λ becomes a rigid rotation $t \mapsto t + \delta$ in the angular variable t. It suffices to find the modulus k and shift δ associated to each convex caustic C_λ.

Lemma 2.1 ([5]). *Once fixed a caustic parameter $\lambda \in (0, b)$, we set the modulus $k \in (0, 1)$ and the shift $\delta \in (0, 2K)$ by the formulas*

$$(2.5) \qquad k^2 = (a^2 - b^2)/(a^2 - \lambda^2), \qquad \delta = 2F\big(\operatorname{asin}(\lambda/b), k\big).$$

The segment joining $\tilde{q}_0(t)$ and $\tilde{q}_0(t + \delta)$ is tangent to the caustic C_λ for all $t \in \mathbb{R}$.

Let m and n be two relatively prime integers such that $1 \le m < n/2$. Let $\rho(\lambda)$ be the rotation number of the convex caustic C_λ. We want to characterize the convex caustic C_λ whose tangent billiard trajectories form closed polygons with n sides that makes m turns inside Q_0 or, equivalently, the caustic parameter $\lambda \in (0, b)$ such that $\rho(\lambda) = m/n$. Such caustic parameter is unique because $\rho : (0, b) \to \mathbb{R}$ is an increasing analytic function such that $\rho(0) = 0$ and $\rho(b) = 1/2$, see [4]. Any

(m, n)-periodic billiard trajectory gives rise to a $(n-m, n)$-periodic one by inverting the direction of motion. Hence, a convex caustic is (m, n)-resonant if and only if it is also $(n-m, n)$-resonant. This explains why we can assume that $m < n/2$.

The caustic C_λ is the (m, n)-resonant convex caustic if and only if

$$(2.6) \qquad n\delta = 4Km.$$

This identity has the following geometric interpretation. When a billiard trajectory makes one turn around C_λ, the old angular variable φ changes by 2π, so the new angular variable t changes by $4K$. On the other hand, we have seen that the variable t changes by δ when a billiard trajectory bounces once. Hence, a billiard trajectory inscribed in Q_0 and circumscribed around C_λ makes exactly m turns around C_λ after n bounces if and only if (2.6) holds.

From now on, k and δ will denote the modulus and the shift defined in (2.5). We will also assume that relation (2.6) holds, since we only deal with resonant caustics. We will skip the dependence of the Jacobian elliptic functions on the modulus.

The billiard dynamics through the foci of the ellipse can also be simplified by using a suitable variable $s \in \mathbb{R}$. If a billiard trajectory passes alternatively through the foci, its segments tend to the major axis of the ellipse both in future and past. We consider the change of variables $(-\pi/2, \pi/2) \ni \varphi \mapsto s \in \mathbb{R}$ given by

$$(2.7) \qquad \tanh s = \sin \varphi, \qquad \operatorname{sech} s = \cos \varphi,$$

in order to give explicit formulas for this dynamics. If φ moves from $-\pi/2$ to $\pi/2$, then s moves from $-\infty$ to $+\infty$. Thus, any 2π-periodic function in φ generates a non-periodic function in s. We will usually denote the function in s by putting a hat above the name of the function in φ. For instance, the parametrization of the upper semi-ellipse $Q_0^+ = Q_0 \cap \{y > 0\}$ given by

$$(2.8) \qquad \hat{q}_0 : \mathbb{R} \to Q_0^+, \qquad \hat{q}_0(s) = (a \tanh s, b \operatorname{sech} s),$$

is obtained from parameterization (2.4). Clearly, $\hat{q}_0(s) = q_0(\varphi)$.

The billiard dynamics through the foci becomes a constant shift $s \mapsto s + h$ in the variable $s \in \mathbb{R}$ for a suitable shift $h > 0$.

Lemma 2.2 ([8]). *Once fixed the semi-lengths $0 < b < a$, let $c = \sqrt{a^2 - b^2}$ be the semi-focal distance and let $h > 0$ be the quantity determined by*

$$(2.9) \qquad \sinh(h/2) = c/b, \qquad \cosh(h/2) = a/b, \qquad \tanh(h/2) = c/a.$$

The segment from $\hat{q}_0(s)$ to $-\hat{q}_0(s+h)$ passes through the focus $(-c, 0)$ for all $s \in \mathbb{R}$.

Note that $\lim_{s \to \pm\infty} \hat{q}_0(s) = (\pm a, 0)$, which shows up that the trajectories through the foci tend to bounce between the vertices of the major axis of the ellipse. It is known that these vertices form a two-periodic hyperbolic trajectory whose *characteristic exponent* is h. That is, the eigenvalues of the differential of the billiard map at the two-periodic hyperbolic points are $\lambda = \mathrm{e}^h$ and $\lambda^{-1} = \mathrm{e}^{-h}$. Following a standard terminology in problems of splitting of separatrices, we will say that the parameterizations (2.3) and (2.8) are *natural parameterizations* of the billiard dynamics tangent to the convex caustic C_λ and through the foci, respectively.

Remark 2.3. We can associate four different billiard trajectories to each $s \in \mathbb{R}$. The first two ones are $\big((-1)^n \hat{q}_0(s+nh)\big)_{n \in \mathbb{Z}}$ and $\big((-1)^n \hat{q}_0(s-nh)\big)_{n \in \mathbb{Z}}$, which have the same starting point $\hat{q}_0(s) \in Q_0^+$ but are traveled in opposite directions. The last two ones are their symmetric images with respect to the origin: $\big((-1)^{n+1} \hat{q}_0(s+nh)\big)_{n \in \mathbb{Z}}$
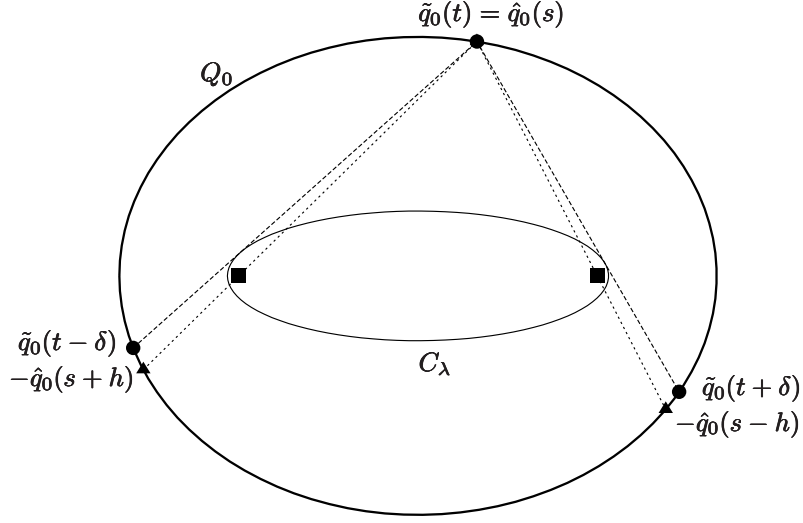
FIGURE 1. A billiard trajectory (dashed line) tangent to the ellipse C_λ tends to a billiard trajectory (dotted line) through the foci (the two solid squares) as $\lambda \to b^-$. The values of t and s are chosen in such a way that $\tilde{q}_0(t) = \hat{q}_0(s)$.

and $\left((-1)^{n+1}\hat{q}_0(s - nh)\right)_{n\in\mathbb{Z}}$, which start in a point on the lower semi-ellipse Q_0^-. Hence, there is a one-to-four correspondence between s and the homoclinic billiard trajectories inside the ellipse Q_0. Indeed, we should consider s defined modulo h, because s and $s + h$ give rise to the same set of four homoclinic trajectories.

The billiard dynamics through the foci corresponds to the caustic parameter $\lambda = b$, so it should be obtained as the limit of the billiard dynamics tangent to the convex caustic C_λ when $\lambda \to b^-$. See Figure 1. Note that C_λ flattens into the segment of the x-axis enclosed by the foci of the ellipse when $\lambda \to b^-$. We confirm this idea in the following lemma. We also stress that $\lim_{\lambda \to b^-} \delta \neq h$. This has to do with the minus sign that appears in Lemma 2.2 in front of the point $\hat{q}_0(s + h)$.

Lemma 2.4. *Let* $k \in (0, 1)$, $K = K(k) > 0$, $K' = K'(k) = K(\sqrt{1 - k^2}) > 0$, *and* $\delta \in (0, 2K)$ *be the modulus, the complete elliptic integral of the first kind, the complete elliptic integral of the first kind of the complementary modulus, and the constant shift associated to a convex caustic* C_λ. *Set* $\zeta = 2K - \delta \in (0, 2K)$. *Then:*

$$\lim_{\lambda \to b^-} k = 1, \qquad \lim_{\lambda \to b^-} K = +\infty, \qquad \lim_{\lambda \to b^-} K' = \pi/2, \qquad \lim_{\lambda \to b^-} \zeta = h,$$

where $h > 0$ *is the characteristic exponent defined in* (2.9). *Besides,*

$$\lim_{\lambda \to b^-} \tilde{q}_0(t) = \hat{q}_0(t), \qquad \lim_{\lambda \to b^-} \tilde{q}_0(t \pm \delta) = -\hat{q}_0(t \mp h),$$

and both limits are uniform on compacts sets of $\mathbb{R}$, *but not on* $\mathbb{R}$.

Proof. The first limit follows from the definition $k^2 = (a^2 - b^2)/(a^2 - \lambda^2)$. We know that $\lim_{k\to 1^-} K(k) = +\infty$ and $K'(1) = K(0) = \pi/2$, which gives the second and

third limits. Let $\alpha, \varphi, \psi, \psi_0 \in (0, \pi/2)$ be the angles determined by relations

$$\sin \alpha = k, \quad \sin \varphi = \lambda/b, \quad \sin \psi = \sqrt{1 - \lambda^2/a^2}, \quad \sin \psi_0 = \sqrt{1 - b^2/a^2} = c/a.$$

We note that $\cos \alpha \tan \varphi \tan \psi = 1$, which implies that $F(\varphi, k) + F(\psi, k) = K$. See [1, Formula 17.4.13]. Besides, $\delta/2 = F(\mathrm{asin}(\lambda/b), k) = F(\varphi, k)$. Therefore,

$$\lim_{\lambda \to b^-} \zeta/2 = \lim_{\lambda \to b^-} (K - \delta/2) = \lim_{\lambda \to b^-} \big(K - F(\varphi, k)\big) = \lim_{\lambda \to b^-} F(\psi, k)$$
$$= F(\psi_0, 1) = \mathrm{artanh}(\sin \psi_0) = \mathrm{artanh}(c/a) = h/2.$$

The property $\lim_{\lambda \to b^-} \tilde{q}_0(t) = \hat{q}_0(t)$ is a direct consequence of the limits

$$\lim_{k \to 1^-} \mathrm{sn}(t, k) = \tanh t, \qquad \lim_{k \to 1^-} \mathrm{cn}(t, k) = \mathrm{sech}\, t,$$

which can be found in [1]. Finally,

$$\lim_{\lambda \to b^-} \tilde{q}_0(t \pm \delta) = - \lim_{\lambda \to b^-} \tilde{q}_0(t \pm \delta \mp 2K) = -\hat{q}_0(t \mp h),$$

where we have used that $\tilde{q}_0(t)$ is $2K$-antiperiodic and $\lim_{\lambda \to b^-} \zeta = h$.   □

3. AN ELLIPSE UNDER THE CURVATURE FLOW

Let $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$. Let $Q_0 = q_0(\mathbb{T})$, $q_0 : \mathbb{T} \to \mathbb{R}^2$, be a closed smooth embedded curve in the plane. This curve may not be an ellipse. The t-time curvature flow of Q_0 is the curve $Q_t = q_t(\mathbb{T}) = q(\mathbb{T}; t)$ where the map $q : \mathbb{T} \times [0, \tau) \to \mathbb{R}^2$, $q = q(\varphi; t)$, satisfies the initial value problem

$$(3.1) \qquad \frac{\partial q}{\partial t} = \kappa N, \qquad q(\cdot, 0) = q_0.$$

Here, κ and N are the curvature and the unit inward normal vector, respectively. Observe that φ is not, in general, the arc-length parameter.

M. Gage and R. Hamilton [10] showed that if Q_0 is strictly convex, then the curvature flow is defined for $t \in [0, \tau)$, where $\tau = A_0/2\pi$ and A_0 is the area enclosed by Q_0. Besides, Q_t shrinks to a point and becomes more circular as $t \to \tau^-$.

Let Q_0 be the ellipse (2.1). We want to study a small deformation of Q_0 under the curvature flow. Henceforth, in order to emphasize that we are only interested in infinitesimal deformations of Q_0, we will denote the infinitesimally deformed ellipse by the symbol Q_ϵ, instead of Q_t.

We consider the elliptic coordinates (μ, φ) associated to the ellipse Q_0. That is, (μ, φ) are defined by relations

$$(3.2) \qquad x = c \cosh \mu \sin \varphi, \qquad y = c \sinh \mu \cos \varphi,$$

where $c = \sqrt{a^2 - b^2}$ is the semi-focal distance of Q_0. The ellipse Q_0 in these elliptic coordinates reads as $\mu \equiv \mu_0$, where $\cosh \mu_0 = a/c$ and $\sinh \mu_0 = b/c$. Therefore, the deformation Q_ϵ of the ellipse Q_0 can be written in elliptic coordinates as

$$(3.3) \qquad \mu = \mu_\epsilon(\varphi) = \mu_0 + \epsilon \mu_1(\varphi) + \mathrm{O}(\epsilon^2),$$

for some 2π-periodic smooth function $\mu_\epsilon : \mathbb{R} \to \mathbb{R}$. If a curve is symmetric with respect to a line, so is its curvature flow deformation, as long as it exists. Thus, the deformation Q_ϵ has the axial symmetries of the ellipse Q_0 with respect to both coordinates axis. This means that $\mu_\epsilon(\varphi)$ is even and π-periodic. Next, we compute the first order term of this function. That is, we compute the function $\mu_1(\varphi)$.

Lemma 3.1. *Let Q_ϵ be the deformation under the ϵ-time curvature flow of the ellipse* (2.1). *If we write the deformed ellipse Q_ϵ as in equation* (3.3), *then*

$$(3.4) \qquad \mu_1(\varphi) = \frac{-ab}{(a^2 \cos^2 \varphi + b^2 \sin^2 \varphi)^2}.$$

Proof. Let $q : \mathbb{T} \times [0, \tau) \to \mathbb{R}$, $q = q_t(\varphi) = q(\varphi; t)$, be the solution of the initial value problem (3.1), where $q_0(\varphi) = (a \sin \varphi, b \cos \varphi)$. On the one hand, we obtain from (3.1) that $q_\epsilon(\varphi) = q_0(\varphi) + \epsilon q_1(\varphi) + \mathrm{O}(\epsilon^2)$, where $q_1(\varphi) = \kappa_0(\varphi) N_0(\varphi)$,

$$\kappa_0(\varphi) = \frac{ab}{(a^2 \cos^2 \varphi + b^2 \sin^2 \varphi)^{3/2}}$$

is the curvature of the ellipse Q_0 at the point $q_0(\varphi)$, and

$$N_0(\varphi) = \frac{-1}{(a^2 \cos^2 \varphi + b^2 \sin^2 \varphi)^{1/2}} (b \sin \varphi, a \cos \varphi)$$

is the inward unit normal vector of the ellipse Q_0 at the point $q_0(\varphi)$.

On the other hand, we deduce from the elliptic coordinates (3.2) that

$$q_\epsilon(\varphi) = (c \cosh \mu_\epsilon(\varphi) \sin \varphi, c \sinh \mu_\epsilon(\varphi) \cos \varphi)$$
$$= (a \sin \varphi, b \cos \varphi) + \epsilon \mu_1(\varphi)(b \sin \varphi, a \cos \varphi) + \mathrm{O}(\epsilon^2).$$

By combining these two results, we get that

$$\frac{-ab}{(a^2 \cos^2 \varphi + b^2 \sin^2 \varphi)^2} (b \sin \varphi, a \cos \varphi) = \mu_1(\varphi)(b \sin \varphi, a \cos \varphi),$$

which implies formula (3.4). $\qquad \square$

4. Subharmonic and homoclinic Melnikov potentials

For the sake of brevity, we introduce both Melnikov potentials just for billiards by means of a variational approach. A more powerful geometric approach for the study of the break up of resonant invariant curves and the splitting of separatrices of general area-preserving twist maps can be found in [8, 9, 17].

We begin with some basic properties of billiards described in [13, 19].

Let $Q = q(\mathbb{T})$, $q : \mathbb{T} \to \mathbb{R}^2$, be a closed smooth embedded curve in the plane. This curve may not be an ellipse. The associated billiard problem consists in the free motion of a particle inside Q that is ellastically reflected at the impacts with Q. We model this dynamics by means of the *billiard map* $f : \mathbb{T} \times (0, \pi) \to \mathbb{T} \times (0, \pi)$, $f(\varphi, \theta) = (\varphi', \theta')$, defined as follows. If the particle hits Q at a point $q(\varphi)$ under an angle of incidence $\theta \in (0, \pi)$, then $q(\varphi')$ is the next impact point and $\theta' \in (0, \pi)$ is the next angle of incidence.

The billiard map is an exact twist map with Lagrangian $h(\varphi, \varphi') = |q(\varphi) - q(\varphi')|$. This means that the billiard dynamics satisfies the following variational principle. Billiard trajectories inside Q are in one-to-one correspondence with the (formal) stationary configurations $\Phi = (\varphi_j)_{j \in \mathbb{Z}}$ of the *action functional*

$$(4.1) \qquad W[\Phi] = \sum_{j \in \mathbb{Z}} h(\varphi_j, \varphi_{j+1}).$$

The series for $W[\Phi]$ may be divergent, but $\frac{\partial W}{\partial \varphi_j}$ only involves two terms of the series, and so ∇W is always well defined. Indeed, $W[\Phi]$ can be written as a convergent series when dealing with periodic and homoclinic trajectories. Let us explain this.

Let m and n be relatively prime integers such that $1 \leq m < n/2$. If the impact points $q(\varphi_j)$, $j \in \mathbb{Z}$, form a (m,n)-periodic billiard trajectory, then

$$h(\varphi_{j+n}, \varphi_{j+n+1}) = h(\varphi_j + 2\pi m, \varphi_{j+1} + 2\pi m) = h(\varphi_j, \varphi_{j+1}),$$

so there are only n different terms in the action functional (4.1), which encode the (m,n)-periodic dynamics. In particular, any (m,n)-periodic billiard trajectory is in correspondence with a stationary configuration $\Phi = (\varphi_0, \ldots, \varphi_{n-1})$ of the (m,n)-*periodic action*

$$(4.2) \qquad W^{(m,n)}[\Phi] = h(\varphi_0, \varphi_1) + \cdots + h(\varphi_{n-1}, \varphi_0 + 2\pi m).$$

If there is a (m,n)-resonant caustic C inside Q, then there is a continuum of (m,n)-periodic billiard trajectories inscribed in Q and circumscribed around C, and so, a continuum $\Phi = \Phi(\varphi) = (\varphi_0(\varphi), \ldots, \varphi_{n-1}(\varphi))$, $\varphi_0(\varphi) = \varphi$, of stationary configurations of the periodic action (4.2). Hence, the existence of C implies that

$$L^{(m,n)} : \mathbb{T} \to \mathbb{R}, \qquad L^{(m,n)}(\varphi) = W^{(m,n)}[\Phi(\varphi)] = \sum_{j=0}^{n-1} h(\varphi_j(\varphi), \varphi_{j+1}(\varphi))$$

is a constant function, because $\nabla W^{(m,n)}[\Phi(\varphi)] \equiv 0$. Here, $\varphi_n(\varphi) = \varphi + 2\pi m$.

Next, we consider a perturbative situation. Let $Q_\epsilon = q_\epsilon(\mathbb{T})$, $q_\epsilon : \mathbb{T} \to \mathbb{R}^2$, be a perturbed ellipse that has the form (3.3) in the elliptic coordinates (3.2). We do not assume now that this perturbed ellipse is obtained through the curvature flow. Let $W_\epsilon^{(m,n)} = W_0^{(m,n)} + \epsilon W_1^{(m,n)} + \mathrm{O}(\epsilon^2)$ be the (m,n)-periodic action associated to the perturbed ellipse Q_ϵ. That is,

$$W_k^{(m,n)}[\Phi] = h_k(\varphi_0, \varphi_1) + \cdots + h_k(\varphi_{n-1}, \varphi_0 + 2\pi m), \qquad k = 0, 1,$$

where $h_\epsilon = h_0 + \epsilon h_1 + \mathrm{O}(\epsilon^2)$ is the Lagrangian of the billiard map inside Q_ϵ. If the (m,n)-resonant convex caustic C_λ persists under the deformation Q_ϵ, then there exists a continuum $\Phi_\epsilon = \Phi_\epsilon(\varphi) = (\varphi_{\epsilon,0}(\varphi), \ldots, \varphi_{\epsilon,n-1}(\varphi))$, $\varphi_{\epsilon,0}(\varphi) = \varphi$, of critical configurations of $W_\epsilon^{(m,n)}$, and so the function $L_\epsilon^{(m,n)} : \mathbb{T} \to \mathbb{R}$,

$$L_\epsilon^{(m,n)}(\varphi) = W_\epsilon^{(m,n)}[\Phi_\epsilon(\varphi)] = W_0^{(m,n)}[\Phi_0(\varphi)] + \epsilon W_1^{(m,n)}[\Phi_0(\varphi)] + \mathrm{O}(\epsilon^2),$$

should be constant, although this constant may depend on ϵ. We have used that $\nabla W_0^{(m,n)}[\Phi_0(\varphi)] \equiv 0$ in the second equality. Hence, the first-order term $L_1^{(m,n)}(\varphi) = W_1^{(m,n)}[\Phi_0(\varphi)]$ should also be constant. This function has the form

$$(4.3) \qquad L_1^{(m,n)} : \mathbb{T} \to \mathbb{R}, \qquad L_1^{(m,n)}(\varphi) = \sum_{j=0}^{n-1} h_1(\varphi_j, \varphi_{j+1}) = 2\lambda \sum_{j=0}^{n-1} \mu_1(\varphi_j).$$

The last equality is proved in the middle of the proof of Proposition 4.1 in [17]. The points $q_j = (a \sin \varphi_j, b \cos \varphi_j)$ are the consecutive impact points of the billiard trajectory inscribed in the ellipse Q_0 and circumscribed around the caustic C_λ starting at $q_0 = (a \sin \varphi, b \cos \varphi)$. That is, the sequence $\varphi_0 = \varphi, \varphi_1, \ldots, \varphi_{n-1}, \varphi_n$ is given by the unperturbed billiard dynamics around the (m,n)-resonant caustic.

Definition 4.1. The function (4.3) is the (m,n)-*subharmonic Melnikov potential* for the billiard dynamics inside the perturbed ellipse (3.3).

Corollary 4.2. *If the (m,n)-subharmonic Melnikov potential is not constant, then the (m,n)-resonant caustic does not persist under perturbation* (3.3).

Remark 4.3. The following result is contained in [17], but we do not need it. If the (m, n)-subharmonic Melnikov potential does not have degenerate critical points and $\epsilon > 0$ is small enough, then there is a correspondence between its critical points and the (m, n)-periodic Birkhoff billiard orbits inside the deformed ellipse (3.3).

The persistence of a continuum of homoclinic billiard trajectories (that is, the persistence of a *separatrix*) can be studied by using the same variational technique.

For simplicity, we assume that the perturbation (3.3) preserves the axial symmetries of the ellipse Q_0. This means that $\mu_\epsilon(\varphi) = \mu_0 + \epsilon\mu_1(\varphi) + \mathrm{O}(\epsilon^2)$ is an even π-periodic smooth function. In particular,

$$\breve{\mu}_\infty := \mu_1(-\pi/2) = \mu_1(\pi/2).$$

There are four separatrices, but symmetric perturbations cause the same effect on any of them. Therefore, it suffices to study a single homoclinic Melnikov potential.

The homoclinic Melnikov potential (4.4) is defined in a completely analogous way to the subharmonic Melnikov potential (4.3). There are just three differences. First, the caustic parameter is $\lambda = b$ in the homoclinic case. Second, the sum for the homoclinic Melnikov potential is infinite, because homoclinic billiard trajectories have infinite impact points. Third, we are forced to substract the constant $\breve{\mu}_\infty$ from each term $\mu_1(\varphi_j)$ in order to get a convergent series.

Definition 4.4. The *homoclinic Melnikov potential* for the billiard dynamics inside the symmetrically perturbed ellipse (3.3) is the function

$$(4.4) \qquad L_1 : (-\pi/2, \pi/2) \to \mathbb{R}, \qquad L_1(\varphi) = 2b \sum_{j \in \mathbb{Z}} \big(\mu_1(\varphi_j) - \breve{\mu}_\infty\big),$$

where $q_0(\varphi_j) = (-1)^j(a \sin\varphi_j, b \cos\varphi_j)$ are the consecutive impact points of the billiard trajectory inside the ellipse Q_0 such that the segment from $q_0(\varphi)$ to $-q_0(\varphi_1)$ passes through the focus $(-c, 0)$.

The series in the homoclinic Melnikov potential converges uniformly on compact subsets of $(-\pi/2, \pi/2)$ because $\mu_1(\varphi_j)$ tends geometrically fast to $\breve{\mu}_\infty$ as $j \to \pm\infty$.

Proposition 4.5 ([9])**.** *If the homoclinic Melnikov potential (4.4) is not constant, the separatrices of the unperturbed billiard map do not persist under the perturbation (3.3). If the homoclinic Melnikov potential does not have degenerate critical points and $\epsilon > 0$ is small enough, then there is a one-to-four correspondence between its critical points (modulo the billiard dynamics) and the transverse primary homoclinic billiard trajectories inside the deformed ellipse (3.3).*

This proposition follows directly from Theorem 2.1 in [9], where the splitting of separatrices and the appearance of perturbed transverse primary homoclinic orbits are studied in the more general setting of exact symplectic maps.

Remark 4.6. The correspondence is one-to-four because each critical point gives rise to two different homoclinic "paths" (mirrored by the central symmetry with respect to the origin) and each "path" can be traveled in two directions. See Remark 2.3.

5. Computations with elliptic functions

Let us assume that the perturbed ellipse (3.3) is the ϵ-time curvature flow of the ellipse (2.1), so that the first order term $\mu_1(\varphi)$ has the form (3.4). We can not apply the result about non-persistence of resonant convex caustics established in [17] or

the result about splitting of separatrices established in [8] to this curvature flow setting, because the function (3.4) is not entire in the variable φ. Nevertheless, many of the ideas developed in [8, 17] are still useful.

Let $\tilde{\mu}_1^{(m,n)} : \mathbb{R} \to \mathbb{R}$ be the function defined by $\tilde{\mu}_1^{(m.n)}(t) = \mu_1(\varphi)$, so

$$(5.1) \qquad \tilde{\mu}_1^{(m,n)}(t) = \frac{-ab}{(a^2 \operatorname{cn}^2 t + b^2 \operatorname{sn}^2 t)^2}.$$

Here, C_λ is the (m,n)-resonant convex caustic inside Q_0, the modulus $k \in (0,1)$ and the shift $\delta \in (0, 2K)$ are defined in (2.5), and variables φ and t are related through relation $t = F(\varphi)$, so identities (2.2) hold. We skip the dependence of the Jacobian elliptic functions on the modulus k.

Analogously, let $\hat{\mu}_1 : \mathbb{R} \to \mathbb{R}$ be the function defined by $\hat{\mu}_1(s) = \mu_1(\varphi) - \breve{\mu}_\infty$, so

$$(5.2) \qquad \hat{\mu}_1(s) = \frac{a}{b^3} - \frac{ab}{(a^2 \operatorname{sech}^2 s + b^2 \tanh^2 s)^2}.$$

Here, variables φ and s are related through identities (2.7).

The key observation in what follows is that (5.1) can be analytically extended to an elliptic function defined over $\mathbb{C}$, whereas (5.2) can be analytically extended to a meromorphic function over $\mathbb{C}$. We list below the main properties of these extensions.

Lemma 5.1. *Let m and n be two relatively prime integers such that $1 \leq m < n/2$. Let C_λ be the (m,n)-resonant elliptical caustic of the ellipse Q_0. Let $\delta \in (0, 2K)$ be the shift defined by $\operatorname{sn}(\delta/2) = \lambda/b$, so relation (2.6) holds. Set $\zeta = 2K - \delta \in (0, 2K)$. The function (5.1) is an even elliptic function of order four, periods $2K$ and $2K'\mathrm{i}$, and double poles in the set*

$$T = T_- \cup T_+, \qquad T_\pm = t_\pm + 2K\mathbb{Z} + 2K'\mathrm{i}\mathbb{Z}, \qquad t_\pm = \pm\zeta/2 + K'\mathrm{i}.$$

It has no other poles. There exist two Laurent coefficients $\alpha_2, \alpha_1 \in \mathbb{C}$, with $\alpha_2 \neq 0$, such that

$$\tilde{\mu}_1^{(m,n)}(t_\pm + \tau) = \frac{\alpha_2}{\tau^2} \pm \frac{\alpha_1}{\tau} + \mathrm{O}(1), \qquad \tau \to 0.$$

Proof. We know that the square of the elliptic cosine is an even elliptic function of order two and periods $2K$ and $2K'\mathrm{i}$. Thus, the function

$$f(t) = a^2 \operatorname{cn}^2 t + b^2 \operatorname{sn}^2 t = b^2 + (a^2 - b^2) \operatorname{cn}^2 t$$

has the same properties. (We have used the identity $\operatorname{sn}^2 + \operatorname{cn}^2 \equiv 1$.) Hence, the function $f(t)$ has exactly two roots (counted with multiplicity) in the complex cell

$$C = \{t \in \mathbb{C} : -K \leq \Re t < K, \ 0 \leq \Im t < 2K'\}.$$

Let us find them. On the one hand, the values of the Jacobian elliptic functions at $t = K$ are

$$\operatorname{sn} K = 1, \qquad \operatorname{cn} K = 0, \qquad \operatorname{dn} K = \sqrt{1 - k^2} = \sqrt{(b^2 - \lambda^2)/(a^2 - \lambda^2)}.$$

On the other hand, the values of the Jacobian elliptic functions at $t = \delta/2$ are $\operatorname{sn}(\delta/2) = \lambda/b$,

$$\operatorname{cn}(\delta/2) = b^{-1}\sqrt{b^2 - \lambda^2}, \qquad \operatorname{dn}(\delta/2) = ab^{-1}\sqrt{(b^2 - \lambda^2)/(a^2 - \lambda^2)}.$$

Therefore, the addition formula for the elliptic sine implies that

$$\operatorname{sn}(\zeta/2) = \operatorname{sn}(K - \delta/2) = \frac{\operatorname{sn} K \operatorname{cn}(\delta/2) \operatorname{dn}(\delta/2) - \operatorname{sn}(\delta/2) \operatorname{cn} K \operatorname{dn} K}{1 - k^2 \operatorname{sn}^2 K \operatorname{sn}^2(\delta/2)} = \sqrt{a^2 - \lambda^2}/a.$$

Next, we check that the function $f(t)$ vanishes at the points $t = t_\pm$:

$$f(t_\pm) = b^2 + (a^2 - b^2)\operatorname{cn}^2(\pm\zeta/2 + K'\mathrm{i})$$
$$= b^2 + (a^2 - b^2)\left(1 - k^{-2}\operatorname{sn}^{-2}(\pm\zeta/2)\right) = 0.$$

We note that $t_\pm = \pm\zeta/2 + K'\mathrm{i} \in C$, so these are the two roots we were looking for and, in addition, they are simple roots. From the parity and periodicity of $f(t)$, we also deduce that

$$f'(t_-) = f'(-\zeta/2 + K'\mathrm{i}) = -f'(\zeta/2 - K'\mathrm{i}) = -f'(\zeta/2 + K'\mathrm{i}) = -f'(t_+).$$

Finally, all the properties of the function (5.1) follow directly from the fact that $\tilde{\mu}_1^{(m,n)} = -ab/f^2$. It suffices to take $\alpha_2 = -ab/(f'(t_+))^2 = -ab/(f'(t_-))^2 \neq 0$. $\quad\square$

Lemma 5.2. *Let $h > 0$ be the characteristic exponent (2.9). The function (5.2) is an even meromorphic $\pi\mathrm{i}$-periodic function with double poles in the set*

$$S = S_- \cup S_+, \qquad S_\pm = s_\pm + \pi\mathrm{i}\mathbb{Z}, \qquad s_\pm = \pm h/2 + \pi\mathrm{i}/2.$$

It has no other poles. There exist two Laurent coefficients $\beta_2, \beta_1 \in \mathbb{C}$, with $\beta_2 \neq 0$, such that

$$\hat{\mu}_1(s_\pm + \sigma) = \frac{\beta_2}{\sigma^2} \pm \frac{\beta_1}{\sigma} + \mathrm{O}(1), \qquad \sigma \to 0.$$

Proof. The square of the hyperbolic secant is an even meromorphic $\pi\mathrm{i}$-periodic function. Thus, the function

$$g(s) = a^2\operatorname{sech}^2 s + b^2\tanh^2 s = b^2 + c^2\operatorname{sech}^2 s$$

has the same properties. We have used the identities $\operatorname{sech}^2 + \tanh^2 \equiv 1$ and $c^2 = a^2 - b^2$. Next, we look for all the roots of $g(s)$. We note that

$$g(s) = 0 \Leftrightarrow \cosh^2 s = -c^2/b^2 = -\sinh^2(h/2) = \cosh^2(h/2 + \pi\mathrm{i}/2) \Leftrightarrow s \in S.$$

We have used that $\cosh^2 s = \cosh^2 r$ if and only if $s - r \in \pi\mathrm{i}\mathbb{Z}$ or $s + r \in \pi\mathrm{i}\mathbb{Z}$. These roots are simple. In fact, if $s_* \in S$, then $\cosh^2 s_* = -c^2/b^2$ and $\sinh^2 s_* = -a^2/b^2$, so $g(s_*) = 0$ and $g'(s_*) = -2c^2\sinh s_*/\cosh^3 s_* \neq 0$. From the parity and periodicity of $g(s)$, we deduce that $g'(s_-) = -g'(h/2 - \pi\mathrm{i}/2) = -g'(s_+)$. Finally, all the properties of $\hat{\mu}_1(s)$ follow directly from the fact that $\hat{\mu}_1 = a/b^3 - ab/g^2$. It suffices to take $\beta_2 = -ab/(g'(s_+))^2 = -ab/(g'(s_-))^2 \neq 0$. $\quad\square$

Next, we are going to prove that the (m,n)-subharmonic Melnikov potential (4.3) is not constant. We recall that the billiard dynamics inside the ellipse Q_0 around the (m,n)-resonant caustic C_λ becomes a rigid rotation $t \mapsto t+\delta$ in the new variable $t = F(\varphi)$. Thus, the (m,n)-subharmonic Melnikov potential (4.3) becomes

$$(5.3) \quad \tilde{L}_1^{(m,n)}(t) = 2\lambda\sum_{j=0}^{n-1}\tilde{\mu}_1^{(m,n)}(t + j\delta), \qquad \tilde{\mu}_1^{(m,n)}(t) = \frac{-ab}{(a^2\operatorname{sn}^2 t + b^2\operatorname{cn}^2 t)^2},$$

in the variable t.

Proposition 5.3. *Let $\alpha_2 \neq 0$ be the dominant Laurent coefficient introduced in Lemma 5.1. The (m,n)-subharmonic Melnikov potential (5.3) is an even elliptic function of order two with periods ζ and $2K'\mathrm{i}$, poles in the set*

$$(5.4) \qquad\qquad \mathcal{T} = t_\star + \zeta\mathbb{Z} + 2K'\mathrm{i}\mathbb{Z}, \qquad t_\star = \zeta/2 + K'\mathrm{i},$$

and principal parts

$$\tilde{L}_1^{(m,n)}(t_\star + \tau) = \begin{cases} 4\lambda\alpha_2\tau^{-2} + \mathrm{O}(1) \ \ as \ \tau \to 0, & if \ n \ is \ odd, \\ 8\lambda\alpha_2\tau^{-2} + \mathrm{O}(1) \ \ as \ \tau \to 0, & if \ n \ is \ even. \end{cases}$$

In particular, it is not constant. Besides, its only real critical points are the points of the set $\zeta\mathbb{Z}/2$, and all of them are nondegenerate.

Proof. We skip the dependence of $\tilde{\mu}_1^{(m,n)}(t)$ and $\tilde{L}_1^{(m,n)}(t)$ on (m,n) for simplicity. The finite sum $\tilde{L}_1(t) = 2\lambda \sum_{j=0}^{n-1} \tilde{\mu}_1(t + j\delta)$ can be analytically extended to an elliptic function $\tilde{L}_1 : \mathbb{C} \to \mathbb{C}$ defined over the whole complex plane, see Lemma 5.1. The point $t_+ \in \mathbb{C}$ is a singularity of $\tilde{\mu}_1(t + j\delta)$ if and only if $t_+ + j\delta \in T = T_- \cup T_+$. Besides,

$$t_+ + j\delta \in T_+ \Leftrightarrow j\delta \in 2K\mathbb{Z} \Leftrightarrow 2jm \in n\mathbb{Z} \Leftrightarrow j \in \{0, n/2\},$$
$$t_+ + j\delta \in T_- \Leftrightarrow (j-1)\delta \in 2K\mathbb{Z} \Leftrightarrow 2(j-1)m \in n\mathbb{Z} \Leftrightarrow j - 1 \in \{0, n/2\}.$$

We have used that $\delta = 4Km/n$, $t_- = t_+ + \delta - 2K$, and $\gcd(m,n) = 1$. Equalities $j = n/2$ and $j - 1 = n/2$ only can take place when n is even. Hence, we distinguish two cases:

- If n is odd, then $\tilde{\mu}_1(t)$ and $\tilde{\mu}_1(t + \delta)$ are the only terms in the sum that have a singularity at $t = t_+$, so that

$$\begin{aligned} \tilde{L}_1(t_+ + \tau) &= 2\lambda\tilde{\mu}_1(t_+ + \tau) + 2\lambda\tilde{\mu}_1(t_+ + \delta + \tau) + \mathrm{O}(1) \\ &= 2\lambda\tilde{\mu}_1(t_+ + \tau) + 2\lambda\tilde{\mu}_1(t_- + \tau) + \mathrm{O}(1) \\ &= 4\lambda\alpha_2\tau^{-2} + \mathrm{O}(1) \quad \text{as } \tau \to 0. \end{aligned}$$

- If n is even, then $n\delta/2 = 2Km$ and $\tilde{L}_1(t) = 4\lambda \sum_{j=0}^{n/2-1} \tilde{\mu}_1(t + j\delta)$. We note that $\tilde{\mu}_1(t)$ and $\tilde{\mu}_1(t + \delta)$ are the only terms in this new sum that have a singularity at $t = t_+$, so

$$\begin{aligned} \tilde{L}_1(t_+ + \tau) &= 4\lambda\tilde{\mu}_1(t_+ + \tau) + 4\lambda\tilde{\mu}_1(t_+ + \delta + \tau) + \mathrm{O}(1) \\ &= 4\lambda\tilde{\mu}_1(t_+ + \tau) + 4\lambda\tilde{\mu}_1(t_- + \tau) + \mathrm{O}(1) \\ &= 8\lambda\alpha_2\tau^{-2} + \mathrm{O}(1) \quad \text{as } \tau \to 0. \end{aligned}$$

Thus, the analytic extension $\tilde{L}_1 : \mathbb{C} \to \mathbb{C}$ has a double pole at $t = t_+$ in both cases, which implies that $\tilde{L}_1 : \mathbb{R} \to \mathbb{R}$ is not constant.

Next, let us prove that the points in the set $\zeta\mathbb{Z}/2$ are the only real critical points of $\tilde{L}_1(t)$, and all of them are nondegenerate. The derivative $\tilde{L}_1'(t)$ is odd, has periods ζ and $2K'\mathrm{i}$, has triple poles in the set (5.4), and vanishes at the points in the set $\{0, \zeta/2, K'\mathrm{i}\} + \zeta\mathbb{Z} + 2K'\mathrm{i}\mathbb{Z}$ due to its symmetry and periodicities. These critical points are nondegenerate and they are the only critical points because $\tilde{L}_1'(t)$ is an elliptic function of order three. $\qquad\square$

Let us check that the homoclinic Melnikov potential (4.4) is not constant. We recall that the billiard dynamics inside the ellipse Q_0 through the foci becomes a constant shift $s \mapsto s + h$ in the variable s defined by (2.7). Thus, the homoclinic Melnikov potential (4.4) becomes

$$(5.5) \qquad \hat{L}_1(s) = 2b \sum_{j\in\mathbb{Z}} \hat{\mu}_1(s + jh), \qquad \hat{\mu}_1(s) = \frac{a}{b^3} - \frac{ab}{(a^2 \operatorname{sech}^2 s + b^2 \tanh^2 s)^2},$$

in the variable s.

Proposition 5.4. *Let $\beta_2 \neq 0$ be the dominant Laurent coefficient introduced in Lemma 5.2. The homoclinic Melnikov potential (5.5) is an even elliptic function of order two with periods h and $\pi\mathrm{i}$, poles in the set*

$$\mathcal{S} = s_\star + h\mathbb{Z} + \pi\mathrm{i}\mathbb{Z}, \qquad s_\star = h/2 + \pi\mathrm{i}/2,$$

and principal parts

$$\hat{L}_1(s_\star + \sigma) = 4b\beta_2\sigma^{-2} + \mathrm{O}(1), \qquad \sigma \to 0.$$

In particular, it is not constant. Besides, its only real critical points are the points of the set $h\mathbb{Z}/2$, and all of them are nondegenerate.

Proof. The series $\hat{L}_1(s) = 2b\sum_{j=0}^{n-1}\hat{\mu}_1(s + jh)$ can be analytically extended to a meromorphic function $\hat{L}_1 : \mathbb{C} \to \mathbb{C}$ defined over the whole complex plane, see Lemma 5.2. The point $s_+ \in \mathbb{C}$ is a singularity of the j-th term $\hat{\mu}_1(s + jh)$ if and only if $s_+ + jh \in S = S_- \cup S_+$. Besides,

$$s_+ + jh \in S_+ \Leftrightarrow j = 0, \qquad s_+ + jh \in S_- \Leftrightarrow j = -1,$$

Here, we have used that $h \in \mathbb{R}$ and $s_- = s_+ - h$. Hence, $\hat{\mu}_1(s - h)$ and $\hat{\mu}_1(s)$ are the only terms in the sum that have a singularity at $s = s_+$, so that

$$\begin{aligned}
\hat{L}_1(s_+ + \sigma) &= 2b\hat{\mu}_1(s_+ - h + \sigma) + 2b\hat{\mu}_1(s_+ + \sigma) + \mathrm{O}(1) \\
&= 2b\hat{\mu}_1(s_- + \sigma) + 2b\hat{\mu}_1(s_+ + \sigma) + \mathrm{O}(1) \\
&= 4b\beta_2\sigma^{-2} + \mathrm{O}(1) \quad \text{as } \sigma \to 0.
\end{aligned}$$

Therefore, the analytic extension $\hat{L}_1 : \mathbb{C} \to \mathbb{C}$ has a double pole at $s = s_+$, which implies that the homoclinic Melnikov potential $\hat{L}_1 : \mathbb{R} \to \mathbb{R}$ is not constant.

Finally, the points in the set $h\mathbb{Z}/2$ are the only real critical points of $\hat{L}_1(s)$, and all of them are nondegenerate. This is proved following the same argument at the end of the proof of the previous proposition. $\qquad\square$

Finally, we establish the relation between the homoclinic Melnikov potential and the limit of the (m,n)-subharmonic Melnikov potential when $m/n \to 1/2$ or, equivalently, when $\lambda \to b^-$. We still assume that the perturbed ellipse is obtained by using the curvature flow, so this is a very specific result. The relation depends on the parity of the period n, which is a phenomenon that, up to our knowledge, never takes place in continuous systems. This is the reason for our interest in it.

Proposition 5.5. *If m and n are relatively prime integers such that $1 \leq m < n/2$,*

$$\lim_{\frac{m}{n} \to \frac{1}{2}} \tilde{L}_1^{(m,n)}(t) = constant + \begin{cases} \hat{L}_1(t), & \text{if } n \text{ is odd}, \\ 2\hat{L}_1(t), & \text{if } n \text{ is even}, \end{cases}$$

uniformly on compact subsets of $\mathbb{R}$.

Proof. The proof is based on the fact that any elliptic function is determined, up to an additive constant, by its periods, its poles, and the principal parts of its poles. The periods, poles, and principal parts of the subharmonic and homoclinic Melnikov potentials $\tilde{L}_1^{(m,n)}(t)$ and $\hat{L}_1(s)$ are listed in Propositions 5.3 and 5.4, respectively. We only have to see that the former ones tend to the later ones.

Let $\lambda \in (0, b)$ be the caustic parameter such that C_λ is an (m,n)-resonant caustic. It is known that if $m/n \to 1/2$, then $\lambda \to b^-$. See [4, Proposition 10]. Besides, we have seen in Lemma 2.4 that $\lim_{\lambda \to b^-} K' = \pi/2$ and $\lim_{\lambda \to b^-} \zeta = h$. Thus, it

suffices to check that $\lim_{\lambda \to b^-} \alpha_2 = \beta_2$, where α_2 and β_2 are the Laurent coefficients introduced in Lemmas 5.1 and 5.2. This limit is a straightforward computation.   □

6. Proof of Theorem 1.1

The claims of Theorem 1.1 about the break up of all resonant convex caustics and the splitting of the separatrices in a transverse way are a direct consequence of the results above. Let us explain this.

First, let C_λ be the (m,n)-resonant convex caustic of the billiard dynamics inside the unperturbed ellipse Q_0, and let $\tilde{L}_1^{(m,n)}(t)$ be the (m,n)-subharmonic Melnikov potential for the billiard dynamics inside the ellipse Q_0 deformed by the curvature flow. We have seen in Proposition 5.3 that $\tilde{L}_1^{(m,n)}(t)$ is not constant. Hence, the caustic C_λ does not persist under the curvature flow; see Corollary 4.2.

Second, let $\hat{L}_1(s)$ be the homoclinic Melnikov potential for the billiard dynamics inside the ellipse Q_0 deformed by the curvature flow. We have seen in Proposition 5.4 that $\hat{L}_1(s)$ is not constant, its only real critical points are the points of the set $h\mathbb{Z}/2$, and all of them are nondegenerate. Thus, the separatrices do not persist under the curvature flow and the billiard map associated to the deformed ellipse has eight primary transverse homoclinic trajectories; see Proposition 4.5.

Next, we recall two classical results about surface diffeomorphisms possesing homoclinic points with a topological (that is, not necessarily transverse) crossing.

Let $W^s(p)$ and $W^u(p)$ be the stable and unstable invariant curves of a hyperbolic fixed point p of a diffeomorphism $f : M \to M$ defined on a surface M. The points $q \in W^s(p) \cap W^u(p)$, with $q \neq p$, are called *homoclinic*. If the diffeomorphism is analytic, then $W^s(p)$ and $W^u(p)$ are analytic inmersed curves that have a transverse intersection at p. In particular, if $W^s(p)$ and $W^u(p)$ intersect at some homoclinic point q, either both curves coincide along some of their branchs (and so, the map has a separatrix) or they have a contact of finite order at q. The next theorem says that the only integrable analytic area-preserving maps with homoclinic points are the ones with separatrices. An analytic map $f : M \to M$ is called *integrable* if there exists an analytic nonconstant function $H : M \to \mathbb{R}$ such that $H \circ f = H$.

Theorem 6.1 ([7, Section 3]). *Any analytic area-preserving diffeomorphism possessing a homoclinic point with finite order contact is nonintegrable.*

The splitting of separatrices, besides nonintegrability in analytic area-preserving maps, also gives rise to positive topological entropy in surface C^1 diffeomorphisms.

Theorem 6.2 ([3, Section 2]). *Any surface C^1 diffeomorphism possessing a homoclinic point with a topological crossing (possibly with infinite order contact), has positive topological entropy.*

We can apply both theorems to our perturbed billiard maps, because those maps are analytic area-preserving diffeomorphisms defined on the annulus $\mathbb{T} \times (0, \pi)$ possesing transverse homoclinic points.

This ends the proof of Theorem 1.1.

References

1. M. Abramowitz M and I. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1972.
2. G. D. Birkhoff. *Dynamical Systems*. A. M. S. Colloquium Publications, Providence, RI, 1966 (Original ed. 1927).

3. K. Burns and H. Weiss. A geometric criterion for positive topological entropy. *Comm. Math. Phys.* **172** (1995), 95–118.
4. P. S. Casas and R. Ramírez-Ros. The frequency map for elliptic billiards. *SIAM J. Appl. Dyn. Syst.* **10** (2011), 278–324.
5. S-J. Chang and R. Friedberg. Elliptical billiards and Poncelet's theorem. *J. Math. Phys.* **29** (1988), 1537–1550.
6. K.-S. Chou and X.-P. Zhu. *The Curve Shortening Problem*. Chapman & Hall/CRC, Boca Raton, 2001.
7. R. Cushman. Examples of nonintegrable analytic Hamiltonian vectorfields with no small divisors. *Trans. Amer. Math. Soc.* **238** (1978), 45–55.
8. A. Delshams and R. Ramírez-Ros. Poincaré-Melnikov-Arnold method for analytic planar maps. *Nonlinearity* **9** (1996), 1–26.
9. A. Delshams and R. Ramírez-Ros. Melnikov potential for exact symplectic maps. *Comm. Math. Phys.* **190** (1997), 213–245.
10. M. Gage and R. S. Hamilton. The heat equation shrinking convex plane curves. *J. Differential Geom.* **23** (1986), 69–96.
11. M. A. Grayson. The heat equation shrinks embedded plane curves to round points. *J. Differential Geom.* **26** (1987), 285–314.
12. J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, volume 42 of *Applied Mathematical Sciences*. Springer-Verlag, New York, Heidelberg, Berlin, 1983.
13. V. V. Kozlov and D. Treshchëv. *Billiards: a Genetic Introduction to the Dynamics of Systems with Impacts*. Transl. Math. Monographs **89**, AMS, Providence, RI, 1991.
14. D. Jane. An example of how the Ricci flow can increase topological entropy. *Ergodic Theory Dynam. Systems* **27** (2007), 1919–1932.
15. A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Cambridge Univ. Press, Cambridge, UK, 1995.
16. V. F. Lazutkin. The existence of caustics for a billiard problem in a convex domain. *Math. USSR Izvestija* **7** (1973), 185–214.
17. S. Pinto-de-Carvalho and R. Ramírez-Ros. Non-persistence of resonant caustics in perturbed elliptic billiards. *Ergodic Theory Dynam. Systems* **33** (2013), 1876–1890.
18. R. Ramírez-Ros. Break-up of resonant invariant curves in billiards and dual billiards associated to perturbed circular tables. *Phys. D* **214** (2006), 78–87.
19. S. Tabachnikov. *Billiards*. Coll. Panorama et Synthèses, SMF, Paris, 1995.
20. E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis*. Cambridge University Press, Cambridge, UK, 1927.

Departamento de Matemática, Universidade Federal de Ouro Preto, 35.400–000, Ouro Preto, Brazil
E-mail address: josue@iceb.ufop.br

Departamento de Matemática, ICEx, Universidade Federal de Minas Gerais, 30.123–970, Belo Horizonte, Brazil
E-mail address: carneiro@mat.ufmg.br

Departament de Matemàtiques, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, Spain
E-mail address: rafael.ramirez@upc.edu