HOMOCLINIC AND HETEROCLINIC INTERSECTIONS FOR LEMON BILLIARDS

XIN JIN AND PENGFEI ZHANG

Abstract. In this paper we study the dynamical billiards on a symmetric lemon table $Q(b)$, where $Q(b)$ is the intersection of two unit disks with center distance $b$. We show that there exists $\delta_0 > 0$ such that for all $b \in (1.5, 1.5 + \delta_0)$ (except possibly a discrete subset), the billiard map $F_b$ on the lemon table $Q(b)$ admits crossing homoclinic and heteroclinic intersections. In particular, such lemon billiards have positive topological entropy.

Contents

1. Introduction ............................................. 2
2. Preliminaries ........................................... 4
   2.1. Homoclinic intersections and topological entropy .... 4
   2.2. Twist coefficients and nonlinear stability ........... 5
   2.3. Prime ends ........................................... 5
   2.4. Dynamical billiards ................................... 6
   2.5. Lemon billiards ...................................... 7
3. Elliptic and Hyperbolic periodic orbits of period 6 .......... 8
   3.1. The elliptic periodic orbit of period 6 ............... 8
   3.2. The hyperbolic periodic orbit of period 6 .......... 10
4. Periodic 6 orbits near the center of the phase space ....... 12
   4.1. The combinatorial type of the periodic orbit ....... 12
5. Trajectories with two parallel segments ................. 17
6. A generalized billiard map ................................ 27
   6.1. Basic properties of the generalized maps ......... 30
   6.2. The set of points with parallel trajectories ....... 33
7. Heteroclinic intersections and topological entropy ....... 39
   7.1. Some characterizations of the system for $b$ close to 1.5 .... 40
   7.2. Local phase portrait ................................. 43
   7.3. Hypothesis for Contradiction ...................... 46

2020 Mathematics Subject Classification. 37C83 37C29 37E30.
Key words and phrases. lemon billiards, elliptic periodic point, nonlinear stability, hyperbolic periodic point, homoclinic intersection, heteroclinic intersection, topological entropy.

X. J. is supported in part by the NSF Grant DMS-1854232.
1. INTRODUCTION

Topological entropy of a dynamical system measures the uncertainty of the system, where positive topological entropy usually means chaotic behaviors and zero topological entropy is associated with integrable behaviors. There have been many results on computing the topological entropy of a dynamical system, see [16] for a survey of related results. In this paper we will study the topological entropy of lemon billiards.

Let \( Q(b) \) be the planar domain obtained as the intersection of two unit disks \( D(O_L, 1) \) and \( D(O_R, 1) \), where \( b = |O_L O_R| \in (0, 2) \) measures the distance between their centers, see Fig. 2. The lemon shaped billiards have been studied by Heller and Tomsovic [11], in which they demonstrate a clear connection between the classical mechanics and the quantum mechanics: the eigenstates of the quantum billiards are large only at where the periodic trajectories of the classical billiards go. Numerical results in [6, Section IV] suggest that

**Conjecture 1.1.** Let \( 0 < b < 2 \) and \( F_b : M_b \to M_b \) be the billiard map on the lemon table \( Q(b) \). Then \( h_{\text{top}}(F_b) > 0 \).

There are many results about the topological entropy and the metric entropy for dynamical systems with certain geometric or topological structure for dynamical systems under certain generic conditions. For example, the existence of transverse homoclinic intersections for some hyperbolic fixed point implies that the system has positive topological entropy [17, Section 6.5], and the existence of an invariant and eventually strictly invariant cone-field implies that the system has positive metric entropy [31]. On the other hand, it is a rather difficult task to determine the entropy of a specific dynamical system beyond the above mentioned classes. One of the simplest models with chaotic behaviors is the standard map (also called Chirikov–Taylor map)

\[
f_\lambda : \mathbb{R}/2\pi \times \mathbb{R} \to \mathbb{R}/2\pi \times \mathbb{R}, \quad (x, y) \mapsto (x + y + \lambda \sin x, y + \lambda \sin x),
\]

where \( \lambda \in \mathbb{R} \) is a parameter. It is conjectured that the map \( f_\lambda \) has positive metric entropy for all \( \lambda \neq 0 \), see [28, Lecture 13]. Despite its simplicity and numerous effort for the past thirty years, this conjecture is still open and is considered to be one of most resistant problems in dynamical systems [15]. The metric entropy of the standard map is positive for sufficiently large \( \lambda \) after a super-exponentially small random perturbation [2] or after certain non-generic perturbations [1].

In this paper we obtain a partial result for Conjecture 1.1.

**Theorem 1.2.** Let \( F_b : M_b \to M_b \) be the billiard map on the lemon table \( Q(b) \). Then there exists \( \delta_0 > 0 \) such that the set \( \{ b \in (1.5, 1.5 + \delta_0) : h_{\text{top}}(F_b) = 0 \} \) has no limiting point in \((1.5, 1.5 + \delta_0)\).

In other words, \( h_{\text{top}}(F_b) > 0 \) for most \( b \in (1.5, 1.5 + \delta_0) \). See Fig. 1 for phase portraits of the lemon billiards with \( b = 1.51 \) and \( b = 1.54 \), respectively. Only the part on the right arc of the lemon table \( Q(b) \) is shown here, since the phase portrait is the same on the left arc. See (7.3) for the choice of the number \( \delta_0 \).
What we actually prove is the existence of crossing homoclinic intersections of hyperbolic periodic points of period 6 for such lemon billiards. Since the existence of crossing homoclinic intersections is an open property, it follows that the topological entropy for lemon-like billiards is also positive after we make small perturbation of the boundary $Q(b)$.

There are several classes of billiards that are hyperbolic and have positive metric entropy, see [7] for more details. The study of strictly convex billiards is rather limited. For example, it is only proved recently [32] that $C^r$-generic convex planar billiards have positive topological entropy, and the first class of hyperbolic strictly convex billiards is obtained in [4, 12]. Unlike [32], here the object, the lemon billiards, is given, and no perturbation is allowed. With limited tools, we have to combine both algebraic and geometric aspects of the lemon billiards to prove the existence of crossing homoclinic intersections. We introduce a new coordinate system under which the symmetries of the lemon billiards become more transparent. There are several places that we have to lean on the geometric meaning of certain sets of points that separate the phase space, since the algebraic formula for such sets are quite involved.

For comparison, it is well known that the standard map (1.1) has positive topological entropy. In fact, Lazutkin [18] showed that the separatrices split for all hyperbolic periodic points of a globally defined complex analytic diffeomorphism. However, Lazutkin’s result does not apply to the lemon billiards since the billiard map is not globally defined. A closely related problem is the Cohen map $F(x, y) = (y, -x + \sqrt{y^2 + 1})$. The question about the integrability of the Cohen map was first conjectured by H. Cohen and communicated by Y. Colin de Verdiere to J. Moser in 1993. Rychlik and Torgerson [26] showed that the Cohen map $F$ does not have an algebraic integral. Based on Lowther’s post [20], Cima, Gasull and Manosa [8] proved that the Cohen map is actually not $C^6$-locally integrable. It is still open whether the Cohen map has positive topological or metric entropy.

**Organization of the paper.** In Section 2, we provide some preliminary results on homoclinic and heteroclinic intersections, Moser’s Twist Mapping Theorem, Brouwer’s Fixed Point Theorem, and Mather’s results on Prime-end extensions. Some elementary results about lemon billiards are also given. In Section 3, we give explicit formulas for two periodic
orbits of period 6 that bifurcate from the periodic orbit \(O_2(b)\) of period 2 of the lemon billiards \(Q(b)\) when the parameter \(b\) crosses \(b = 1.5\). A priori, there might be other such periodic orbits. In Section 4 we obtain a characterization of the configurations of the periodic orbits of period 6 that bifurcate from \(O_2(b)\). This leads to the geometric construction in Section 5 of a family of (not necessarily periodic) orbits that can be used to connect the orbits in Section 3. In Section 6, we introduce a new coordinate system \((d_\ell, d_r)\) for the lemon billiards under which one can take advantage of the various symmetries of the lemon billiards. Finally in Section 7 we show that the periodic orbits found in Section 3 are the only ones in a neighborhood of the orbit \(O_2(b)\) for \(b\) sufficiently close to 1.5, and use this to prove the main theorem that \(h_{\text{top}}(F_b) > 0\) for most \(b\) close to 1.5.

2. Preliminaries

Let \(U \subseteq \mathbb{R}^2\) be an open neighborhood of the origin \(P = (0, 0) \in \mathbb{R}^2\), \(f : U \to \mathbb{R}^2\) be a symplectic embedding fixing the origin. The tangent map \(D_P f\) is a \(2 \times 2\) matrix with determinant 1 and two eigenvalues \(\lambda\) and \(1/\lambda\). Then the fixed point \(P\) is said to be hyperbolic (parabolic and elliptic, respectively) if \(|\lambda| > 1\) (\(\lambda = \pm 1\) and \(\lambda \in S^1 \setminus \{\pm 1\}\), respectively).

2.1. Homoclinic intersections and topological entropy. Let \(p\) be a hyperbolic fixed point of \(f\). Then the stable manifold \(W^s(p)\) and the unstable manifold \(W^u(p)\) are immersed curves in \(M\) as smooth as the map \(f\). In particular, if \(f\) is analytic, then both \(W^s(p)\) and \(W^u(p)\) are analytic immersed curves. The set \(W^s(p) \setminus \{p\}\) has two components, each of which is called a stable branch of \(p\). Consider the case that a stable branch \(\Gamma^s(p)\) of \(p\) intersects an unstable branch \(\Gamma^u(q)\) of another hyperbolic fixed point \(q\). When \(f\) is analytic, we have the following dichotomy for the intersection \(\Gamma^s(p) \cap \Gamma^u(q)\):

1. either \(\Gamma^s(p) = \Gamma^u(q)\): it is called a heteroclinic connection between \(p\) and \(q\);
2. or \(\Gamma^s(p) \cap \Gamma^u(q)\) is countable: each point in \(\Gamma^s(p) \cap \Gamma^u(q)\) is called a heteroclinic intersection between \(p\) and \(q\).

A similar distinction can be made for the case that \(q = p\), in which case we have a homoclinic loop and a homoclinic intersection, respectively. For brevity we will call it a saddle connection if it is either a homoclinic loop or a heteroclinic connection.

Poincaré discovered that the existence of a transverse homoclinic intersection of a hyperbolic fixed point leads to complicated dynamical behaviors. Later Smale [29] showed that it actually leads to the existence of an isolated hyperbolic set, the so-called horseshoe. In particular, such a dynamical system has positive topological entropy. The transversality condition can be replaced by a much weaker condition [5, Theorem 2.1].

**Proposition 2.1.** Let \(f\) be a diffeomorphism on a surface \(M\) with a hyperbolic fixed point \(p\). If a stable branch \(\Gamma^s(p) \subseteq W^s(p)\) and an unstable branch \(\Gamma^u(p) \subset W^u(p)\) has a topological crossing, then some power of \(f\) admits a full-shift factor. In particular, \(h_{\text{top}}(f) > 0\).

**Brouwer’s Fixed Point Theorem.** Let \(f : \mathbb{R}^2 \to \mathbb{R}^2\) be an orientation-preserving homeomorphism. Brouwer [3] proved the following dichotomy: either \(f\) has a fixed point, or \(f\) has no periodic point at all. A slightly stronger result is the following [10, Corollary 1.3].

**Proposition 2.2.** Let \(f : \mathbb{R}^2 \to \mathbb{R}^2\) be an orientation-preserving homeomorphism that has no fixed point. Then any topological disk \(D\) satisfying \(f(D) \cap D = \emptyset\) is a wandering domain.
The conclusion of Proposition 2.2 also holds when one replaces \( \mathbb{R}^2 \) by a topological disk, since a topological disk is homeomorphic to \( \mathbb{R}^2 \).

2.2. **Twist coefficients and nonlinear stability.** An elliptic fixed point \( P \) is said to be non-resonant if its eigenvalue \( \lambda \) satisfies \( \lambda^n \neq 1 \) for any \( n \geq 3 \). More generally, an elliptic fixed point \( P \) is non-resonant up to order \( N \) if \( \lambda^n \neq 1 \) for each \( 3 \leq n \leq N \). If \( P \) is non-resonant up to order \((2N + 2)\), then there exist an open neighborhood \( U_N \subset U \) of \( P \) and a coordinate transformation

\[
h_N : U_N \to \mathbb{R}^2, \quad \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x + p_2(x,y) + \cdots + p_{2N+1}(x,y) \\ y + q_2(x,y) + \cdots + q_{2N+1}(x,y) \end{bmatrix} + O(r^{2N+2}),
\]

where \( p_n \) and \( q_n \) are polynomials of degree \( n \) for each \( 2 \leq n \leq 2N + 1 \), such that

\[
h^{-1}_N \circ f \circ h_N \left( \begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} \cos \Theta(r^2) & -\sin \Theta(r^2) \\ \sin \Theta(r^2) & \cos \Theta(r^2) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + O(r^{2N+2}),
\]

where \( r^2 = x^2 + y^2 \), \( \Theta(r^2) = \theta + \tau_1 r^2 + \tau_2 r^4 + \cdots + \tau_N r^{2N} \), and \( \theta \) satisfies \( \lambda = e^{i\theta} \). The function \( \Theta(r^2) \) measures the amount of rotations of points around the fixed point \( P \). The coordinate transformation (2.1) is called Birkhoff transformation, the resulting form (2.2) is called Birkhoff Normal Form, and the coefficient \( \tau_k \) in the function \( \Theta(r^2) \) is called the \( k \)-th twist coefficient of \( f \) at \( P \) for each \( 1 \leq k \leq N \). See [27, 25] for more details. Recall that an elliptic fixed point \( P \) is nonlinearly stable there are nesting invariant circles accumulating on \( P \).

**Moser’s Twist Mapping Theorem.** [25, Theorem 2.13] Let \( P \) be an elliptic fixed point of \( f \) and is non-resonant up to order \((2N + 2)\). If \( \tau_k \neq 0 \) for some \( 1 \leq k \leq N \), then \( P \) is nonlinearly stable.

2.3. **Prime ends.** Let \( U \subset \mathbb{R}^2 = \mathbb{C} \) be a bounded, simply connected domain. The boundary \( \partial U \) can be very complicated. The prime-end compactification gives one way to compactify \( U \) by adding a circle to it. More precisely, pick a conformal map \( h : \mathbb{D} \to U \) from the unit disk \( \mathbb{D} \subset \mathbb{C} \) to \( U \). This induces a topology on the space \( \hat{U} = U \cup S^1 \) via the extended homeomorphism \( \hat{h} : \mathbb{D} \to \hat{U} \), such that \( \hat{h}|_\mathbb{D} = h \) and \( \hat{h}|_{S^1} = \text{Id} \). The topology on \( \hat{U} \) is independent of the choice of \( h \), and the space \( \hat{U} \) is called the prime-end compactification of \( U \). See [21, 22].

Let \( f : U \to U \) be a homeomorphism. Then there exists a unique extension \( \hat{f} : \hat{U} \to \hat{U} \), which is called the prime-end extension of \( f \). If \( f \) is orientation-preserving, then the restriction \( \hat{f}|_{S^1} \) is an orientation-preserving circle homeomorphism. Therefore, the rotation number of \( \hat{f}|_{S^1} \) is well defined. Such a number is denoted by \( \rho(f, U, \partial U) \) and is called the *Carathéodory rotation number* associated to the triple \((f, U, \partial U)\). Recall that a fixed connection is an arc \( \gamma : [0, 1] \to U \) such that \( f(\gamma) = \gamma \) and \( f(\gamma(j)) = \gamma(j), j = 0, 1 \).

**Proposition 2.3.** [21, Theorem 5.1] Let \( U \subset \mathbb{R}^2 \) be a simply connected domain with smooth boundary, \( f : U \to U \) be an orientation-preserving homeomorphism. If \( f \) has no fixed connection, then \( \rho(f, U, \partial U) \neq 0 \).

Let \( M \subset \mathbb{R}^2 \) be a simply connected domain with piecewise smooth boundary, \( f : M \to \mathbb{R}^2 \) be a symplectic diffeomorphism onto its image \( f(M) \). Let \( \text{Fix}(f) \subset M \) be the set of points
fixed by \( f \). We will call \( f \) a general symplectic map on \( M \) if it satisfies the following assumptions:

1. \(|\text{Fix}(f)| < \infty \) and each point \( p \in \text{Fix}(f) \) is either elliptic or hyperbolic;
2. every elliptic fixed point \( p \) is nonlinearly stable;
3. every hyperbolic fixed point \( p \) has positive eigenvalues.

A hyperbolic fixed point with positive eigenvalues is usually called positive hyperbolic. It is equivalent to say that the map \( f \) fixes each of the four branches of the stable and unstable manifolds of \( p \). In [21] Mather considered a more general class of maps and a more general class of fixed points. The above definition is sufficient for this paper.

Let \( L \) be an unstable branch of a positive hyperbolic fixed point \( p \). For each point \( x \in L \), consider the unstable arc \([x, fx]_u \subset L\) and the limit set \( \omega(L, f) \) of the sequence \( f^{n}[x, fx]_u, n \geq 1 \). Note that this limit set \( \omega(L, f) \) is independent of the choices of \( x \in L \) and is called the omega-set of the branch \( L \). Then \( L \) is said to be recurrent if \( L \subset \omega(L, f) \). For a stable branch \( L \), we can define the limit set \( \alpha(L, f) = \omega(L, f^{-1}) \). Then a stable branch \( L \) is said to be recurrent if \( L \subset \alpha(L, f) \).

**Proposition 2.4.** Let \( M \subset \mathbb{R}^2 \) be a simply connected domain with smooth boundary, \( f \) be a general symplectic diffeomorphism on \( M \) with no fixed connection. Then for every hyperbolic fixed point \( p \in \text{Fix}(f) \), each branch of the stable and unstable manifolds of \( p \) is recurrent.

See [21, Theorem 5.2]. A slightly stronger statement is proved in [32, Proposition 3.3].

**Proposition 2.5.** Let \( M \subset \mathbb{R}^2 \) be a simply connected domain with smooth boundary, \( f \) be a general symplectic diffeomorphism on \( M \). Then for every hyperbolic fixed point \( p \in \text{Fix}(f) \), for every branch \( L \) of the (un)stable manifold of \( p \), either \( L \) is a saddle connection, or \( L \) is recurrent.

Let \( p \) be a hyperbolic fixed point, \( U \) be a small neighborhood of \( p \). Then the local stable and unstable manifolds of \( p \) divide \( U \) into four open quadrants. Let \( K \) and \( L \) be an adjacent pair of branches of the stable and unstable manifolds of \( p \), and \( Q(K, L) \) be the open quadrant between them. Then \((K, L)\) is said to accumulate on each other on the quadrant between them if both \( p \in K \cap Q(K, L) \) and \( p \in L \cap Q(K, L) \) hold.

**Proposition 2.6.** Let \( M \subset \mathbb{R}^2 \) be a simply connected domain with smooth boundary, \( f \) be a symplectic diffeomorphism on \( M \), \( p \) be a hyperbolic fixed point of \( f \) that admits an adjacent pair of branches of its stable and unstable manifolds accumulating on each other on the quadrant between them. Then there exists a crossing homoclinic intersection of \( p \).

See [32, Theorem 4.3]. Although the property ‘crossing’ is not explicitly stated in the cited theorem, it does holds for symplectic surface maps, see the proof of [5, Theorem 3.1].

### 2.4. Dynamical billiards

Let \( Q \subset \mathbb{R}^2 \) be a connected and compact domain with (piece-wise) smooth boundary \( \partial Q \), \( |\partial Q| \) be the arc-length of \( \partial Q \) and \( s \) be an arc-length parameter of \( \partial Q \), \( 0 \leq s < |\partial Q| \) (oriented in such a way that \( Q \) is on the left side of \( \partial Q \)). Let \( \gamma(s) \in \partial Q \) and \( \gamma'(s) \) be the positive unit tangent vector of \( \partial Q \) at \( \gamma(s) \). Then for each \( 0 < \theta < \pi \), \((s, \theta)\) determines a trajectory on \( Q \) with initial position \( \gamma(s) \) and initial velocity \( R_\theta(\gamma(s)) \), where \( R_\theta \) is the rotation matrix of rotating \( \theta \) counterclockwise. Let \( \gamma(s_1) \in \partial Q \) be the first intersection of the trajectory \((s, \theta)\) with the boundary \( \partial Q \), \( \theta_1 \in [0, \pi] \) be the angle
from $\gamma(s_1)$ to the direction of the trajectory. This defines a map $F : (s, \theta) \mapsto (s_1, \theta_1)$ on the space $M = \partial Q \times [0, \pi]$, which is called a billiard map. See [7] for more details. Let $L(s, s_1) = |\gamma(s_1) - \gamma(s)|$ be the Euclidean distance between two points on the boundary $\partial Q$, which turns out to be a generating function of the billiard map [30, Chapter 3]:

$$dL = -\cos \theta \, ds + \cos \theta_1 \, ds_1.$$ 

Since $d(dL) = 0$, it follows that $\sin \theta_1 ds_1 \wedge \theta_1 = \sin \theta ds \wedge d\theta$. That is, the billiard map $F$ preserves the 2-form $\sin \theta ds \wedge d\theta$ and is symplectic. There is a convenient expression of the tangent map $DF$ when the domain $Q$ is convex [19]. More precisely, let $L(s, \theta) = L(s, s_1(s, \theta))$ be the distance depending on $(s, \theta)$, $R(s)$ the radius of curvature of $\partial Q$ at $\gamma(s)$, and $d(s, \theta) = R(s) \sin \theta$. Then the tangent map $DF$ of the billiard map $F : (s, \theta) \mapsto (s_1, \theta_1)$ is given by

$$D_{(s, \theta)}F = \frac{1}{d(s_1, \theta_1)} \begin{bmatrix} L(s, \theta) - d(s, \theta) & L(s, \theta) \\ L(s, \theta) - d(s, \theta) - d(s_1, \theta_1) & L(s, \theta) - d(s_1, \theta_1) \end{bmatrix}.$$ (2.3)

A tangent vector $v \in T_x M$ can be interpreted as a beam of trajectories $\eta : (-\epsilon, \epsilon) \to M$ with $\eta(0) = p$ and $\dot{\eta}(0) = v$. Each line $\eta(t)$ intersects $\eta(0)$ at a point $P(t)$ (could be at infinity), $-\epsilon < t < \epsilon$. Let $f^+(v)$ be the limit of the signed distance from the base point of $\eta(0)$ to $P(t)$ as $t \to 0$. Note that $f^+(v)$ is independent of the choice of the curve $\eta$ and is called the forward focusing distance of the tangent vector $v$. Similarly one can define the backward focusing distance $f^-(v)$. The mirror equation in optics states that

$$\frac{1}{f^+(v)} + \frac{1}{f^-(v)} = \frac{2}{d(x)}.$$ (2.4)

2.5. Lemon billiards. Let $Q(b)$ be the lemon table, $M_b$ be the phase space and $F_b$ be the billiard map. There exists a unique periodic orbit $O_2(b) = \{P, F_b(P)\}$ of period 2, the orbit bouncing back and forth on the table $Q(b)$ along the line $O_b O_r$. Note that $D_P F_b = D_{F_b(P)} F_b = \begin{bmatrix} 1 - b & 2 - b \\ -b & 1 - b \end{bmatrix}$. In particular, the orbit $O_2(b)$ is an elliptic periodic orbit for each $b \in (0, 2) \setminus \{1\}$. Moreover, for $1 < b < 2$, the tangent matrix $D_P F_b^2$ is conjugate to the rotation matrix $R_{\theta_b}$, where $\theta = \arccos(2(1 - b)^2 - 1)$, which decreases from $\pi$ to 0. It follows from [14] that the first twist coefficient of the billiard map $F_b$ at the elliptic periodic orbit $O_2(b)$ satisfies $\tau_1(F_b^2, P) = \frac{1}{4}$. Note that $\theta(1.5) = \frac{2\pi}{3}$, or equally, $\lambda^3 = 1$. So the periodic orbit $O_2(1.5)$ is 3-resonant. Still Moser’s theorem is applicable to this resonance case because of the symmetry of the lemon table, see [13] for more details. Therefore, the periodic point $O_2(b)$ is nonlinearly stable for every $1 < b < 2$. Pick a Diophantine number $\rho > \frac{1}{3}$ that is sufficient close to $\frac{1}{3}$ such that there exists an invariant curve of $F_{1.5}$ with rotation number $\rho$. Then there exists a positive number $\delta_\rho > 0$ such that for each $1.5 \leq b \leq 1.5 + \delta_\rho$, there exists an $F_b$-invariant curve $C_\rho(b)$ of the same rotation number $\rho$ surrounding the periodic orbit $O_2(b)$. Moreover, the invariant curve $C_\rho(b)$ depends smoothly on the parameter $b$. Denote by $D_\rho(b)$ the disk bounded by $C_\rho(b)$, which also depends continuously on the parameter $1.5 \leq b \leq 1.5 + \delta_\rho$. For convenience, we state this fact in the following

**Proposition 2.7.** Given a Diophantine number $\rho > \frac{1}{3}$ that is sufficiently close to $\frac{1}{3}$, there exists a positive number $\delta_\rho > 0$ such that for any $b \in (1.5, 1.5 + \delta_\rho)$, the billiard map $F_b$ admits an invariant curve $C_\rho(b)$ of rotation number $\rho$ surrounding the elliptic periodic orbit $O_2(b)$. 
In this section we show the existence of elliptic and hyperbolic periodic orbits of period 6 bifurcated from the periodic orbit $O_2(b)$ of period 2 when $b$ passes the value $b = 1.5$. Note that such orbits bounce alternatively between the left and the right arcs of the table $Q(b)$. Let $\phi_A(b)$ be the position angle of the corner point $A$ of the table $Q(b)$ with respect to the center $O_\ell$. It is easy to see that $\cos \phi_A(b) = \frac{b}{2}$.

3.1. The elliptic periodic orbit of period 6. Consider the periodic orbit of period 6 given in Fig. 2.

$$\text{Figure 2. A periodic orbit of period 6 on the lemon table } Q(b) \text{ for } b = 1.53.$$ 

To find an explicit formula for the periodic orbit showing in Fig. 2, we set $P_2(\cos \phi, \sin \phi)$, where $\phi = \phi(b)$ is to be determined. Because of symmetry, $P_1P_2$ is parallel to the axis $O_\ell O_r$. Then $\angle P_2O_\ell P_3 = \angle O_\ell P_2P_1$. Since $\angle O_\ell P_2P_1 = \angle O_\ell P_2P_3$, we have $|P_2P_3| = |O_\ell P_3| = b - 1$. That is, $(b - 1)^2 = (b - 1 - \cos \phi)^2 + \sin^2 \phi = (b - 1)^2 - 2(b - 1) \cos \phi + 1$.

It follows that $\cos \phi = \frac{1}{2(b-1)}$. Note that $\cos \phi < 1$ holds only if $b > 1.5$. On the other hand, $\cos \phi > \cos \phi_A(b) = \frac{b}{2}$ holds only if $b < \frac{1 + \sqrt{5}}{2}$. Therefore, the domain of existence for this periodic orbit is

$$1.5 < b < \frac{1 + \sqrt{5}}{2}. \quad (3.1)$$

Let $e_j(b) \in M_b$ be the point in the phase space corresponding to the trajectory starting at $P_j$ and pointing to $P_{j+1}$, $0 \leq j \leq 5$. This leads to a periodic orbit $O_6(b) = \{e_j(b) : 0 \leq j \leq 5\}$. One can get another periodic orbit with exactly the same properties by reversing the direction of the trajectory. We only need to consider the orbit $O_6(b)$.

Now we show that the orbit $O_6(b)$ is elliptic. For this reason we will rename it as $O'_6(b)$. Let $L_j = |P_jP_{j+1}|$ and $d_j = d(e_j(b))$ be the corresponding terms given in Section 2.4. Note that $d_1 = d_2 = d_4 = d_5 = \cos \phi = \frac{1}{2(b-1)}$, $L_0 = L_2 = L_3 = L_5 = b - 1$,

$$d_0 = d_3 = \cos 2\phi = \frac{1}{2(b-1)^2} - 1, \quad L_1 = L_4 = 2 \cos \phi - b = \frac{1}{b - 1} - b. \quad (3.2)$$
It follows that, as $2 \times 2$-matrices, $D_{e_{0}(b)}F_{b}^{6} = (D_{e_{0}(b)}F_{b}^{3})^{2}$, and
\[
D_{e_{0}(b)}F_{b}^{3} = \frac{1}{d_{0}^{2}d_{1}} \begin{bmatrix}
L_{0} - d_{1} & L_{0} \\
L_{0} - d_{1} - d_{0} & L_{0} - d_{0}
\end{bmatrix} \begin{bmatrix}
L_{1} - d_{1} & L_{1} \\
L_{1} - 2d_{1} & L_{1} - d_{1}
\end{bmatrix} \begin{bmatrix}
L_{0} - d_{0} & L_{0} \\
L_{0} - d_{0} - d_{1} & L_{0} - d_{1}
\end{bmatrix}.
\]

So it is sufficient to consider the tangent matrix $D_{e_{0}(b)}F_{b}^{3}$, for which we have
\[
f(b) := \frac{1}{2} \text{Tr}(D_{e_{0}(b)}F_{b}^{3}) = 16b^{5} - 48b^{4} + 40b^{3} - 4b^{2} - b + 1 + \frac{b}{2b^{2} - 4b + 1}. \tag{3.3}
\]

Note that $|f(b)| < 1$ for $1.5 < b < \frac{1 + \sqrt{5}}{2} \approx 1.618$. It follows that the periodic orbit is elliptic. For reasons that will be clear later, we will allow $b$ to go beyond $\frac{1 + \sqrt{5}}{2}$ and consider the corresponding periodic orbit of some algebraically defined (generalized billiard) map in Section 6. Let $b_{\text{crit}} \approx 1.63477$ be the solution of $f(b) = -1$, which is the critical parameter when the generalized orbit becomes parabolic. This parameter will come up frequently in the following sections, see Lemma 5.4 and Lemma 6.4.

![Figure 3. The graph of the function $f(b) = \text{Tr}(D_{e_{0}(b)}F_{b}^{3})/2$.](image)

Next we will show that there exists $\delta > 0$ such that the first twist coefficient $\tau_{1}(F_{b}^{6}, e_{0}(b)) \neq 0$ for each $1.5 < b < 1.5 + \delta$. Although we won’t need it, direct computation shows that $\tau_{1}(F_{b}^{6}, e_{0}(b)) \neq 0$ for all $1.5 < b < \frac{1 + \sqrt{5}}{2}$.

Due to the symmetry of the lemon table and the symmetry of the orbit $O_{0}^{6}(b)$, it suffices to compute $\tau_{1}(F_{b}^{3}, e_{0}(b))$. Note that Moser’s Theorem is applicable as long as $e_{0}(b)$ is non-resonant of order 4 for $F_{b}^{3}$, which holds whenever $f(b) \neq -\frac{1}{2}, 0$, or equally, $b \neq 1.61663, 1.6272$. The elliptic periodic orbit $O_{0}^{6}(b)$ is just a triple iterate of the orbit $O_{2}(b)$ for $b = 1.5$. It follows from [14] (see also [13]) that the first twist coefficient $\tau_{1}(F_{b}, P)$ of the one-step billiard map at the point $P$ is given by $\tau_{1}(F_{b}, P) = \frac{1}{8}$. Therefore $\tau_{1}(F_{1.5}^{3}, e_{0}(1.5)) = \frac{3}{8}$. Note that the elliptic periodic point $e_{0}(b)$ is non-resonant of order 4 for $F_{b}^{3}$ for all $1.5 < b < 1.61663$, and the first twist coefficient $\tau_{1}(F_{b}^{3}, e_{0}(b))$ depends continuously on the parameter $b$ for $1.5 \leq b < 1.61663$. In particular, there exists $\delta > 0$ such that $\tau_{1}(F_{b}^{3}, e_{0}(b)) > 0$ for each $1.5 \leq b < 1.5 + \delta$. To distinguish this constant from other constants that will be introduced later, we will denote it as $\delta_{M}$. It follows from Moser’s Twist Map Theorem that the elliptic periodic point $e_{0}(b)$ is nonlinearly stable for every $1.5 < b < 1.5 + \delta_{M}$. 

3.2. The hyperbolic periodic orbit of period 6. Consider the periodic orbit of period 6 showing in Fig. 4.

To find an explicit formula for the periodic orbit showing in Fig. 2, we set $P_0(\cos \phi, \sin \phi)$ for some $\phi \in (0, \phi_A(b))$, and $P_1(r \cos \phi, r \sin \phi)$ for some $r \in (0, 1)$. Then the remaining two points are $P_2(b - r \cos \phi, -r \sin \phi)$ and $P_3(b - \cos \phi, -\sin \phi)$. The point $P_1$ satisfies $|O_rP_1| = 1$, or equally,

$$ (r \cos \phi - b)^2 + r^2 \sin^2 \phi = r^2 - 2br \cos \phi + b^2 = 1. $$

(3.4)

Note that $\angle P_1O_\ell P_2 = \angle P_1P_2O_\ell$ and hence $|O_\ell P_1| = |P_1P_2|$, which leads to

$$ 3r^2 - 4br \cos \phi + b^2 = 0. $$

(3.5)

Eliminating $\cos \phi$ in (3.4) using (3.5), we have $r^2 = b^2 - 2$. Plugging $r = \sqrt{b^2 - 1}$ back into (3.4), we have

$$ \cos \phi = \frac{r^2 + b^2 - 1}{2br} = \frac{2b^2 - 3}{2b\sqrt{b^2 - 2}}. $$

(3.6)

This determines another periodic orbit of period 6. Note that $\cos \phi > \frac{b}{2}$ when $1.5 < b < \sqrt{3}$. So the domain of existence of this periodic orbit is

$$ 1.5 < b < \sqrt{3}. $$

(3.7)

Denote $h_j(b) \in M_b$ the point starting at $P_j$ and pointing at $P_{j+1}$, and $h_{3+j}(b) \in M_b$ the point starting at $P_{3-j}$ and pointing at $P_{2-j}$, for $0 \leq j \leq 2$. This leads to a periodic orbit $O_b(b) = \{h_j(b) : 0 \leq j \leq 5\}$. One can get another periodic orbit with exactly the same properties by reflecting the trajectory about the axis $AB$ through the two corners $A$ and $B$. We only need to consider the orbit $O_6(b)$.

Next we will show that this periodic orbit $O_b(b)$ is hyperbolic. For this reason we will rename it as $O'_6(b)$. Since $P_1 = r(\cos \phi, \sin \phi) = (b - \cos \phi_1, \sin \phi_1)$, we have

$$ \sin \phi_1 = r \sin \phi = \sqrt{b^2 - 2} \cdot \frac{\sqrt{4b^2 - 9}}{2b\sqrt{b^2 - 2}} = \sqrt{1 - \frac{9}{4b^2}}. $$

(3.8)
That is, $\cos \phi_1 = \frac{3}{2b}$. The orbit lengths are $L_0 := |P_0P_1| = 1 - r = 1 - \sqrt{b^2 - 2}$ and $L_1 := |P_1P_2| = |O,P_1| = r = \sqrt{b^2 - 2}$. Moreover, $d_0 = d_3 = 1$ while

$$d_1 = d_2 = d_4 = d_5 = \cos \angle P_0O,P_2 = \frac{2 - |P_0P_2|^2}{2}$$

$$= 1 - \frac{1}{2}(b^2 - 2(1 + r)b \cos \phi + (1 + r)^2) = \frac{1}{2\sqrt{b^2 - 2}}. \quad (3.9)$$

It follows that the tangent matrix $D_{h_0(b)}F_b^6 = (D_{h_0(b)}F_b^3)^2$, where

$$D_{h_0(b)}F_b^3 = \frac{1}{d_1^2} \begin{bmatrix} L_0 - d_1 & L_0 - 1 & L_0 - 1 & L_0 - 1 \\ L_0 - 1 - d_1 & L_0 - 1 & L_0 - 1 & L_0 - d_1 \\ L_0 - 1 - 2d_1 & L_0 - d_1 & L_0 - 1 & L_0 - d_1 \\ L_0 & L_0 - d_1 & L_0 & L_0 - d_1 \end{bmatrix}. \quad (3.10)$$

After simplifications, we have

$$\frac{1}{2} \text{Tr}(D_{h_0(b)}F_b^3) - 1 = -16b^4 + 68b^2 - 72 + (16b^4 - 64b^2 + 63)\sqrt{b^2 - 2}$$

$$= (2b - 3)(2b + 3)\sqrt{b^2 - 2} - 2(4b^2 - 7 - 4\sqrt{b^2 - 2}) > 0 \quad (3.11)$$

on the domain $1.5 < b < \sqrt{3}$. It follows that the periodic orbit $O(b)$ is hyperbolic. In fact, it says something a bit stronger. Note that the hyperbolic periodic orbit $O(b)$ can only be positive hyperbolic because of the symmetry $D_{h_0(b)}F_b^6 = (D_{h_0(b)}F_b^3)^2$. Eq. (3.11) says that, modulo the symmetry, $h_0(b)$ is already positive hyperbolic with respect to $F_b^3$.

We will show in Proposition 7.2 that there exists $\delta > 0$ such that the periodic orbits we have obtained in this section are the only periodic orbits of period 6 that are contained in a small neighborhood of the periodic orbit $O_2(b) = \{P, F_b(P)\}$ for $b \in (1.5, 1.5 + \delta_0)$. In [23] Meyer studied the generic $k$-bifurcations of periodic points for a smooth family of maps. We will only state the case for $k = 3$. More precisely, let $f_s : U \to \mathbb{R}^2$ be a family of symplectic maps fixing the origin $P = (0, 0)$, $\lambda(s)$ be an eigenvalue of $D_Pf_s$. Suppose $\lambda(0) = e^{2\pi i/3}$, where $l = 1$ or 2. Then $P$ is a 3-bifurcation point for $f_s$ at $s = 0$ if there are constants $\alpha \neq 0$ and $\gamma \neq 0$, and symplectic action-angle coordinates $(I, \phi)$ such that

$$f_s(I, \phi) = (I - \frac{2\gamma}{3}I^{3/2}\sin 3\phi + \cdots, \phi + \frac{2\pi l}{3} + \alpha s + \gamma I^{1/2}\cos 3\phi + \cdots). \quad (3.12)$$

Meyer showed that for a 3-bifurcation point $P$, there exists exactly one periodic orbit of period 3 that bifurcates from it. Moreover, this periodic orbit is hyperbolic. See also [24, Section 11.1]. Consider the periodic orbit $O_2(b) = \{P, F_b(P)\}$ for the lemon billiards. Note that $d = 1$ and $L = 2 - b$. It follows from (2.3) that

$$D_PF_b^2 = \begin{bmatrix} 2(b - 1)^2 - 1 & 2(b - 1)(b - 2) \\ 2b(b - 1) & 2(b - 1)^2 - 1 \end{bmatrix}.$$
4. Periodic 6 orbits near the center of the phase space

In this section we obtain some results about the possible configurations of periodic orbits with period 6 that are contained in a small neighborhood of the periodic orbit \( O_2(b) \) on the lemon table \( Q(b) \) for \( b \) close to 1.5. Recall that the lemon table \( Q(b) \) is the intersection of two unit disks centered at \( O_\ell = (0,0) \) and \( O_r = (b,0) \). We will call the two circular arcs of the boundary \( \partial Q(b) \) as \( \Gamma_\ell \) (the one on the left) and \( \Gamma_r \) (the one on the right), respectively. Note that there is a mismatch of the labeling between the circular arcs and their centers.

A periodic orbit of period 6 in a small neighborhood of \( O_2(b) \) has three reflections on each arc of the table and alternates between the two arcs. Denote by \( O \) \( \partial \) the boundary of the lemon table and three (possibly repeated) points \( P_1, P_2, P_3 \) (indexed from top to bottom) on the left arc \( \Gamma_\ell \) and three (possibly repeated) points \( P_1^r, P_2^r, P_3^r \) on the right arc \( \Gamma_r \) (indexed from top to bottom). Let \( \varphi_i^r \) be the position angle of the point \( P_i^r \) on the arc \( \Gamma_r \) with respect to the center \( O_\ell \), \( 1 \leq j \leq 3 \). Similarly, let \( \varphi_i^\ell \) be the position angle of the point \( P_i^\ell \) on the arc \( \Gamma_\ell \), \( 1 \leq j \leq 3 \). For convenience, we introduce \( \varphi_j^r := \pi - \varphi_j^\ell \). Then \( \varphi_j^r > 0 \) (or \( \varphi_j^\ell > 0 \)) means the point \( P_j^r \) (or \( P_j^\ell \)) has positive \( y \)-coordinate.

The segment \( O_\ell O_r \) divides the table \( Q(b) \) into two parts: the upper half and the lower half. A broken segment \( P_{j1}^\ell P_{j2}^r P_{k1}^r P_{k2}^\ell \) on the lemon table is said to be of the type \((+, -, +)\) if \( P_{j1}^\ell \) is on the lower half while both \( P_{j2}^r \) and \( P_{k1}^r \) are on the upper half of the table. Similarly one can define segments of type \((-, +, +)\). It is easy to see that trajectories on the lemon table cannot be of these two types unless all three points are on the segment \( O_\ell O_r \). The following gives a characterization of the three points on \( \Gamma_\ell \) (or \( \Gamma_r \)).

**Lemma 4.1.** The periodic orbit \( O_6 \) satisfies the following dichotomy:

(i) either \( \varphi_1^r > 0 > \varphi_3^r \) and \( \varphi_1^\ell > 0 > \varphi_3^\ell \);

(ii) or \( \varphi_j^r = \varphi_j^\ell = 0 \) for all \( 1 \leq j \leq 3 \): it is the periodic orbit \( O_2(b) \) repeated three times.

*Proof.* It suffices to prove that (ii) holds if one of the two statements in (i) fails. Without loss of generality, we assume \( \varphi_1^r > 0 > \varphi_3^r \) does not hold. Then we have either \( \varphi_3^\ell \geq 0 \) or \( \varphi_1^\ell \leq 0 \). We assume \( \varphi_3^\ell \geq 0 \). Our argument works the same in the case \( \varphi_1^\ell \leq 0 \). Note that \( \varphi_3^\ell \geq 0 \) (otherwise the two segments containing \( P_3^\ell \) would be of type \((+, -, +)\)). There are two cases: \( \varphi_3^\ell \geq \varphi_3^r \) or \( \varphi_3^\ell \leq \varphi_3^r \). We assume \( \varphi_3^\ell \geq \varphi_3^r \). Pick an index \( 1 \leq j \leq 3 \) such that \( P_j^r P_3^r \) belongs to \( O_6 \). Since \( \varphi_j^r \geq \varphi_3^r \geq \varphi_j^\ell \), the reflection at \( P_j^r \) of the trajectory \( P_j^r P_3^r \) intersects \( \Gamma_\ell \) at a point \( P_k^\ell \) with \( \varphi_k^\ell \leq \varphi_3^r \). Since \( P_3^r \) is a lowest point among all six points, it follows that \( k = 3 \) and \( \varphi_3^\ell = \varphi_3^r \). Then \( P_3^r P_3^\ell \) is part of the the orbit \( O_6 \), is horizontal and lies in the upper half of the table. Since \( P_3^r \) and \( P_3^\ell \) are the lowest points on each side, the reflection of \( P_3^r P_3^\ell \) cannot be any lower than these two points. This can only happen when \( \varphi_3^\ell = \varphi_3^r = 0 \), and \( P_3^r P_3^\ell \) is along the segment \( O_\ell O_r \). That is, \( P_3^r P_3^\ell \) is part of the orbit \( O_2(b) \) and \( O_6 \) is the periodic orbit \( O_2(b) \) (repeating three times). \( \square \)

4.1. The combinatorial type of the periodic orbit. The trajectories of the orbit \( O_6 \) from the three points on \( \Gamma_r \) to the three points on \( \Gamma_\ell \) (resp. \( \Gamma_\ell \) to \( \Gamma_r \)) gives a permutation \( \sigma_{\ell r} \in S_3 \) (resp. \( \tau_{r\ell} \in S_3 \)). Note that the composition \( \mu := \tau_{r\ell} \sigma_{\ell r} \) must be a 3-cycle since they form one periodic orbit of period 6. Therefore, \( \mu = (123) \) or \( (132) \). Alternatively, given
\(\sigma_{\ell t} \in S_3\), then \(\tau_{\ell t}\) is determined by \(\mu\) and \(\sigma_{\ell t}\). So there are \(|S_3| \times 2 = 12\) possibilities in total.

On the other hand, there are three symmetries of the lemon billiards:

1. the orientation reversal of the trajectory, which amounts to
   \[
   (\sigma_{\ell t}, \tau_{\ell t}) \mapsto (\sigma'_{\ell t} = \tau_{\ell t}^{-1}, \tau'_{\ell t} = \sigma_{\ell t}^{-1}),
   \mu \mapsto \mu' = \mu^{-1};
   \]
2. the reflection about the line \(x = \frac{b}{2}\), which amounts to exchanging \(\tau_{\ell t}\) and \(\sigma_{\ell t}\);
3. the reflection about the \(x\)-axis, which amounts to reversing the order of \(\{1, 2, 3\}\), or equally, taking conjugate by (13) on both \(\sigma_{\ell t}\) and \(\tau_{\ell t}\).

Modulo the above symmetries, we have 4 different configurations classified by the length of \(\sigma_{\ell t}\) and \(\tau_{\ell t}\) with respect to the transpositions (12) and (23) (i.e. the minimal length of the expression of \(\sigma_{\ell t}\) and \(\tau_{\ell t}\) using only (12) and (23)):

\[
(\sigma_{\ell t}, \tau_{\ell t}) = \begin{cases}
(O, \text{II}) : ((id, (123)) \sim ((123), id) \sim (id, (132)) \sim ((132), id), \\
(I, \text{I}) : ((12), (23)) \sim ((23), (12)), \\
(I, \text{III}) : ((12), (13)) \sim ((13), (12)) \sim ((23), (13)) \sim ((13), (23)), \\
(\text{II}, \text{II}) : ((123), (123)) \sim ((132), (132)).
\end{cases}
\tag{4.1}
\]

Note that certain degeneracy of \(O_6\) can happen in the above configurations: two neighboring points on \(\Gamma_t\) and/or \(\Gamma_r\) collide. The degeneracy cases are easier to dealt with. It follows directly from Lemma 4.1 that

**Lemma 4.2.** If the orbit \(O_6\) has a triple collision on either side, then \(O_6\) is the periodic orbit \(O_2(b)\) (repeating three times).

Next we consider the double collisions that have no triple collisions.

**Lemma 4.3.** Suppose \(P^\ell_1 = P^\ell_2 \neq P^\ell_3\). Then \(P^r_1 \neq P^r_2 = P^r_3\), and reflections at the points \(P^\ell_3\) and \(P^r_1\) are perpendicular.

**Proof.** Denote \(P^\ell := P^\ell_1 = P^\ell_2\). Then the two reflections at the double point \(P^\ell\) reach all three points \(P^r_j\), \(1 \leq j \leq 3\). The bisecting lines of these two reflections intersect at both \(P^\ell\) and \(O_r\) and hence coincide. It follows that either \(P^r_1 = P^r_2\) or \(P^r_2 = P^r_3\). We just need to show that the first case is impossible. Suppose on the contrary that \(P^r_1 = P^r_2 =: P^r\). Then there are two cases:

(a1) the segment \(P^\ell_3P^r\) is not part of the orbit \(O_6\); then the reflections at both \(P^\ell_3\) and \(P^r\) bounce right back and hence are perpendicular. So the points \(P^r\) and \(P^\ell\) are on the radii \(O_rP^\ell_3\) and \(O_rP^r\), respectively. It follows that all six points are contained in the lower half of the lemon table, which can only be a triple collision along the horizontal axis by Lemma 4.1, contradicting the assumption that \(P^\ell \neq P^r\).

(a2) the segment \(P^\ell_3P^r\) is part of the orbit \(O_6\); among the six trajectory segments of \(O_6\), exactly three of them are connected to \(P^\ell_3\) or \(P^r\) (maybe both). It follows that \(P^\ell P^r\) is a multiple edge with multiplicity 3. This can happen only when the reflections at both points \(P^\ell\) and \(P^r\) are perpendicular, which leads to the periodic orbit \(O_2(b)\) and hence a triple collision, contradicting the assumption that \(P^\ell \neq P^r\).

In the following we assume \(P^r := P^r_2 = P^r_3\). There are two cases:
(b1) the segment $P^r \ell_3 P^r$ is not part of the orbit $O_6$: then the trajectory $P^r \ell_3 P^r$ has a reflection at $P^r_3$ and bounces right back to $P^r$. It follows that the reflections at $P^r_3$ is perpendicular. So is the reflection at $P^r_1$.

(b2) the segment $P^r \ell_3 P^r_1$ is part of the orbit $O_6$: this case is similarly to Case (a2): there are exactly three edges connected to $P^r_3$ or $P^r_1$ (maybe both), and hence $P^r \ell P^r$ is a multiple edge with multiplicity 3. Again it leads to a triple collision, contradicting the assumption that $P^r \neq P^r_3$.

So the only possible degeneracy with double collisions is of the type with a double point and a perpendicular point on the one arc and a perpendicular point and a double point on the other arc. This completes the proof. □

An example of the above double degeneracy configuration is the hyperbolic periodic orbit $O^6_h(b)$ in Section 3.2. It can be viewed as the limit of configuration type (I, I). In the following we will deal with the non-degenerate cases.

An orbit segment $P^r \ell j_1 P^r k_1 P^r \ell j_2 P^r k_2$ on a lemon table is said to be “going-up” if $P^r \ell j_2$ is above $P^r \ell j_1$, $P^r k_2$ is above $P^r k_1$, and all four points are contained in the upper (or lower) half of lemon table. See Fig. 5 for an illustration. The illustration looks pathological and we are going to show it is impossible.

**Figure 5.** An illustration of a “going-up” orbit $P^r \ell j_1 P^r k_1 P^r \ell j_2 P^r k_2$ (red).

**Lemma 4.4.** There is no “going-up” orbit on the lemon billiards for any $1.5 \leq b \leq 1 + 2^{-1/2}$.

The number $1 + 2^{-1/2}$ will appear several times later in this paper. See the discussion right after (5.4) and the discussion right before Lemma 6.4.

**Proof.** Suppose on the contrary that $P^r \ell j_1 P^r k_1 P^r j_2 P^r k_2$ is a “going-up” orbit that is contained in the upper half of the lemon table $Q(b)$. Let $B_{\ell} = (b - 1, 0)$ and $Q_{\ell}$ be the extension of the segment $O_{\ell} P^r_{k_1}$ to the arc $\Gamma_{\ell}$. To admit such an orbit, we must have

(i) $P^r j_2$ is above $Q_{\ell}$: otherwise $P^r_{k_2}$ would be below or at $P^r k_1$ after a reflection at $P^r j_2$, a contradiction;

(ii) $P^r j_1$ is below $Q_{\ell}$, since it is the reflection of the segment $P^r j_2 P^r k_1$ with respect to the line $O_{\ell} P^r_{k_1}$.

It follows that

$$\angle O_{\ell} P^r_{k_1} Q_{\ell} < \angle O_{\ell} P^r_{k_1} P^r j_2 = \angle O_{\ell} P^r_{k_1} P^r j_1 < \angle B_{\ell} P^r_{k_1} O_{\ell}. \quad (4.2)$$
For convenience, we introduce the following three vectors:
\[ v_1 = \overrightarrow{O \ell} P_{r1}^\ell = (\cos \varphi_{r1}, -\sin \varphi_{r1}); \]
\[ v_2 = \overrightarrow{P_{r1}^\ell O \ell} = (-\cos \varphi_{r1}, -\sin \varphi_{r1}); \]
\[ v_3 = \overrightarrow{P_{r1}^\ell B \ell} = (-\cos \varphi_{r1} + b - 1, -\sin \varphi_{r1}). \]
Then it follows from (4.2) that
\[ \cos \angle \overrightarrow{O \ell P_{r1}^\ell Q^\ell} = \frac{v_1 \cdot v_2}{|v_1| \cdot |v_2|} > \cos \angle \overrightarrow{B \ell P_{r1}^\ell O \ell} = \frac{v_2 \cdot v_3}{|v_2| \cdot |v_3|}. \]
Simplifying the above expression, we obtain
\[ \frac{-1 + b \cos \varphi_{r1}}{(1 + b^2 - 2b \cos \varphi_{r1})^{1/2}} > \frac{1 - (b - 1) \cos \varphi_{r1}}{(1 + (b - 1)^2 - 2(b - 1) \cos \varphi_{r1})^{1/2}}, \]
which is equivalent to
\[ (1 - \cos^2 \varphi_{r1})(1 + 2b^2 \cos \varphi_{r1} - 2b(1 + \cos \varphi_{r1})) > 0. \] (4.3)
Since \( \cos^2 \varphi_{r1} \leq 1 \) and \( b > 1.5 \), a necessary condition for (4.3) is
\[ \cos \varphi_{r1} > \frac{2b - 1}{2b(b - 1)}. \]
However, \( \frac{2b - 1}{2b(b - 1)} \geq 1 \) for any \( 1.5 \leq b \leq 1 + 2^{-1/2} \), a contradiction. This completes the proof.

\[ \text{Figure 6. A type (II, II) configuration.} \]

**Lemma 4.5.** It is impossible to have a periodic orbit of type (II, II) on the lemon table.

**Proof.** Suppose on the contrary we have a periodic orbit of type (II, II), see Fig. 6 for an illustration. It follows from Lemma 4.1 that \( \varphi_1 > 0 > \varphi_3 \) and \( \varphi_1^l > 0 > \varphi_3^r \). Then \( \varphi_3^l > 0 \) since \( P_2^l P_1^r P_3^l \) cannot be of type \((-,-,+,-)\). It implies that \( \varphi_3 > 0 \) since \( P_1^r P_3^l P_2^l \) cannot be \((+,-,+,-)\), a contradiction. This completes the proof.

Using a similar argument, we can exclude the Type (O, II) configurations. See Fig. 7 for an illustration.

**Lemma 4.6.** It is impossible to have a periodic orbit of type (O, II) on the lemon table.
Proof. We only need to consider the configuration of the type \((id, (123))\), see the left figure in Fig. 7 for an illustration. It follows from Lemma 4.1 that \(\varphi^r_1 > 0 > \varphi^r_3\) and \(\hat{\varphi}^\ell_1 > 0 > \hat{\varphi}^\ell_3\). If \(\hat{\varphi}^\ell_2 \geq 0\), then \(\varphi^r_2 \geq 0\) (otherwise \(P^\ell_2 P^r_2 P^\ell_1\) would be of type \((+, -, +)\)). Then \(P^\ell_2 P^r_2 P^\ell_1\) is a going-up orbit, contradicting Lemma 4.4. It follows that \(\hat{\varphi}^\ell_2 < 0\). Similarly, we have \(\phi^r_2 > 0\) (otherwise \(P^\ell_3 P^r_3 P^\ell_2 P^r_2\) would be a going-up orbit, contradicting Lemma 4.4). An illustration of such an orbit on the billiard table is given in Fig. 8. Note that

(a) the distance \(d(O^\ell, P^\ell_1 P^r_2) > d(O^\ell, P^\ell_1 P^r_1)\) since \(P^r_2\) is below \(P^r_1\);
(b) the distance \(d(O^\ell, P^\ell_3 P^r_1) > d(O^\ell, P^\ell_3 P^r_3)\) since \(P^r_2\) is below \(P^r_1\) and \(P^\ell_3\) is below \(P^\ell_2\);
(c) the distances \(d(O^\ell, P^\ell_3 P^r_1) = d(O^\ell, P^\ell_1 P^r_1)\) and \(d(O^\ell, P^\ell_3 P^r_2) = d(O^\ell, P^\ell_1 P^r_2)\), since they are reflections at \(P^r_1\) and \(P^\ell_2\), respectively.

Collecting terms, we have a contradiction between (a), (b) and (c). This completes the proof. \(\Box\)

Figure 7. Type (O, II) configurations: \((id, (123))\) (left) and \((id, (132))\) (right)

Figure 8. A possible configuration of type \((id, (123))\) on the lemon table

Now consider the type (I, III) configurations, see Fig. 9 for an illustration.

**Lemma 4.7.** It is impossible to have a periodic orbit of type (I, III) on the lemon table.

Proof. By symmetry we only need to consider the configuration \(((12), (13))\), see the left-side picture in Fig. 9. It follows from Lemma 4.1 that \(\varphi^r_1 > 0 > \varphi^r_3\) and \(\hat{\varphi}^\ell_1 > 0 > \hat{\varphi}^\ell_3\). Then \(\hat{\varphi}^\ell_2 > 0\) (otherwise \(P^\ell_2 P^r_2 P^\ell_1\) would be of the type \((-+, +-)\)). Using the same argument we can show \(\varphi^r_2 > 0\). The trajectory \(P^r_2 P^\ell_3 P^r_1\) being on the upper half plane implies that \(\varphi^r_2 < \hat{\varphi}^\ell_2\). Similarly, the trajectory \(P^\ell_2 P^r_2 P^\ell_1\) being on the upper half plane implies that \(\hat{\varphi}^\ell_2 < \varphi^r_2\), a contradiction. This completes the proof. \(\Box\)
Collecting results from Lemma 4.5, 4.6, 4.7, we have excluded the possibilities of three of the four types defined in (4.1), with Type (I, I) being the only remaining type, see Fig. 10. Both periodic orbits described in Section 3 are of this type, see Fig. 2 and 4 (with double collisions $P^\ell_1 = P^\ell_2 = P_1$ and $P^r_2 = P^r_3 = P_2$).

Now consider a periodic orbit of type (I, I). It follows from Lemma 4.1 that $\varphi^r_1 > 0 > \varphi^r_3$ and $\varphi^\ell_1 > 0 > \varphi^\ell_3$. Since the segment $P^\ell_1 P^r_1$ is part of the orbit $O_6$ and the reflections at both points cannot go higher, both centers $O^\ell$ and $O^r$ are below (or on) the line $P^\ell_1 P^r_1$. Similarly, both centers $O^\ell$ and $O^r$ are above (or on) the line $P^\ell_3 P^r_3$. Let $d^\ell,j$ and $d^r,j$ be the (signed) distance from $O^\ell$ and $O^r$ to the line $P^\ell_1 P^r_1$, respectively. Similarly, we define the distances $d^\ell,j$ and $d^r,j$ from $O^\ell$ and $O^r$ to the line $P^\ell_3 P^r_3$, respectively. It follows that

$$d^\ell,j \geq 0, \quad d^r,j \geq 0, \quad j = 1, 3. \quad (4.4)$$

5. Trajectories with two parallel segments

In this section we will consider a special class of billiard trajectories as showing in Fig. 11. This construction plays an important role in Sections 6 and 7. For convenience, we denote

$$D = \left\{ (\alpha, \beta) : 0 \leq \alpha, \beta \leq \frac{\pi}{2}, \alpha + \beta < \frac{\pi}{2} \right\} \setminus \{(0, 0)\}. \quad (5.1)$$

For each point $(\alpha, \beta) \in D$, we construct a trajectory on a lemon table $Q(|O^\ell O^r|)$ with the left center $O^\ell$ at $(0, 0)$ and the to-be-determined right center $O^r$. Denote by $P^r_1$ the point on $\Gamma_r$ with coordinate angle $\alpha$ and by $P^r_3$ the point on $\Gamma_r$ with coordinate angle $-\beta$. Let $Q^\ell_1$ and $Q^r_3$ be the projections of $O^\ell$ to the horizontal lines through $P^r_1$ and $P^r_3$, respectively. See Fig. 11.
Suppose that $P_r^3Q_3$ is the post-collision trajectory of the incoming trajectory $Q_1P_1$ for some lemon table $Q(\{O_fO_r\})$ with the to-be-determined right center $O_r$. Then the intermediate reflection point $P_{\ell 2}$ is the intersection of the post-collision ray $\gamma_1(t)$ of $Q_1P_1$ and the pre-collision ray $\gamma_2(s)$ of $P_r^3Q_3$, where

$$\gamma_1(t) = (\cos \alpha - t \cos 2\alpha, \sin \alpha - t \sin 2\alpha) \quad \text{and} \quad \gamma_2(s) = (\cos \beta - s \cos 2\beta, -\sin \beta + s \sin 2\beta),$$

respectively. The intersection of the two rays $\gamma_1(t)$ and $\gamma_2(s)$ is characterized by

$$\begin{bmatrix} -\cos 2\alpha & \cos 2\beta \\ -\sin 2\alpha & -\sin 2\beta \end{bmatrix} \begin{bmatrix} t \\ s \end{bmatrix} = \begin{bmatrix} \cos \beta - \cos \alpha \\ -\sin \beta - \sin \alpha \end{bmatrix},$$

which leads to the following

$$\begin{bmatrix} t \\ s \end{bmatrix} = \frac{1}{\sin(2\alpha + 2\beta)} \begin{bmatrix} \sin(\alpha + 2\beta) - \sin \beta \\ \sin(2\alpha + \beta) - \sin \alpha \end{bmatrix} = \frac{2 \sin \frac{\alpha + \beta}{2}}{\sin(2\alpha + 2\beta)} \left[ \frac{\cos \frac{\alpha + 3\beta}{2}}{\cos \frac{3\alpha + \beta}{2}} \right]. \quad (5.2)$$

Note that $t + s = \frac{\cos \frac{\alpha + \beta}{2}}{\cos \frac{3\alpha + \beta}{2}} \geq 1$. It follows from (5.2) that the point $P_{\ell 2}^r$ is given by

$$P_{\ell 2}^r = \frac{1}{\sin(2\alpha + 2\beta)} (\sin \beta \cos 2\alpha + \sin \alpha \cos 2\beta, \sin 2\alpha \sin \beta - \sin \alpha \sin 2\beta).$$

We have the following observations:

1. the $x$-component of $P_{\ell 2}^r$ is always positive: this is clear if both $\alpha$ and $\beta$ are less than $\pi/4$. If one of them is larger than $\pi/4$, say $\beta > \pi/4$, then $\alpha < \pi/4$ and $\beta < \pi/2 - \alpha$, which implies

$$\sin \beta \cos 2\alpha + \sin \alpha \cos 2\beta \geq \sin \beta \cos 2\alpha - \sin \alpha \cos 2\alpha > 0.$$

2. for the $y$-component of $P_{\ell 2}^r$, note that

$$\sin 2\alpha \sin \beta - \sin \alpha \sin 2\beta = 2 \sin \alpha \sin \beta (\cos \alpha - \cos \beta),$$

Figure 11. The construction of an orbit segments with parallel incoming and outgoing trajectories.
which is non-negative if \( \alpha \leq \beta \), non-positive if \( \alpha \geq \beta \), and zero if either \( \alpha = 0 \), \( \beta = 0 \) or \( \alpha = \beta \).

It is possible that some of the reflection points are not contained in the table constructed. More precisely, on the domain \( D \) given in (5.1), we have

1. \( \alpha + 3\beta < \pi \) and \( 3\alpha + \beta < \pi \): both \( t > 0 \) and \( s > 0 \), and the trajectory is generated by regular reflections.
2. \( \alpha + 3\beta = \pi \): then \( t = 0 < s \). It means \( P_2^t = P_1^r \), and a reflection at \( P_1^r \) is followed immediately by a reflection at \( P_2^t \).
3. \( \alpha + 3\beta > \pi \): then \( 3\alpha + \beta < \pi \) and hence \( t < 0 < s \). It means that the trajectory travels in the opposite direction by \( |t| \) distance after a regular reflection at \( P_1^r \), have a regular reflection at \( P_2^t \) and travels in the opposite direction by \( s \) distance to reach \( P_3^r \).
4. \( 3\alpha + \beta \geq \pi \): it is similar to the previous cases.

See the green and blue segments in Fig. 12. Note that all three lines \( \alpha + \beta = \frac{\pi}{2} \), \( \alpha + 3\beta = \pi \) and \( 3\alpha + \beta = \pi \) go through the point \( (\frac{\pi}{4}, \frac{\pi}{4}) \), which is an essential singularity of our construction.

The center \( O_\ell \) lies on the ray starting at \( P_2^t \) and bisecting the angle \( \angle P_1^r P_2^t P_3^r \). Therefore, the coordinate of the center \( O_\ell \) is given by

\[
(X_{\ell r}, Y_{\ell r}) := \left[ \frac{1}{\sin(2\alpha + 2\beta)}(\sin \beta \cos 2\alpha + \sin \alpha \cos 2\beta) + \cos(\alpha - \beta) \right],
\]

(5.3)

Let \( b(\alpha, \beta) := |O_{\ell} O_r| \). After some simplification, we get

\[
b(\alpha, \beta)^2 = 1 + \frac{\sin \alpha + \sin \beta}{\sin(\alpha + \beta)} + \frac{\sin \alpha \sin \beta}{\sin^2(\alpha + \beta)} + \frac{(\sin \alpha - \sin \beta)^2}{\sin^2(2\alpha + 2\beta)}.
\]

(5.4)

Among the four terms in the function \( b^2 \), the second term is bounded from below by 1 and the sum of the last two terms is bounded from below by 1/4. It follows that \( b(\alpha, \beta) \geq 1.5 \) for \( (\alpha, \beta) \in D \). Using the linear approximation \( \sin x \approx x \), we have \( b(\alpha, \beta) \rightarrow 1.5 \) as \( (\alpha, \beta) \in D \rightarrow (0,0) \). It is easy to see that \( b(\alpha, \beta) \rightarrow \infty \) when \( \alpha + \beta \nearrow \frac{\pi}{2} \) (except the point \( (\frac{\pi}{4}, \frac{\pi}{4}) \), at which the limit is path-dependent). In fact, along the diagonal \( \alpha = \beta \), we have \( b(\alpha, \beta) = 1 + \frac{1}{2\cos \beta} \rightarrow b_{\max} := 1 + \frac{1}{\sqrt{2}} \) when \( \beta \nearrow \frac{\pi}{4} \). Note that \( b_{\max} \) also appears in Lemma 4.4 and in Section 6, see the discussion right before Lemma 6.4. See Fig. 12 for two level curves with \( b = 1.6 \) (red), \( b = b_{\max} \) (blue) and \( b = 2 \) (brown). Note that the level curve \( b^{-1}(b) \) is smooth for \( 1.5 < b < b_{\max} \) and has a singularity at the point \( (\frac{\pi}{4}, \frac{\pi}{4}) \) for \( b \geq b_{\max} \). For reasons that will clear later, we only need to consider the part of domain with \( b \leq b_{\max} \), on which we do have \( t > 0 \) and \( s > 0 \).

Remark 5.1. Note that we have excluded the origin \( (\alpha, \beta) = (0,0) \) from the definition (5.1) of the domain \( D \). In principle, one can extend our construction to the case \( (\alpha, \beta) = (0,0) \) by taking limit \( (\alpha, \beta) \in D \rightarrow (0,0) \). Note that this would correspond to the trajectory of the periodic orbit \( O_2(b) \) for one particular parameter \( b = 1.5 \). The trajectory of \( O_2(b) \) for \( b \neq 1.5 \) does not come from the above construction. We have excluded \((0,0)\) from consideration to avoid possible ambiguity.

The above construction can be viewed as a deformation between the elliptic and hyperbolic periodic orbits of period 6 in Section 3:
Lemma 5.2. The \( z \)-type segment is part of the elliptic periodic orbit \( O_e^b(b) \) when \( \alpha = \beta \), and is part of the hyperbolic periodic orbit \( O_h^b(b) \) when \( \alpha = 0 \) or \( \beta = 0 \).

Proof. Suppose \( \alpha = \beta \). Then \( b = b(\beta, \beta) = 1 + \frac{1}{2 \cos \beta} \), or equally, \( \cos \beta = \frac{1}{2(b-1)} \). This is exactly the parameter for the elliptic periodic orbit \( O_e^b(b) \) found in Section 3.1 with \( \beta = \phi \).

Next we assume \( \alpha = 0 \). Then \( b^2 = b(\beta, \beta)^2 = 2 + \frac{1}{4 \cos^2 \beta} \), or equally, \( \cos \beta = \frac{1}{2 \sqrt{b^2-2}} \). This is exactly the parameter for the hyperbolic periodic orbit \( O_h^b(b) \) found in Section 3.2 with \( \beta = \angle P_0O_P_2 \), see also Eq. (3.9). The case \( \beta = 0 \) is the same. This completes the proof. \( \square \)

Flipping the table about the horizontal line through \( O_\ell \) if necessary, we may assume \( \alpha \leq \beta \). In the following we will frequently restrict our discussion to the sub-domain \( D' = \{ (\alpha, \beta) \in D : \alpha \leq \beta \} \). All of our conclusions hold for the domain \( D \setminus D' \) with by switching the variables. One can even rotate the table and consider orbits coming from the right-hand side, see the trajectory colored in cyan in Fig. 14.

At the end of Section 4, we have introduced the following signed distances, see also (4.4). More precisely, let \( d_{\ell,1} \) and \( d_{r,1} \) be the (signed) distances from \( O_\ell \) and \( O_r \) to the line \( Q_1P_1^r \), respectively. Similarly, we define the distances \( d_{\ell,3} \) and \( d_{r,3} \) from \( O_\ell \) and \( O_r \) to the line \( P_3^rQ_3 \), respectively. Note that they are functions of \( (\alpha, \beta) \), \( d_{\ell,1} = \sin \alpha \) and \( d_{r,1} = \sin \beta \). Moreover, \( d_{\ell,1} + d_{\ell,3} = d_{r,1} + d_{r,3} \), which is the distance between the two lines \( Q_1P_1^r \) and \( P_3^rQ_3 \). Alternatively, we have

\[
d_{\ell,3} - d_{r,3} = d_{r,1} - d_{\ell,1}. \tag{5.5}
\]

This characterizes the trajectory corresponding to each \( (\alpha, \beta) \in D \).

Lemma 5.3. \( d_{\ell,1} \leq d_{r,3} \) for each \( (\alpha, \beta) \in D' \), and the equality hold if and only if one of the following holds:

1. \( \alpha = 0 \): then \( d_{\ell,1} = d_{r,3} = 0 \);
2. \( \alpha = \beta \): then \( d_{\ell,1} = d_{r,3} = \frac{1}{2} |Q_1Q_3| \).
Proof. Since \((\alpha, \beta) \in D'\), it follows from (5.3) that \(Y_{\ell r} \leq 0\) and
\[
d_{r,3} = \sin \beta + Y_{\ell r} = \sin \beta + \frac{2 \sin \alpha \sin \beta (\cos \alpha - \cos \beta)}{\sin (2\alpha + 2\beta)} + \sin (\alpha - \beta). \tag{5.6}
\]
Then the inequality \(d_{\ell,1} \leq d_{r,3}\) is equivalent to
\[
\sin (2\alpha + 2\beta) (\sin (\beta - \alpha) + \sin \alpha - \sin \beta) \leq 2 \sin \alpha \sin \beta (\cos \alpha - \cos \beta),
\]
\[
\sin (2\alpha + 2\beta) \sin \frac{\beta - \alpha}{2} (\cos \frac{\beta - \alpha}{2} - \cos \frac{\beta + \alpha}{2}) \leq 2 \sin \alpha \sin \beta \sin \frac{\beta - \alpha}{2} \sin \frac{\beta + \alpha}{2}. \tag{5.7}
\]
It is easy to see that the equality in (5.7) holds if and only if \(\alpha = 0\) or \(\alpha = \beta\). In the following we assume \(0 < \alpha < \beta\). Then the inequality (5.7) is equivalent to
\[
\sin (2\alpha + 2\beta) \sin \frac{\alpha}{2} \sin \frac{\beta}{2} \leq \sin \alpha \sin \beta \sin \frac{\beta + \alpha}{2},
\]
\[
\cos (\alpha + \beta) \cos \frac{\alpha + \beta}{2} \leq \cos \frac{\alpha}{2} \cos \frac{\beta}{2}.
\]
The last one does hold for all \((\alpha, \beta) \in D'\). This completes the proof. \(\square\)

Combining with (5.6), we have \(d_{\ell,1} \leq d_{r,3} \leq d_{\ell,3}\) for every \((\alpha, \beta) \in D'\), since \(Y_{\ell r} \leq 0\). Similarly, \(d_{\ell,1} \leq d_{r,1} \leq d_{r,3}\) since \(d_{\ell,1} + d_{\ell,3} = d_{r,1} + d_{r,3}\). There is no definite inequality between \(d_{r,1}\) and \(d_{r,3}\). More precisely, \(d_{r,1} = d_{r,3}\) if and only if \(d_{r,3} = \frac{\sin \alpha + \sin \beta}{2}\), or equally,
\[
2 \sin \alpha \sin \beta (\cos \alpha - \cos \beta) = \sin (2\alpha + 2\beta) (\sin (\beta - \alpha) + \frac{\sin \alpha - \sin \beta}{2}),
\]
\[
\sin \alpha \sin \beta = \cos (\alpha + \beta) \cos \frac{\beta + \alpha}{2} (2 \cos \frac{\beta - \alpha}{2} - \cos \frac{\beta + \alpha}{2}).
\]
See the red curve in Fig. 13 for a plot of the solutions of the above equation. It intersects the diagonal at \(\alpha \approx 0.611517\), which is slightly smaller than \(\pi/5\).

**Figure 13.** The equality \(d_{r,1} = d_{r,3}\) holds on the curve in red, while \(Y_{\ell r} = 0\) holds on the union of the curve in green and the diagonal on the domain \(D\).
For comparison, since \( \sin(2\alpha + 2\beta) > 0 \) on \( D \), the \( y \)-component of the center \( O_r \) satisfies
\[
Y_{tr} = 0 \quad \text{if and only if} \quad 2 \sin \alpha \sin \beta (\cos \alpha - \cos \beta) - \sin(2\alpha + 2\beta) \sin(\beta - \alpha) = 0,
\]
or equally,
\[
\sin \frac{\alpha + \beta}{2} \sin \frac{\alpha - \beta}{2} (\cos \alpha + \cos \beta \cos(\alpha + \beta) - \sin \alpha \sin \beta) = 0.
\]
In particular, \( Y_{tr} = 0 \) if either \( \alpha = \beta \) or
\[
F_J(\alpha, \beta) := (\cos \alpha + \cos \beta \cos(\alpha + \beta) - \sin \alpha \sin \beta) = 0.
\]
Denote by \( J \) the curve defined by (5.9). It intersects the diagonal \( \alpha = \beta \) at \( \alpha = \alpha_0 \), where \( \alpha_0 \) is the solution of \( 2 \cos \alpha \cos 2\alpha - \sin^2 \alpha = 0 \). Note that \( \alpha_0 \approx 0.663742 \), which is larger than \( \pi/5 \). We record the following simple fact for easy reference. Recall the number \( b_{\text{crit}} \approx 1.63477 \) is the solution of the equation \( f(b) = -1 \), where \( f(b) \) is given in (3.3).

**Lemma 5.4.** Let \( (\alpha_0, \alpha_0) \) be the intersection point of the curve \( J \) with the diagonal. Then \( b(\alpha_0, \alpha_0) = b_{\text{crit}}. \)

**Proof.** Note that \( \alpha_0 \) satisfies the equation \( 2 \cos \alpha \cos 2\alpha - \sin^2 \alpha = 0 \), or equally, \( x_0 = \cos \alpha_0 \) satisfies \( 4x^3 + x^2 - 2x - 1 = 0 \). On the other hand, plugging \( b = b(\alpha, \alpha) = 1 + \frac{1}{2 \cos \alpha} = 1 + \frac{1}{2x} \) into the function \( f(b) \) in Eq. (3.3), we have
\[
f(1 + \frac{1}{2x}) + 1 = -\frac{(4x^3 + x^2 - 2x - 1)(4x^4 - 2x^3 - 3x^2 + 2x + 1)}{2x^5(1 - 2x^2)}.
\]
Hence \( x_0 \) satisfies \( f(1 + \frac{1}{2x}) + 1 = 0 \). It follows that \( b(\alpha_0, \alpha_0) = 1 + \frac{1}{2x_0} = b_{\text{crit}}. \)

**Lemma 5.5.** The curve \( J \) is a smooth curve contained in \( \{(\alpha, \beta) \in D : \alpha + \beta \geq \frac{\pi}{2}\} \) with an endpoint in \( \{\beta = 0\} \) (resp. \( \{\alpha = 0\} \)) with coordinate \( (0, \frac{\pi}{2}) \) (resp. \( (0, \frac{\pi}{2}) \)). Moreover, the normal directions of \( J \) (defined by the gradient of \( F_J \) along \( J \)) is contained in the positive open cone generated by the rays \([-1, -1]\) and \([-1, 1]\) (resp. \([1, -1]\)) for \( \alpha < \beta \) (resp. \( \alpha > \beta \)).

**Proof.** It is easy to see that for each \( (\alpha, \beta) \in D \),
\[
(\partial_\alpha + \partial_\beta)F_J = \frac{1}{2}(-2 \sin(\alpha + \beta) - 3 \sin(2\alpha + \beta) - 3 \sin(\alpha + 2\beta) - \sin \alpha - \sin \beta < 0.
\]
It follows that \( \text{grad} F_J \neq 0 \) on \( D \) and hence the level curve \( J \) is smooth. Similarly, we have
\[
(\partial_\alpha - \partial_\beta)F_J = (\sin(\alpha - \beta) + (\sin \beta - \sin \alpha) \cos(\alpha + \beta)
\]
\[
= 2 \sin(\frac{\alpha - \beta}{2})(\cos(\frac{\alpha - \beta}{2}) - \cos(\frac{\alpha + \beta}{2}) \cos(\alpha + \beta)) < 0,
\]
for each \( (\alpha, \beta) \in D' \). It follows that \( \text{grad} F_J \) is contained in the cone \( \{v = (a, b) : a + b < 0, a - b < 0\} \) on \( D' \). Let \( V_F \) be the unit vector field on \([0, \frac{\pi}{2}]^2 \cap \{\alpha + \beta \leq \frac{\pi}{2}\} - \{0, 0\}\) defined by that \( \{\text{grad} F_J, V_F\} \) is everywhere orthogonal and gives the positive orientation. Then \( J \) is an integrating curve of \( V_F \) and is symmetric about the diagonal \( \{\alpha = \beta\} \). Since \( V_F \) is everywhere contained in the open positive cone generated by \([1, -1]\) and \([-1, -1]\) (resp. \([1, 1]\)) for \( \alpha < \beta \) (resp. \( \alpha > \beta \)), we get \( J \) intersects the diagonal exactly once at \( (\alpha_0, \alpha_0) \), and is contained in the upper right corner cut out by the tangent line at \( (\alpha_0, \alpha_0) \).
In Fig. 14 we have two orbits: the $\Sigma$-orbit on right side (cyan) is obtained from the $\Sigma$-orbit on the left (red) by rotating the table around the midpoint of $O_{\ell}O_{r}$ by $\pi$. So these two orbits have exactly the same characterizations, and the following result works for both orbits. For later application, we will work with the $\Sigma$-orbit. For any point $p \in M_b$, let $v_p \in T_p M_b$ be the tangent vector with backward focusing distance $f^{-}(v_p) = \infty$, see §2.4 for more details.

**Lemma 5.6.** Let $(\alpha, \beta) \in D$, $b = b(\alpha, \beta)$, $p = p(\alpha, \beta) \in M_b$ be the point with first reflection at $P^1_{\ell}$ coming from its right side, $v_p \in T_p M_b$ be the vector with $f^{-}(v_p) = \infty$, and $f^\pm_j(v_p) = f^\pm_j(DF^{-}_{b} - 1(v_p))$, $1 \leq j \leq 3$. Then $f^+_3(v_p) = \infty$ if and only if

$$F_T(\alpha, \beta) := \cos^2(\alpha + \beta)((\cos \alpha + \cos \beta) \cos(\alpha + \beta) - \sin \alpha \sin \beta) + (1 + \cos(\alpha + \beta))(\sin \alpha - \sin \beta)^2 = 0.$$  

(5.11)

**Proof.** Let $\tau_1 = |\bar{P}^1_{\ell}U|$ and $\tau_2 = |\bar{U}V|$, see Fig. 14. Suppose $f^+_3(v_p) = \infty$. It follows from (2.4) that $f^-_3(v_p) = \frac{\cos \alpha}{2}$, and

$$f^-_2(v_p) = \tau_1 - f^+_1(v_p) = \tau_1 - \frac{\cos \beta}{2},$$

$$f^+_2(v_p) = \tau_2 - f^-_3(v_p) = \tau_2 - \frac{\cos \alpha}{2}.$$  

It follows that (2.4) again, we get

$$\frac{1}{\tau_1 - \frac{\cos \beta}{2}} + \frac{1}{\tau_2 - \frac{\cos \alpha}{2}} = \frac{2}{\cos(\alpha + \beta)}.$$  

(5.12)
On the other hand, we have (similarly to Eq. (5.2), just with $\alpha$ and $\beta$ switched)
\[
\tau_1 \sin 2\beta + \tau_2 \sin 2\alpha = \sin \alpha + \sin \beta,
\cos \beta - \tau_1 \cos 2\beta = \cos \alpha - \tau_2 \cos 2\alpha,
\]
from which we solve for $\tau_1$, $\tau_2$:
\[
\begin{bmatrix}
\tau_1 \\
\tau_2
\end{bmatrix} = \frac{1}{\sin(2\alpha + 2\beta)} \begin{bmatrix}
\sin(2\alpha + \beta) - \sin \alpha \\
\sin(2\beta + \alpha) - \sin \beta
\end{bmatrix}.
\]
Plugging $\tau_1$ and $\tau_2$ into (5.12) and simplifying it, we have
\[
4 \sin^4 \left( \frac{\alpha + \beta}{2} \right) \left( 4 \cos \alpha + \cos \beta \right) \cos^3(\alpha + \beta)
- (\sin(2\alpha + \beta) + 2 \sin \alpha - 3 \sin \beta)(\sin(\alpha + 2\beta) - 3 \sin \alpha + 2 \sin \beta) = 0.
\]
In the prescribed range of $\alpha$ and $\beta$, it is equivalent to
\[
4 \cos(\alpha + \cos \beta) \cos^3(\alpha + \beta)
- (\sin(2\alpha + \beta) + 2 \sin \alpha - 3 \sin \beta)(\sin(\alpha + 2\beta) - 3 \sin \alpha + 2 \sin \beta) = 0. \tag{5.13}
\]
Using
\[
\sin(2\alpha + \beta) + 2 \sin \alpha - 3 \sin \beta = 2 \sin \alpha \cos(\alpha + \beta) + 2 \sin \alpha - 2 \sin \beta
\sin(2\alpha + \beta) + 2 \sin \beta - 3 \sin \alpha = 2 \sin \beta \cos(\alpha + \beta) + 2 \sin \beta - 2 \sin \alpha,
\]
we get exactly Eq. (5.11). \hfill \Box

Let $T$ denote for the curve defined by Eq. (5.11). It is of the shape of a “teardrop” contained in $D$, see Figure 15, in which it has been extended to be a nodal curve inside the square $[0, \frac{\pi}{2}]^2$.

Lemma 5.7. The curve $T$ is a closed curve contained in the region bounded by $J$ and $\{\alpha + \beta = \frac{\pi}{2}\}$, and intersects the curve $J$ tangentially at a single point $(\alpha_0, \alpha_0)$.

Proof. It follows from the definition of the functions $F_J$ in Eq. (5.9) and $F_T$ in Eq. (5.11) that the intersection of $J$ and $T$ happens only at their intersection with the diagonal $\alpha = \beta$, which corresponds to the point $(\alpha_0, \alpha_0)$. To show that $T$ is closed and is contained in the region bounded by $J$ and the line $\alpha + \beta = \frac{\pi}{2}$, we only need to compute the intersection of $T$ with the line $\alpha + \beta = \frac{\pi}{2}$. Then we get $\alpha = \beta$, and there is exactly one solution $(\alpha, \beta) = (\frac{\pi}{4}, \frac{\pi}{4})$. \hfill \Box

Next we will show that the level curves of the function $b = b(\alpha, \beta)$ are smooth.

Lemma 5.8. $(\partial_\alpha - \partial_\beta)b^2 > 0$ for $\alpha > \beta$ and $(\partial_\alpha - \partial_\beta)b^2 < 0$ for $\alpha < \beta$.

Proof. By direct computation, we have
\[
(\partial_\alpha - \partial_\beta)b^2 = -\frac{\sin(\alpha - \beta)}{\sin^2(\alpha + \beta)} + \frac{2(\sin \alpha - \sin \beta)(\cos \alpha + \cos \beta)}{\sin^2 2(\alpha + \beta)} + \frac{\cos \alpha - \cos \beta}{\sin(\alpha + \beta)}
= \frac{8 \sin \frac{\alpha + \beta}{2} \sin \frac{\alpha - \beta}{2}}{\sin^2(2\alpha + 2\beta)} \left( \sin \frac{3\alpha + 3\beta}{2} \cos \frac{\alpha - \beta}{2} - \sin(\alpha + \beta) \cos^2(\alpha + \beta) \right).
\]
Note that the term in the parentheses of the last term is strictly positive. Then the lemma follows. \hfill \Box
Corollary 5.9. The gradient $\nabla b^2$ is nowhere vanishing on the domain $D$, and the level curves of $b^2$ are smooth and tangent to the lines $\alpha + \beta = c$ exactly when crossing the diagonal $\alpha = \beta$.

Proof. By Lemma 5.8, only need to check if $\nabla b^2 \neq 0$ along the diagonal $\alpha = \beta$. Recall that $b^2(\beta, \beta) = (1 + \frac{1}{2\cos \beta})^2$. In particular, $(\partial_\alpha + \partial_\beta)b^2 > 0$ along the diagonal. Collecting terms, we have $\nabla b^2 \neq 0$ on $D$. The remaining properties follows from the characterizations of $(\partial_\alpha - \partial_\beta)b^2$. This completes the proof. \hfill $\square$

Recall $b_{\text{max}} = 1 + 2^{-1/2} = \lim_{\beta \to \frac{\pi}{2}} b(\beta, \beta)$ and $b_{\text{crit}} = b(\alpha_0, \alpha_0)$, see Lemma 5.4.

Lemma 5.10. The level curve $b(\alpha, \beta) = b$ does not intersect the curve $J$ for $1.5 < b < b_{\text{crit}}$, and intersects $J$ exactly once in the upper wedge $\alpha < \beta$, crosses the diagonal and then intersects $J$ exactly once in the lower wedge $\alpha > \beta$ for $b_{\text{crit}} < b < b_{\text{max}}$.

Proof. Note that a tangent vector of a level curve of $b^2$ at $(0, \beta)$ is given by $(1, -\frac{2\cos^3 \beta - 2\cos^2 \beta + 1}{1 + \cos \beta})$, for which the second component is between $-1$ and $0$. It follows from Lemma 5.8 that the minimum of $b(\alpha, \beta)$ can be only obtained along the diagonal, along which we have $b(\beta, \beta) = 1 + \frac{1}{2\cos \beta}$. Moreover, in the upper wedge $\alpha < \beta$, each level curve intersects the lines $\alpha + \beta = c$ transversely. In particular, it intersects each line $\alpha + \beta = c$ at most once. It follows that each level curve intersects $J$ at most once in the upper wedge. Then by symmetry, a level curve crosses the diagonal and then intersects the curve $J$ at most once in the lower wedge $\alpha > \beta$. \hfill $\square$
In order to obtain a better understanding of the level curves of \( b \), we will study \((\partial_\alpha + \partial_\beta) b^2\). We will consider the two parts in \( b^2 = X_{\ell r}^2 + Y_{\ell r}^2\) separately, see (5.3).

\[
(\partial_\alpha + \partial_\beta) X_{\ell r}^2 = \frac{2}{\sin(2\alpha + 2\beta)} (\cos(2\alpha + \beta) + \cos(\alpha + 2\beta)) - 4\cot(2\alpha + 2\beta)(\sin \alpha \cos 2\beta + \cos 2\alpha \sin \beta) - \sin 2\alpha \sin \beta - \sin \alpha \sin 2\beta) \\
\quad \left( \cos(\alpha - \beta) + \frac{1}{\sin(2\alpha + 2\beta)}(\sin \alpha \cos 2\beta + \cos 2\alpha \sin \beta) \right),
\]

(5.15)

\[
(\partial_\alpha + \partial_\beta) Y_{\ell r}^2 = \frac{\sin^2(\alpha - \beta) \sin(\alpha + 2\beta)}{4 \cos^2(\alpha + 2\beta) \cos(\alpha + \beta)} \left( \cos(2\alpha + 2\beta) - 4 \cos(\alpha - \beta) - 3 \cos 2\alpha - 3 \cos 2\beta + 1 \right) \left( \cos(2\alpha - \beta) + \cos(\alpha + 2\beta) - 2 \sin \alpha \sin \beta + \cos \alpha + \cos \beta \right).
\]

(5.16)

For convenience we introduce two subsets of the domain \( D \) given in Eq. (5.1):

\[
D_1 := \{(\alpha, \beta) \in D : \alpha + \beta \geq \frac{\pi}{3}\};
\]

(5.17)

\[
D_2 := \{(\alpha, \beta) \in D : (\cos \alpha + \cos \beta) \cos(\alpha + \beta) - \sin \alpha \sin \beta \leq 0\}.
\]

(5.18)

Note that \( D_2 \subset D_1 \) since \((\cos \alpha + \cos \beta) \cos(\alpha + \beta) - \sin \alpha \sin \beta \geq 0\) for \((\alpha, \beta) \in D\) with \( \alpha + \beta = \frac{\pi}{3} \).

**Lemma 5.11.** (1) \((\partial_\alpha + \partial_\beta) X_{\ell r}^2 > 0\) for every \((\alpha, \beta) \in D_1\).

(2) \((\partial_\alpha + \partial_\beta) Y_{\ell r}^2 \geq 0\) for \((\alpha, \beta) \in D_2\).

(3) \((\partial_\alpha + \partial_\beta) b^2 \geq 0\) for \((\alpha, \beta) \in D_2\).

**Proof.** (1) The expression of \((\partial_\alpha + \partial_\beta) X_{\ell r}^2\) given in (5.15) can be rewritten as

\[
(\partial_\alpha + \partial_\beta) X_{\ell r}^2 = \frac{2X_{\ell r}}{\sin(2\alpha + 2\beta)}(\mathcal{P} \mathcal{X}_1 - 4 \cot(2\alpha + 2\beta) \mathcal{P} \mathcal{X}_2 + \mathcal{P} \mathcal{X}_3),
\]

where

\[
\mathcal{P} \mathcal{X}_1 = \cos(2\alpha + \beta) + \cos(\alpha + 2\beta);
\]

\[
\mathcal{P} \mathcal{X}_2 = \sin \alpha \cos 2\beta + \cos 2\alpha \sin \beta;
\]

\[
\mathcal{P} \mathcal{X}_3 = -\sin 2\alpha \sin \beta - \sin \alpha \sin 2\beta.
\]

Note that \( X_{\ell r} > 0 \) and \( \sin(2\alpha + 2\beta) > 0 \) for any \((\alpha, \beta) \in D\). To prove \((\partial_\alpha + \partial_\beta) X_{\ell r}^2 \geq 0\), it suffices to show that \( \mathcal{P} \mathcal{X}_1 - 4 \cot(2\alpha + 2\beta) \mathcal{P} \mathcal{X}_2 + \mathcal{P} \mathcal{X}_3 \geq 0\).

Note that for any \((\alpha, \beta) \in D_1\),

\[
(\partial_\alpha - \partial_\beta) \mathcal{P} \mathcal{X}_1 = \sin(\alpha + 2\beta) - \sin(2\alpha + \beta) = 2 \cos \frac{3(\alpha + \beta)}{2} \sin \frac{\beta - \alpha}{2};
\]

\[
(\partial_\alpha - \partial_\beta) \mathcal{P} \mathcal{X}_2 = -\cos 2\alpha \cos \beta + \cos \alpha \cos 2\beta + 4 \sin \alpha \sin \beta(\cos \beta - \cos \alpha) \\
= (\cos \beta - \cos \alpha)(2 \cos \beta \cos \alpha + 1 + 4 \sin \alpha \sin \beta) \\
= (\cos \beta - \cos \alpha)(1 + 3 \cos(\alpha - \beta) - \cos(\alpha + \beta));
\]

\[
(\partial_\alpha - \partial_\beta) \mathcal{P} \mathcal{X}_3 = 2 \left( \sin \alpha \cos 2\beta - \cos 2\alpha \sin \beta \right) + \left( \sin 2\alpha \cos \beta - \cos \alpha \sin 2\beta \right) \\
= 2(\sin \alpha - \sin \beta)(3 + 4 \sin \alpha \sin \beta).
\]
It follows that, for each $j = 1, 2, 3$, $(\partial_\alpha - \partial_\beta)\mathcal{P}X_j < 0$ when $\alpha < \beta$ and $> 0$ when $\alpha > \beta$. Therefore, the minimum of $\mathcal{P}X_j$ along the line $\alpha + \beta = \text{cst}$ is achieved at $\alpha = \beta$. Since $\cot(2\alpha + 2\beta)$ is a negative constant along the line $\alpha + \beta = \text{cst}$, the minimum of $\mathcal{P}X_1 - 4\cot(2\alpha + 2\beta)\mathcal{P}X_2 + \mathcal{P}X_3$ is achieved at $\alpha = \beta$, at where we have

$$(\mathcal{P}X_1 - 4\cot(2\alpha + 2\beta)\mathcal{P}X_2 + \mathcal{P}X_3)|_{\alpha=\beta} = 2\cos 2\beta \sin \beta \tan \beta > 0.$$ 

This completes the proof of (1).

(2) Let $(\alpha, \beta) \in D_2$, and

$$\mathcal{P}Y_1 = -4 \cos(\alpha - \beta) + \cos(2\alpha + 2\beta) - 3\cos 2\alpha - 3\cos 2\beta + 1,$$

$$\mathcal{P}Y_2 = \cos(2\alpha + \beta) + \cos(\alpha + 2\beta) - 2\sin \alpha \sin \beta + \cos \alpha + \cos \beta$$

$$= 2(\cos \alpha + \cos \beta) \cos(\alpha + \beta) - 2\sin \alpha \sin \beta.$$ 

Note that $\mathcal{P}Y_2 \leq 0$ on $D_2$ by the our choice of the domain $D_2$. From the expression (5.16), we have $(\partial_\alpha + \partial_\beta)Y_2^2 \geq 0$ if and only if $\mathcal{P}Y_1 \mathcal{P}Y_2 \geq 0$. Moreover, $(\partial_\alpha + \partial_\beta)Y_{\ell r}^2 = 0$ if and only if either $\alpha = \beta$ or $\mathcal{P}Y_1 \mathcal{P}Y_2 = 0$. Note that

$$(\partial_\alpha - \partial_\beta)\mathcal{P}Y_1 = 8\sin(\alpha - \beta) + 6\sin 2\alpha - 6\sin 2\beta$$

$$= \sin(\alpha - \beta)(8 + 12\cos(\alpha + \beta)).$$

For $\alpha < \beta$ (resp. $\alpha > \beta$) in the the region $D_2$, $(\partial_\alpha - \partial_\beta)\mathcal{P}Y_1 < 0$ (resp. $(\partial_\alpha - \partial_\beta)\mathcal{P}Y_1 > 0$), hence for fixed $\alpha + \beta$, the maximum is obtained when $\alpha = 0$ (or $\beta = 0$):

$$\mathcal{P}Y_1(0, \beta) = -4\cos \beta - 2\cos 2\beta - 2 < 0.$$ 

Therefore $\mathcal{P}Y_1 < 0$ on $D_2$.

(3) It follows directly from (1) and (2). $\square$

It follows from the above proof that for $(\alpha, \beta) \in D_2$, $(\partial_\alpha + \partial_\beta)Y_{\ell r}^2 = 0$ if and only if $Y_{\ell r} = 0$, which is equivalent to either (5.9) or $\alpha = \beta$.

6. A generalized billiard map

Let $Q(b)$ be the lemon table with $1.5 < b < 2$, $x = (p(x), v(x)) \in M_b$ be a point in the phase space of the lemon billiards, where $p(x) \in \partial Q(b)$ is the base point of $x$ and $v(x)$ is the direction of the trajectory of $x$. Let $L(x) = \{p(x) + tv(x) : t \in \mathbb{R}\}$ be the oriented line corresponding to the point $x$. Let $d_\ell(x)$ and $d_r(x)$ be the oriented distances from the center $O_\ell$ and $O_r$ to the line $L(x)$, respectively. After one reflection on the left arc, the quantity $d'_r := d_r(F_b x)$ does not change, while $d'_\ell := d_\ell(F_b x)$ changes. The case with reflections on the right arc is similar. Now we find an explicit formula for $(d'_\ell, d'_r)$ in terms of $(d_\ell, d_r)$.

For certainty, consider a trajectory starting on the right arc and reflecting on the left arc, see Fig. 16 for an illustration. Let $\varphi_\ell$ be the reflection position on the left arc, $\hat{\varphi}_\ell = \pi - \varphi_\ell$, $\theta$ be the angle from the outer normal direction at $p(\varphi_\ell)$ to the direction of the trajectory. Then

1. the angle $\beta$ from the positive horizontal direction to the direction of the trajectory is given by $\beta = \varphi_\ell + \theta$;
2. the normal direction of the trajectory is $(\cos(\beta + \frac{\pi}{2}), \sin(\beta + \frac{\pi}{2}))$;
3. the angle $\beta'$ from the positive horizontal direction to the direction of the reflected trajectory is given by $\beta' = (\varphi_\ell - \pi) - \theta$;
The normal direction of the reflected trajectory is \((\cos(\beta' + \frac{\pi}{2}), \sin(\beta' + \frac{\pi}{2}))\).

It follows that

\[
\begin{align*}
\ell_r &= (-\cos \phi, \sin \phi) \cdot (\cos(\frac{\pi}{2} + \beta), \sin(\frac{\pi}{2} + \beta)); \\
\ell'_r &= (b - \cos \phi, \sin \phi) \cdot (\cos(\frac{\pi}{2} + \beta), \sin(\frac{\pi}{2} + \beta)).
\end{align*}
\]

Note that (6.1) is consistent with \(\ell_r = -\sin \theta\) and is handy for expressing the billiard map in terms of \((\ell, \ell_r)\). Also observe that \(\sin \beta = b \cdot \ell_r - \ell'\). Then after reflecting on the left arc, \(\ell_r\) remains the same, \(\beta' = \beta - 2\theta - \pi\) and

\[
\begin{align*}
\ell'_r &= (b - \cos \phi, \sin \phi) \cdot (\cos(\frac{\pi}{2} + \beta'), \sin(\frac{\pi}{2} + \beta')) \\
&= -b \sin \beta' + \sin(\phi + \beta') = b \sin(\beta - 2\theta) - \sin \theta \\
&= b(\sin \beta(2 \cos^2 \theta - 1) - 2 \cos \beta \sin \theta \cos \theta) - \sin \theta \\
&= \frac{b}{b}(d_r - d_\ell)(2(1 - d_r^2) - 1) - 2 \sqrt{1 - \left(\frac{d_r - d_\ell}{b}\right)^2}d_r \sqrt{1 - d_r^2} + d_r \\
&= (d_r - d_\ell)(1 - 2d_r^2) - 2d_r \sqrt{(b^2 - (d_r - d_\ell)^2)(1 - d_r^2)} + d_r.
\end{align*}
\]

In other words, the reflection on the arc \(\Gamma_\ell\) can be written as \(\mathcal{L}_b(d_\ell, d_r) = (d'_\ell, d_r)\), where \(d'_\ell\) is given by (6.3). The formula for the reflection on the arc \(\Gamma_r\) can be derived in a similar way. To emphasize the difference of reflections on the left arc and the right arc, we denote it by \(\mathcal{R}_b\). Alternatively, we can use the symmetry about the lemon table between the left
arcs and the right arc. Let \( \mathcal{I}(d_\ell, d_r) = (d_\ell, d_\ell) \), which represents the reflection with respect to the diagonal. Then

\[
\mathcal{I} \circ \mathcal{R}_b = \mathcal{L}_b \circ \mathcal{I},
\]

(6.4)

It follows that \( \mathcal{R}_b(d_\ell, d_r) = (d_\ell, d'_r) \), where

\[
d'_r = (d_\ell - d_r)(1 - 2d_\ell^2) - 2d_\ell \sqrt{(b^2 - (d_\ell - d_r)^2)(1 - d_\ell^2)} + d_\ell.
\]

(6.5)

We will restrict ourselves to the set of oriented lines that intersect both circles, which implies \(-1 \leq d_s \leq 1\) for \(* \in \{\ell, r\}\). For convenience, we introduce the ambient phase space \( \mathcal{M} = [-1, 1] \times [-1, 1] \), and view \( \mathcal{L}_b \) and \( \mathcal{R}_b \) as two self-maps on \( \mathcal{M} \) defined by (6.3) and (6.5). The connections between these two maps with the billiard map are discussed in Remark 6.1.

Note that there are points in \( \mathcal{M} \) where the maps \( \mathcal{L}_b \) and \( \mathcal{R}_b \) are not defined and/or not differentiable. More precisely, the branched locus of the maps \( \mathcal{L}_b \) and \( \mathcal{R}_b \) on the domain \( \mathcal{M} \) consists of two segments \( L_b \) and \( L_{-b} \), where

\[
L_{\pm b} = \{(d_\ell, d_r) \in \mathcal{M} : d_r - d_\ell = \pm b\}.
\]

(6.6)

To find the domains on which the two maps are smooth, we note that

\[
\frac{\partial d'_\ell}{\partial d_\ell} = -(1 - 2d_\ell^2) + \frac{2d_r(1 - d_\ell^2)(d_\ell - d_r)}{\sqrt{(b^2 - (d_\ell - d_r)^2)(1 - d_\ell^2)}},
\]

(6.7)

\[
\frac{\partial d'_r}{\partial d_r} = -(1 - 2d_r^2) + \frac{2d_\ell(1 - d_r^2)(d_r - d_\ell)}{\sqrt{(b^2 - (d_\ell - d_r)^2)(1 - d_r^2)}}.
\]

(6.8)

It follows that \( \frac{\partial d'_r}{\partial d_\ell} = 0 \) when

\[
(1 - 2d_\ell^2)(b^2 - (d_\ell - d_r)^2) = 4d_\ell^2(1 - d_\ell^2)(d_r - d_\ell)^2,
\]

which is equivalent to \((d_r - d_\ell)^2 = b^2(1 - 2d_\ell^2)^2\), or equally, \(d_\ell = \pm b(1 - 2d_\ell^2)^2 + d_r\). Moreover, plugging in \(d_\ell = b(1 - 2d_\ell^2) + d_r\) into the right-hand-side of (6.7), we have \(d_r > 0\). Similar, plugging in \(d_\ell = -b(1 - 2d_\ell^2) + d_r\) into the right-hand-side of (6.7), we have \(d_r < 0\). So the map \( \mathcal{L}_b \) is singular at the following curves:

\[
S_{1, \pm}(b) := \{(d_\ell, d_r) \in \mathcal{M} : \pm d_r > 0, d_\ell - d_r = \pm b(1 - 2d_\ell^2)\}.
\]

(6.9)

Similarly, the map \( \mathcal{R}_b \) is singular at the following curves:

\[
S_{2, \pm}(b) = \{(d_\ell, d_r) \in \mathcal{M} : \pm d_\ell > 0, d_r - d_\ell = \pm b(1 - 2d_\ell^2)\}.
\]

(6.10)

Note that the two curves \( S_{j, +}(b) \) and \( S_{j, -}(b) \) are center-symmetric, \( j = 1, 2 \), and \( \mathcal{I}(S_{1, \pm}(b)) = S_{2, \pm}(b) \). Moreover, \( S_{1, +}(b) \cap S_{2, +}(b) = \{S_0\} \), where \( S_0 = (2^{-1/2}, 2^{-1/2}) \) is independent of the parameter \( b \). Let \( \mathcal{D}_b \subset \mathcal{M} \) be the connected domain of \( \mathcal{M} \setminus (L_{\pm b} \cup S_{1, \pm}(b) \cup S_{2, \pm}(b)) \) containing the origin \((0, 0)\). Then both maps \( \mathcal{L}_b \) and \( \mathcal{R}_b \) are well-defined and are nonsingular on the domain \( \mathcal{D}_b \). Individually, either map can be defined on domains larger than \( \mathcal{D}_b \). The reason that we restrict to \( \mathcal{D}_b \) is that we are interested in dynamically invariant subsets (for example, the stable and unstable manifolds of the hyperbolic periodic points of period 6) on which both \( \mathcal{L}_b \) and \( \mathcal{R}_b \) are well defined.

**Remark 6.1.** The two maps \( \mathcal{L}_b \) and \( \mathcal{R}_b \) are induced from the billiard map \( F_b \) on a lemon table \( Q(b) \). However, the connection between the maps is subtle. For example, the restricted domain \( \mathcal{D}_b \) is different from the admissible phase space of the billiard map. Moreover, the
composition $\mathcal{R}_b \circ \mathcal{L}_b$, when exist, is not always the same as the composition of the billiard map. This is due to the fact that $\mathcal{R}_b \circ \mathcal{L}_b$ is the map that reflects on the left and right arc of the table in an alternating manner (even if the reflection point is outside the lemon table) while the billiard map can have multiple consecutive reflections on one side of the table. It is important to note that they do agree if the corresponding billiard trajectory reflects alternatively on the two arcs $\Gamma_\ell$ and $\Gamma_r$ of $\partial Q(b)$. More precisely, let $\text{Cor}_b \subset M$ be the set of oriented lines intersecting one of the two corners $A$ or $B$ on the billiard table $Q(b)$, see the dashed curve in Fig. 17 and Fig. 18. Then the curve $\text{Cor}_b \subset M$ separates the oriented lines intersecting both arcs of $\partial Q(b)$ from those intersecting only one arc of $\partial Q(b)$.

6.1. Basic properties of the generalized maps. Below we list some basic properties of the maps $\mathcal{L}_b$ and $\mathcal{R}_b$:

**Proposition 6.2.** $\mathcal{L}_b \circ \mathcal{L}_b = Id$ and $\mathcal{R}_b \circ \mathcal{R}_b = Id$. Moreover, $\mathcal{L}_b(S_{1,\pm}(b)) \subset L_{\pm b}$ and $\mathcal{R}_b(S_{2,\pm}(b)) \subset L_{\mp b}$.

**Proof.** Given a point $(d_\ell, d_r) \in \mathcal{R}_b$, $\mathcal{L}_b \circ \mathcal{L}_b(d_\ell, d_r)$ just means that the second reflection is at the same arc and undoes the first reflection. So it recovers the initial point $(d_\ell, d_r)$. The reasoning for $\mathcal{R}_b \circ \mathcal{R}_b = Id$ is the same.

The property $\mathcal{L}_b(S_{1,\pm}(b)) \subset L_{\pm b}$ is equivalent to $S_{1,\pm}(b) \subset \mathcal{L}_b(L_{\pm b})$, which follows from direction calculation. The reasoning for $\mathcal{R}_b(S_{2,\pm}(b)) \subset L_{\mp b}$ is similar. $\square$

The symmetries on the lemon billiard map $F_b$ lead to symmetries of the generalized maps $\mathcal{L}_b$ and $\mathcal{R}_b$:

1. $F_b$ commutes with the reflection of the table $Q(b)$ about the $x$-axis, which implies that $\mathcal{L}_b$ and $\mathcal{R}_b$ commute with the center symmetry $(d_\ell, d_r) \mapsto (-d_\ell, -d_r)$;
2. $F_b$ commutes with the reflection of the table $Q(b)$ about the axis through the two corners of $Q(b)$, which implies that $\mathcal{L}_b$ and $\mathcal{R}_b$ commute with $(d_\ell, d_r) \mapsto (-d_r, -d_\ell)$.
Figure 18. Phase portrait of the generalized billiard maps $L_b$ and $R_b$ for $b = 1.56$.

(3) combining these two symmetries, we get that $L_b$ and $R_b$ commute with the reflection $I : (d_\ell, d_r) \mapsto (d_r, d_\ell)$.

Note that a point $(d_\ell, d_r) \in M$ is fixed by $L_b$ if and only if

$$d_\ell = (d_r - d_\ell)(1 - 2d_r^2) - 2d_r \sqrt{(b^2 - (d_r - d_\ell)^2)(1 - d_r^2)} + d_r,$$

or equally, $d_\ell = (1 - b)d_r$. Such a point corresponds to a billiard trajectory intersecting the arc $\Gamma_\ell$ at the point $B_\ell = (b - 1, 0)$ on the table $Q(b)$. Similarly, a point $(d_\ell, d_r) \in M$ is fixed by $R_b$ if and only if $d_r = (1 - b)d_\ell$, which corresponds to a billiard trajectory intersecting the arc $\Gamma_r$ at the point $B_r = (1, 0)$ on the table $Q(b)$.

**Remark 6.3.** It is easy to see that in the new phase space $M$, there are

(1) two elliptic periodic orbits of period 6: one is given by

(a) $E_0 = \left(\left(1 - \frac{1}{4(b-1)^2}\right)^{1/2}, \left(1 - \frac{1}{4(b-1)^2}\right)^{1/2}\right)$,
(b) $E_1 = L_b(E_0) = \left(\frac{1}{1-b}\left(1 - \frac{1}{4(b-1)^2}\right)^{1/2}, \left(1 - \frac{1}{4(b-1)^2}\right)^{1/2}\right) \in \text{Fix}(R_b)$;
(c) $E_2 = R_b(E_1) = E_1$ since $S_1 \in \text{Fix}(R_b)$;
(d) $E_3 = L_b(E_2) = E_0$ back to $E_0$;
(e) $E_4 = R_b(E_0) = \left(\left(1 - \frac{1}{4(b-1)^2}\right)^{1/2}, \frac{1}{1-b}\left(1 - \frac{1}{4(b-1)^2}\right)^{1/2}\right) \in \text{Fix}(L_b)$;
(f) $E_5 = L_b(E_4) = S_4$ since $S_4 \in \text{Fix}(L_b)$;

the other one is the center-symmetric image of the first one. They correspond to the periodic points shown in Fig. 2. See also Fig. 17;

(2) one hyperbolic periodic points of period 6:

(a) $P_0 = (0, \left(1 - \frac{1}{4b^2-8}\right)^{1/2})$,
(b) $Q_1 = L_b(P_0) = \left(-\left(1 - \frac{1}{4b^2-8}\right)^{1/2}, \left(1 - \frac{1}{4b^2-8}\right)^{1/2}\right)$,
Using the reflection symmetry (6.4) we have

\[ P_1 = R_b(Q_1) = \left( -\left( 1 - \frac{1}{4b-3} \right)^{1/2}, 0 \right), \]

\[ Q_0 = \mathcal{L}_b(P_1) = \left( \left( 1 - \frac{1}{4b-3} \right)^{1/2}, 0 \right), \]

\[ P_2 = R_b(Q_0) = \left( \left( 1 - \frac{1}{4b-3} \right)^{1/2}, -\left( 1 - \frac{1}{4b-3} \right)^{1/2}, \right), \]

\[ Q_2 = \mathcal{L}_b(P_2) = (0, -\left( 1 - \frac{1}{4b-3} \right)^{1/2}). \]

It correspond to the periodic points shown in Fig. 4. The indices of the hyperbolic periodic orbit may looks strange. It is because that we will mainly use the two points \( P_0 \) and \( Q_0 \) that are contained in the first quadrant, and the above indexing is compatible in the sense \( P_j = (R_b \mathcal{L}_b)^j P_0 \) and \( Q_j = (R_b \mathcal{L}_b)^j Q_0, \ j \in \mathbb{Z}/3\mathbb{Z}. \) See also Fig. 17.

For convenience, we introduce short notations for some frequently used combinations:

\[ \Phi_b := \mathcal{L}_b R_b \mathcal{L}_b, \quad \Psi_b := \mathcal{R}_b \mathcal{L}_b \mathcal{R}_b, \quad \Theta_b := (\mathcal{R}_b \mathcal{L}_b)^3 = \Psi_b \Phi_b. \quad (6.11) \]

Note that a point \((d_\ell, d_r) \in \mathcal{M}\) is fixed by \( \Phi_b \) if and only if \( R_b \mathcal{L}_b(d_\ell, d_r) = \mathcal{L}_b(d_\ell, d_r). \) That is, \( \mathcal{L}_b(d_\ell, d_r) \in \text{Fix}(\mathcal{R}_b). \) It follows that \( \text{Fix}(\Phi_b) \) is the graph of the following function

\[ d_\ell = \frac{d_r}{b-1} \left( -1 + 2b(1 - d_\ell^2) - 2b\sqrt{((1-b)^2 - d_\ell^2)(1-d_\ell^2)} \right), \quad 1 - b < d_\ell < b - 1. \quad (6.12) \]

Note that \( \text{Fix}(\Phi_b) \) intersects the diagonal \( d_\ell = d_r \) at exactly three points: the origin \((0,0)\), and elliptic points \( \pm E_0(b) \). Similarly, we have \( \text{Fix}(\Psi_b) = \mathcal{R}_b(\text{Fix}(\mathcal{L}_b)) \), which is the reflection of \( \text{Fix}(\Phi_b) \) and intersects the diagonal at the same points as \( \text{Fix}(\Phi_b) \).

Consider the periodic periodic point \( E_0(b) \). As we have seen in Section 3, the point \( E_0(b) \) corresponds to an elliptic periodic point of \( F_b \) of period 6 for \( 1.5 < b < \frac{1+\sqrt{5}}{2} \approx 1.618 \). At \( b = \frac{1+\sqrt{5}}{2} \), the corresponding billiard trajectory hits the corners of the table \( Q(b) \) and ceases to exist for \( b \geq \frac{1+\sqrt{5}}{2} \). However, this point \( E_0(b) \) continues to exist and remains as an \( \Theta_b \)-fixed point for \( b \geq \frac{1+\sqrt{5}}{2} \). At \( b_{\text{max}} := 1 + 2^{-1/2}, E_0(b_{\text{max}}) = (2^{-1/2}, 2^{-1/2}) \) is the intersection of the singular curves \( S_{1+}(b_{\text{max}}) \) and \( S_{2+}(b_{\text{max}}) \). From now on we will assume \( 1.5 < b \leq b_{\text{max}} \). Recall that the parameter \( b_{\text{crit}} \approx 1.63477 \) is the solution of \( f(b) = -1 \) (see Eq. (3.3) for the definition of \( f \)) and is also the critical level curve of the function \( \mathfrak{b}(\alpha, \beta) = b \) intersecting the curve \( \mathcal{J} \), see Lemma 5.4 and Lemma 5.10.

Lemma 6.4. The \( \Theta_b \)-fixed point \( E_0(b) \) is elliptic for \( 1.5 < b < b_{\text{crit}} \), parabolic for \( b = b_{\text{crit}} \), and hyperbolic for \( b_{\text{crit}} < b \leq b_{\text{max}} \).

Proof. Using the reflection symmetry (6.4) we have

\[ \Phi_b = \mathcal{L}_b R_b \mathcal{L}_b = \mathcal{I} R_b \mathcal{L}_b \mathcal{I} = \mathcal{I} \Psi_b \mathcal{I}, \]

\[ \Theta_b = (\mathcal{R}_b \mathcal{L}_b)^3 = \Psi_b \circ \Phi_b = (\Psi_b \mathcal{I})^2. \quad (6.14) \]
Let \( A = (a_{ij}) \) be the Jacobian matrix of \( \Psi_b = R_b \mathcal{L}_b R_b \) at \( E_0 \). Since \( \mathcal{I}(E_0) = E_0 \), we have \( D_{E_0} \Theta_b = (AJ)^2 \). So we just need to find the trace of the matrix \( \tilde{A} = AJ \). Note that
\[
\begin{align*}
a_{11} &= \frac{b(-8b^3 + 24b^2 - 20b + 3)}{b - 1}, \\
a_{12} &= -\frac{2b^2 + 4b - 1}{b - 1}, \\
a_{21} &= \frac{(8b^4 - 24b^3 + 20b^2 - 4b + 1)(8b^4 - 24b^3 + 20b^2 - 2b - 1)}{(b - 1)(2b^2 - 4b + 1)}, \\
a_{22} &= \frac{b(8b^3 - 24b^2 + 20b - 3)}{b - 1}.
\end{align*}
\]
It follows from (6.16) and (6.17) that
\[
\frac{1}{2} \text{tr} \tilde{A} = \frac{a_{12} + a_{21}}{2} = \frac{32b^7 - 160b^6 + 288b^5 - 216b^4 + 54b^3 + 2b^2 - 4b + 1}{2b^2 - 4b + 1}.
\]
Note that \( \frac{1}{2} \text{tr} \tilde{A} = f(b) \), the function given in Eq. (3.3). Let \( b_{\text{crit}} \approx 1.63477 \) be the unique solution of \( f(b) = -1 \) on the interval \( 1.5 < b < b_{\text{max}} \). Then \( -1 < f(b) < 1 \) for \( 1.5 < b < b_{\text{crit}} \) and \( f(b) < -1 \) for \( b_{\text{crit}} < b < b_{\text{max}} \). This completes the proof.

It follows from Lemma 6.4 that there are new periodic points bifurcating from the corresponding point \( E_0(b) \) for \( b_{\text{crit}} < b < b_{\text{max}} \). See Lemma 6.5 for more details. Numerical results suggest that there is no other periodic orbit of period 6 beyond the ones we have found for all \( 1.5 < b < b_{\text{crit}} \). We are able to verify this rigorously for \( 1.5 < b < 1.5 + \delta \) for some \( \delta > 0 \). See Proposition 7.2.

### 6.2. The set of points with parallel trajectories.

Let \( \mathcal{M}^+ = \{(d_\ell, d_r) \in \mathcal{M} : d_\ell \geq 0, d_r \geq 0\} \setminus \{(0,0)\} \) be the first quadrant of the phase space \( \mathcal{M} \). \( \mathcal{C}_{\Phi_b} \) be the set of points \( (d_\ell, d_r) \in \mathcal{M}^+ \) such that \( (d_\ell', d_r') := \Phi_b(d_\ell, d_r) \) satisfies \( d_\ell' - d_r' = d_r - d_\ell \), and \( \mathcal{C}_{\Psi_b} \) be the set of points \( (d_\ell, d_r) \in \mathcal{M}^+ \) such that \( (d_\ell', d_r') := \Psi_b(d_\ell, d_r) \) satisfies \( d_\ell' - d_r' = d_r - d_\ell \). Analytically, we have
\[
\begin{align*}
\mathcal{C}_{\Phi_b} &= \{(d_\ell, d_r) \in \mathcal{M}^+: (1, -1) \cdot \Phi_b(d_\ell, d_r) = d_r - d_\ell\}; \\
\mathcal{C}_{\Psi_b} &= \{(d_\ell, d_r) \in \mathcal{M}^+: (1, -1) \cdot \Psi_b(d_\ell, d_r) = d_r - d_\ell\}.
\end{align*}
\]
Geometrically, these two conditions mean the final trajectory \( (d_\ell', d_r') \) under \( \Phi_b = \mathcal{L}_b \mathcal{R}_b \mathcal{L}_b \) and \( \Psi_b = \mathcal{R}_b \mathcal{L}_b \mathcal{R}_b \) is parallel to the initial trajectory \( (d_\ell, d_r) \) on the billiard table \( \mathcal{Q}(b) \), respectively. This is exactly the set of points studied in Section 5. See also Eq. (5.5). Recall that for a given point \( (\alpha, \beta) \) in the domain \( D \) defined in (5.1), consider the corresponding parallel orbit coming from right (with angle \( \beta \) on top) on the table \( \mathcal{Q}(b) \), where \( b = b(\alpha, \beta) \) is given in (5.4). See Fig. 14. Consider the map on the domain \( D \) induced by the \( \mathcal{L} \)-shaped orbits:
\[
H_\mathcal{L} : D \to \mathcal{M}, \ (\alpha, \beta) \mapsto \left( \sin \beta - \sin(\beta - \alpha) + \frac{\sin 2\alpha \sin \beta - \sin \alpha \sin 2\beta}{\sin(2\alpha + 2\beta)}, \sin \beta \right).
\]
We also consider the map on the domain \( D \) induced by the \( \mathcal{S} \)-shaped orbits:
\[
H_\mathcal{S} : D \to \mathcal{M}, \ (\alpha, \beta) \mapsto \left( \sin \alpha, \sin \alpha + \sin(\beta - \alpha) - \frac{\sin 2\alpha \sin \beta - \sin \alpha \sin 2\beta}{\sin(2\alpha + 2\beta)} \right).
\]
Lemma 6.5. Let \((d'_e, d'_r) = \Phi_b(d_e, d_r)\). Then the point \((d'_e, d'_r) = H_\Sigma(\alpha, \beta)\) is characterized by

\[
d'_e - d'_r = d_r - d_e.
\]

Therefore, we have \(C_{\Phi_b} = H_\Sigma(b^{-1}(b))\). Similarly, we obtain \(C_{\Psi_b} = H_\Sigma(b^{-1}(b))\). It follows directly from the construction of the \(\Sigma\) - and \(\Sigma\)-orbits that \(\Phi_b(C_{\Phi_b}) = C_{\Phi_b}\) and \(\Psi_b(C_{\Psi_b}) = C_{\Psi_b}\), respectively. Rotating the table by \(\pi\), we have \(I(C_{\Phi_b}) = C_{\Phi_b}\) and vice versa. See Fig. 19 for illustrations of the two curves \(C_{\Phi_b}\) and \(C_{\Psi_b}\). Note that the dashed parts of both curves will be omitted from consideration in (7.4) when constructing the set \(C_{\text{prl},b}\) for \(b > b_{\text{crit}}\).

Note that both curves \(C_{\Phi_b}\) and \(C_{\Psi_b}\) end at the corresponding hyperbolic periodic points \(P_0(b)\) and \(Q_0(b)\). For example, for \((0, \beta) \in b^{-1}(b)\), we have \(b^2 = b(0, \beta)^2 = 2 + \frac{1}{4\cos^2\beta}\), and hence \(H_\Sigma(0, \beta) = H_\Sigma(0, \beta) = (0, \sin \beta) = (0, 1 - \frac{1}{4(b^2 - 2)1/2}) = P_0(b)\). In the same way we can check that

1. for \((\alpha, 0) \in b^{-1}(b)\): \(b^2 = 2 + \frac{1}{4\cos^2\alpha}\) and \(H_\Sigma(\alpha, 0) = H_\Sigma(\alpha, 0) = (\sin \alpha, 0) = Q_0(b)\),
2. for \((\alpha, \alpha) \in b^{-1}(b)\): \(b = 1 + \frac{1}{2\cos^2\alpha}\) and \(H_\Sigma(\alpha, \alpha) = H_\Sigma(\alpha, \alpha) = (\sin \alpha, \sin \alpha) = E_0(b)\).

It follows that for each \(1.5 < b < b_{\text{crit}}\), both \(C_{\Phi_b}\) and \(C_{\Psi_b}\) are simple smooth curves in the first quadrant \(\mathcal{M}^+ \subset \mathcal{M}\) connecting \(P_0\) to \(Q_0\) with a simple transverse intersection with the diagonal at the elliptic periodic point \(E_0\). Note that for \(b \in [b_{\text{crit}}, b_{\text{max}}]\), and for each point \((\alpha, \beta) \in b^{-1}(b) \cap \mathcal{J}\), the \(y\)-component of the center \(O_r\) satisfies \(Y_{\ell_r} = 0\), see Fig. 15. Then the line \(O_rO_r\) is parallel with the trajectories \(Q_1P_1^r\) and \(P_3^rQ_3\), see Fig. 11, which implies \(H_\Sigma(\alpha, \beta) \in \{d_e = d_r\}\). It follows that at \(b = b_{\text{crit}}\), the intersection of \(C_{\Phi_b}\) and \(C_{\Psi_b}\) at \(E_0\) becomes tangential while for \(b_{\text{crit}} < b \leq b_{\text{max}}\), there are three intersections between \(C_{\Phi_b}\) and \(C_{\Psi_b}\), (resp.) and the diagonal \(\{d_e = d_r\}\): the first and the last intersections, say \(E_1(b)\) and \(E_2(b)\), correspond to the intersection \(b^{-1}(b) \cap \mathcal{J}\), while the intermediate intersection is the elliptic periodic point \(E_0(b)\) corresponding to the intersection \(b^{-1}(b) \cap \{\beta = \alpha\}\).

**Lemma 6.5.** Let \(b_{\text{crit}} < b \leq b_{\text{max}}\), and \(E_1(b)\) and \(E_2(b)\) be the two intersection points of \(C_{\Phi_b}\) (or equally, \(C_{\Psi_b}\)) with the diagonal. Then \(\Phi_b(E_1(b)) = \Psi_b(E_1(b)) = E_2(b)\) and vice versa. In particular, both points \(E_1(b)\) and \(E_2(b)\) are fixed points of \(\Theta_b\).

**Proof.** Let \(b_{\text{crit}} < b \leq b_{\text{max}}\) be given. There exists a unique pair \((\alpha, \beta) \in D' \cap \mathcal{J}\) such that \(b = b(\alpha, \beta)\). Note that the curve \(\mathcal{J}\) is characterized by \(Y_{\ell_r} = 0\). In particular, the line \(O_rO_r\) is parallel with the trajectories \(Q_1P_1^r\) (corresponding to \(E_1\)) and \(P_3^rQ_3\) (corresponding to
It follows that a tangent vector of \( E_2 \). Then we have \( E_1 = (\sin \beta, \sin \beta), E_2 = (\sin \alpha, \sin \alpha) \) and \( \Psi_b(\sin \alpha, \sin \alpha) = (\sin \beta, \sin \beta) \) and vice versa. Note that the trajectory obtained by applying the reflection of the table \( Q(b) \) about the vertical line \( x = \frac{b}{2} \) is the same (just with the reversed direction). It follows that \( \Phi_b(\sin \alpha, \sin \alpha) = (\sin \beta, \sin \beta) \) and vice versa. This completes the proof by noting that \( \Theta_b = \Psi_b \Phi_b \).

We need to know a little bit more about the stable and unstable directions of the hyperbolic periodic points \( P_0(b) \) and \( Q_0(b) \). Moreover, we will find the tangent directions of the two curves \( C_{\Phi_b} \) and \( C_{\Psi_b} \) at the two endpoints \( P_0 \) and \( Q_0 \) and show that they are contained in the cone of the stable and unstable directions of the hyperbolic periodic points \( P_0 \) and \( Q_0 \) in the first quadrant. We will focus on the point \( P_0 \), since what happens at \( Q_0 \) is just a reflection about the diagonal.

**Lemma 6.6.** Let \( m_{s,u}(b) \) the slopes of the stable and unstable tangent space at \( P_0(b) \), \( m_{\Xi}(b) \) and \( m_{\Xi}(b) \) the the slopes of the two curves \( C_{\Phi_b} \) and \( C_{\Psi_b} \) at \( P_0(b) \), respectively. Then we have

\[
m_u(b) < m_{\Xi}(b) < m_{\Xi}(b) < m_s(b) < 0
\]

(6.25) for any \( 1.5 < b \leq b_{\text{max}} \).

**Proof.** The tangent vector of the level curve \( b^{-1}(b) \) at the point \( (0, \beta) \in b^{-1}(b) \) with normalized first component is given by

\[
T_b(0, \beta) = \left( 1, -\frac{2\cos^3 \beta - 2\cos^2 \beta + 1}{\sin \beta} \tan \frac{\beta}{2} \right) = \left( 1, -\frac{2\eta^3 - 2\eta^2 + 1}{\eta + 1} \right), \tag{6.26}
\]

where \( \eta = \cos \beta = \frac{1}{2(b^2 - 2)^{1/2}} \) varies from 1 to 0.522933 as \( b \) varies from 1.5 to \( b_{\text{max}} \).

1. The tangent map of \( H_{\Xi} \) at \( (0, \beta) \) is given by

\[
DH_{\Xi}(0, \beta) = \begin{bmatrix} \cos \beta + \frac{1}{\cos \beta} - 1 & 0 \\ 0 & \cos \beta \end{bmatrix} = \begin{bmatrix} \eta + \frac{1}{\eta} - 1 & 0 \\ 0 & \eta \end{bmatrix}. \tag{6.27}
\]

It follows that a tangent vector of \( C_{\Phi_b} \) at the point \( P_0 = H_{\Xi}(0, \beta) \) is given by

\[
T_{\Xi,b} := DH_{\Xi}(0, \beta)T_b(0, \beta) = \begin{bmatrix} \eta + \frac{1}{\eta} - 1 & 0 \\ 0 & \eta \end{bmatrix} \begin{bmatrix} \frac{1}{\eta^3 - 2\eta^2 + 1} \\ -\frac{2\eta^3 - 2\eta^2 + 1}{\eta + 1} \end{bmatrix} = \begin{bmatrix} \eta + \frac{1}{\eta} - 1 & 0 \\ 0 & \eta(2\eta^3 - 2\eta^2 + 1) \end{bmatrix}. \tag{6.28}
\]

For an easy comparison, we will normalize the first component of \( T_{\Xi,b} \) and rewrite it as

\[
T_{\Xi,b} = (1, m_{\Xi}(\eta)),
\]

where

\[
m_{\Xi}(\eta) = \frac{-\eta^2(2\eta^3 - 2\eta^2 + 1)}{\eta^3 + 1}. \tag{6.29}
\]

Note that \( m_{\Xi} \to -\frac{1}{2} \) when \( \eta \to 1 \). See Fig. 20 for a plot of the function \( m_{\Xi}(\eta) \).

2. The tangent map of \( H_{\Xi} \) at \( (0, \beta) \) is given by

\[
DH_{\Xi}(0, \beta) = \begin{bmatrix} \cos \beta + \frac{1}{\cos \beta} & 0 \\ 0 & -\cos \beta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\eta - \frac{1}{\eta} & \eta \end{bmatrix}. \tag{6.30}
\]

It follows that a tangent vector of \( C_{\Psi_b} \) at the point \( P_0 = H_{\Xi}(0, \beta) \) is given by

\[
T_{\Xi,b} := DH_{\Xi}(0, \beta)T_b(0, \beta) = \begin{bmatrix} 1 & 0 \\ 2 - \eta - \frac{1}{\eta} & \eta \end{bmatrix} \begin{bmatrix} \frac{1}{\eta^3 - 2\eta^2 + 1} \\ -\frac{2\eta^3 - 2\eta^2 + 1}{\eta + 1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\frac{2\eta^3 - 2\eta^2 + 1}{\eta + 1} \end{bmatrix}. \tag{6.31}
\]
Denote $T_{\Psi, b} = (1, m_\Psi(\eta))$, where

$$m_\Psi(\eta) = \frac{-2\eta^5 + 2\eta^4 - \eta^3 + \eta - 1}{\eta(\eta + 1)}. \quad (6.32)$$

Note that $m_y \to -\frac{1}{2}$ when $\eta \to 1$. See Fig. 20 for a plot of the function $m_\Psi(\eta)$.

(3) Let $A = (a_{ij})_{1 \leq i, j \leq 2} := D_{\Psi_b} \Phi_b$ be the Jacobian matrix of $\Psi_b = R_b L_b R_b$ at $I(P_0) = Q_0$. Since $\Psi_b I(P_0) = P_0$, it follows from (6.13) that $D_{\Psi_b} \Phi_b = JA J$, where $J = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Similarly, it follows from (6.14) that

$$D_{\Psi_b} \Theta_b = D_{\Psi_b} \Phi_b = A J A J = \hat{A}^2,$$

where $\hat{A} = AJ = \begin{bmatrix} a_{12} & a_{11} \\ a_{22} & a_{21} \end{bmatrix}$. Since the point $P_0$ is a hyperbolic periodic point of period 6, the two matrices $A$ and $D_{\Psi_b} \Theta_b$ share the same eigenspaces. After some simplification, we get

$$a_{11} = -\frac{2(b^2 - 2)(16b^4 - 64b^2 + 63)}{2b^2 - 3}; \quad (6.33)$$

$$a_{12} = -\frac{8b^4 - 36b^2 + 39}{2b^2 - 3}; \quad (6.34)$$

$$a_{21} = \frac{2(16b^4 - 64b^2 + 63)\sqrt{b^2 - 2} - (8b^2 - 15)(8b^4 - 32b^2 + 31)}{2b^2 - 3}; \quad (6.35)$$

$$a_{22} = \frac{8b^4 - 36b^2 + 39 - 2(8b^4 - 35b^2 + 36)}{2b^2 - 3}. \quad (6.36)$$

Using the relation $b^2 = 2 + \frac{1}{4\eta^2}$, we can rewrite the matrix $A = (a_{ij})_{1 \leq i, j \leq 2}$ as

$$\frac{1}{\eta^5(1 + 2\eta^2)} \begin{bmatrix} \eta(\eta^4 - 1) & \eta^3(2\eta^4 + 2\eta^2 - 1) \\ 2\eta^7 - 2\eta^6 + 4\eta^5 - \eta^4 - \eta^3 + 2\eta^2 - 2\eta + 1 & -4\eta^8 + 8\eta^7 - 6\eta^6 + 3\eta^5 - 2\eta^3 + \eta^2 \end{bmatrix}.$$ 

Then the stable and unstable eigenvalues $\lambda_s, u(\eta)$ and the corresponding slopes of their eigenvectors $m_s, u(\eta)$ of the matrix $\hat{A} = AJ$ are given by

$$\lambda_s(\eta) = \frac{4\eta^7 - 2\eta^6 + 6\eta^5 - \eta^4 - 2\eta^3 + 2\eta^2 - 2\eta + 1 - \sqrt{\Delta}}{2(2\eta^7 + \eta^5)}, \quad (6.37)$$

$$\lambda_u(\eta) = \frac{4\eta^7 - 2\eta^6 + 6\eta^5 - \eta^4 - 2\eta^3 + 2\eta^2 - 2\eta + 1 + \sqrt{\Delta}}{2(2\eta^7 + \eta^5)}, \quad (6.38)$$

$$m_s(\eta) = -\frac{2\eta^2(1 - \eta - \eta^2 + 2\eta^3 - 4\eta^4 + 4\eta^5)}{(1 + \eta)(1 + \eta^2)(1 - 2\eta + 2\eta^2) + (2\eta^2 + 1)\sqrt{\Gamma}}, \quad (6.39)$$

$$m_u(\eta) = -\frac{2\eta^2(1 - \eta - \eta^2 + 2\eta^3 - 4\eta^4 + 4\eta^5)}{(1 + \eta)(1 + \eta^2)(1 - 2\eta + 2\eta^2) - (2\eta^2 + 1)\sqrt{\Gamma}}, \quad (6.40)$$

where

$$\Delta = (4\eta^7 - 2\eta^6 + 6\eta^5 - \eta^4 - 2\eta^3 + 2\eta^2 - 2\eta + 1)^2 - 4\eta^{10}(4\eta^4 + 4\eta^2 + 1),$$

$$\Gamma = (1 - \eta)(4\eta^6 + 3\eta^5 + \eta^4 + 2\eta^3 - 2\eta^2 - \eta + 1).$$
See Fig. 20 for the plots of the functions \(m_s(\eta)\) and \(m_u(\eta)\). To complete the proof, it suffices to note that

(i) \(m_{\Xi}(\eta) - m_{\Xi}(\eta) = \frac{(1-\eta^2)^2}{\eta^2 + \eta}(2\eta^5 - 2\eta^4 + \eta^3 + \eta^2 + 1) > 0\) for \(0.5 < \eta < 1\).

(ii) To estimate \(m_s(\eta) - m_{\Xi}(\eta)\), note that both denominators are positive. We multiply it by the common denominator and eliminate the common factor \(\eta^2\):

\[
-2(1 - \eta - \eta^2 + 2\eta^3 - 4\eta^4 + 4\eta^5)(\eta^3 + 1) + ((1 + \eta)(1 + \eta^2)(1 - 2\eta + 2\eta^2) + (2\eta^4 + 1)\sqrt{\Gamma})(2\eta^3 - 2\eta^2 + 1) = (2\eta^2 + 1)(1 - \eta)(2\eta^5 + 3\eta^2 - 1) + (2\eta^3 - 2\eta^2 + 1)\sqrt{\Gamma}.
\]

(6.41)

Set \(p_1(\eta) := (1 - \eta)(2\eta^5 + 3\eta^2 - 1)\) and \(q_1(\eta) := (2\eta^3 - 2\eta^2 + 1)\sqrt{\Gamma}\). It is easy to see that

\[
q_1(\eta)^2 - p_1(\eta)^2 = 4\eta^3(1 - \eta^2)(\eta^3 - \eta^2 + 1)(4\eta^5 - 4\eta^4 + 2\eta^3 - \eta^2 - \eta + 1).
\]

(6.42)

Note that all factors on the right side of (6.42) are positive on the interval \((0.5, 1)\). It follows that \(|q_1(\eta)| > |p_1(\eta)|\) on the interval \((0.5, 1)\). Combining with the fact that \(p_1(0.5) = -0.09375\) and \(q_1(0.5) \approx 0.363092\), we have \(q_1(\eta) = |q_1(\eta)| > |p_1(\eta)| \geq -p_1(\eta)\) on the interval \((0.5, 1)\). It follows that \(p_1(\eta) + q_1(\eta) > 0\), or equally, \(m_s(\eta) > m_{\Xi}(\eta)\) for \(0.5 < \eta < 1\).

(iii) To estimate \(m_{\Xi}(\eta) - m_u(\eta)\), we multiply it by the common denominator:

\[
(2\eta^5 - 2\eta^4 - 3\eta^2 + \eta - 1)((1 + \eta)(1 + \eta^2)(1 - 2\eta + 2\eta^2) - (2\eta^2 + 1)\sqrt{\Gamma}) + 2\eta^2(1 - \eta - \eta^2 + 2\eta^3 - 4\eta^4 + 4\eta^5)\eta(\eta + 1) = (1 + 2\eta^2)((1 - \eta^2)(-1 + 2\eta - \eta^2 - \eta^3 + 3\eta^4 - 6\eta^5 + 2\eta^6) + (2\eta^5 - 2\eta^4 + \eta^3 - \eta + 1)\sqrt{\Gamma}.
\]

Let \(p_2(\eta) = (1 - \eta^2)(-1 + 2\eta - \eta^2 - \eta^3 + 3\eta^4 - 6\eta^5 + 2\eta^6)\) and \(q_2(\eta) = (2\eta^5 - 2\eta^4 + \eta^3 - \eta + 1)\sqrt{\Gamma}\). It is easy to see that

\[
q_2(\eta)^2 - p_2(\eta)^2 = 4\eta^7(1 - \eta^2)(\eta^3 - \eta^2 + 1)(4\eta^5 - 4\eta^4 + 2\eta^3 - \eta^2 - \eta + 1)
\]

\[
= \eta^4(q_1(\eta)^2 - p_1(\eta)^2) > 0.
\]

It follows that \(|q_2(\eta)| > |p_2(\eta)|\) on the interval \((0.5, 1)\). Combining with the fact that \(p_2(0.5) = -0.2578125\) and \(q_2(0.5) \approx 0.272319\), we have \(q_2(\eta) = |q_2(\eta)| > |p_2(\eta)| \geq -p_2(\eta)\) on the interval \((0.5, 1)\). It follows that \(p_2(\eta) + q_2(\eta) > 0\), or equally, \(m_{\Xi}(\eta) > m_u(\eta)\) for \(0.5 < \eta < 1\).

Collecting terms, we have that \(m_u(b) < m_{\Xi}(b) < m_{\Xi}(b) - m_s(b) < 0\) for \(1.5 < b < b_{\text{max}}\).

Let \(C(E_{P_b}^c, E_{P_b}^s)\) be the cone in the plane \(T_{P_b}M\) from \(E_{P_b}^n\) to \(E_{P_b}^s\) (counterclockwise). It follows from Lemma 6.6 that the tangent vectors of the two curves \(C_{\Phi_b}\) and \(C_{\Psi_b}\) at \(P_0\) are contained in the cone \(C(E_{P_b}^c, E_{P_b}^s)\). The same conclusion holds at the point \(Q_0\).

Let \(1.5 < b < b_{\text{max}}\), \(D(\Phi_b)\) be the set of points \((d_\ell, d_r) \in M^+\) such that \((d_\ell', d_r') = \Phi_b(d_\ell, d_r)\) is well-defined. It is possible that \(D(\Phi_b)\) has multiple connected components. Note that \(C_{\Phi_b} \subset D(\Phi_b)\) and \(C_{\Phi_b}\) is invariant under \(\Phi_b\). Let \(D_0(\Phi_b) \subset D(\Phi_b)\) be the connected component of \(D(\Phi_b)\) containing \(C_{\Phi_b}\). Then the curve \(C_{\Phi_b}\) divides \(D_0(\Phi_b)\) into two parts: denote the exterior part by \(D_{\text{ext}}(\Phi_b)\), and the interior part by \(D_{\text{int}}(\Phi_b)\).
Lemma 6.7. Let $1.5 < b \leq b_{\text{max}}$. Then

$$d'_\ell - d'_r > d_r - d_\ell$$

for any $(d_\ell, d_r) \in D_{\text{ext}}(\Phi_b)$, where $(d'_\ell, d'_r) = \Phi_b(d_\ell, d_r)$. The reversed inequality holds for points in $D_{\text{int}}(\Phi_b)$.

Proof. Let $1.5 < b \leq b_{\text{max}}$ be fixed. Note that the equality $d'_\ell - d'_r = d_r - d_\ell$ holds if and only if $(d_\ell, d_r) \in C_{\Phi_b}$. Then the inequality $d'_\ell - d'_r \neq d_r - d_\ell$ holds for any $(d_\ell, d_r) \in D_{\text{ext}}(\Phi_b)$, where $(d'_\ell, d'_r) = \Phi_b(d_\ell, d_r)$. Since $D_{\text{ext}}(\Phi_b)$ is connected on which $\Phi_b$ is smooth, the direction of the inequality (6.43) is fixed with respect to the coordinates $(d_\ell, d_r) \in D_{\text{ext}}(\Phi_b)$ and with respect to the parameter $b$. So we only need to check the direction of the inequality at one point $(d_\ell, d_r) \in D_{\text{ext}}(\Phi_b)$ for one $b \in (1.5, b_{\text{max}})$, say $b = 1.6$. See the first picture in Fig. 19 for $b = 1.6$. In the proof of Lemma 6.4, we have obtained the tangent matrix $D_{\text{E}_0} \Psi_b = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. It follows that $D_{\text{E}_0} \Phi_{1.6} = JD_{\text{E}_0} \Psi_{1.6} J = \begin{bmatrix} a_{22}(1.6) & a_{21}(1.6) \\ a_{12}(1.6) & a_{11}(1.6) \end{bmatrix} = \begin{bmatrix} 0.874667 & 0.503482 \\ 0.466667 & -0.874667 \end{bmatrix}$. Then the image of the tangent vector $v = (1, 1) \in T_{\text{E}_0} \mathcal{M}$ satisfies

$$D_{\text{E}_0} \Phi_b(v) = \begin{bmatrix} a_{22}(1.6) + a_{21}(1.6) \\ a_{12}(1.6) + a_{11}(1.6) \end{bmatrix} = \begin{bmatrix} 1.37815 \\ -0.408 \end{bmatrix}.$$  

Then $d'_\ell - d'_r > d_r - d_\ell$ for $(d_\ell, d_r) \in D_{\text{ext}}(\Phi_b)$. The case about points in $D_{\text{int}}(\Phi_b)$ can be proved in the same way. This completes the proof. \qed

There is a similar property as Lemma 6.7 with the curve $C_{\Psi_b}$ and $\Psi_b$ instead of $C_{\Phi_b}$ and $\Phi_b$ for $1.5 < b \leq b_{\text{max}}$. More precisely, let $D(\Psi_b) \subset \mathcal{M}^+$ be the part domain of $\Psi_b$ in the first quadrant $\mathcal{M}^+$. It is an open subset and contains the curve $C_{\Psi_b}$. Let $D_0(\Psi_b) \subset D(\Phi_b)$ be the connected component of $D(\Psi_b)$ containing $C_{\Psi_b}$. Then $C_{\Psi_b}$ divides $D_0(\Psi_b)$ into two parts: denote the exterior part by $D_{\text{ext}}(\Phi_b)$, and the interior part by $D_{\text{int}}(\Phi_b)$.

Lemma 6.8. Let $1.5 < b \leq b_{\text{max}}$. Then

$$d'_\ell - d'_r < d_r - d_\ell$$

for any $(d_\ell, d_r) \in D_{\text{ext}}(\Phi_b)$, where $(d'_\ell, d'_r) = \Psi_b(d_\ell, d_r)$. The reversed inequality holds for points in $D_{\text{int}}(\Phi_b)$.

Figure 20. Plot of slopes from top to bottom: $m_s(\eta)$ (red), $m_{2}(\eta)$ (cyan), $m_{2s}(\eta)$ (orange) and $m_u(\eta)$ (blue) as $0.52 < \eta < 1$. 
The proof is omitted, since it follows from Lemma 6.8 by applying the involution \( \mathcal{I} \).

**Lemma 6.9.** For \( b \in (1.5, b_{\text{max}}) \), there does not exist \((d_\ell, d_r) \in \mathcal{S}_{1,+}(b) \) such that

\[
\mathcal{L}_b \circ \mathcal{I}(d_\ell, d_r) = \mathcal{R}_b \circ \mathcal{L}_b(d_\ell, d_r).
\]

**Proof.** Suppose on the contrary that there is a solution \((d_\ell, d_r) \in \mathcal{S}_{1,+}(b) \) of (6.46). It follows that

\[
\begin{align*}
&d_\ell - d_r = b(1 - 2d_r^2); \\
&d_\ell = \mathcal{R}_b(d_r - b, d_r)_2 = d_r - b(1 - 2(d_r - b)^2) = 2b(d_r - b)^2 + d_r - 2b; \\
&(d_\ell - d_r)(1 - 2d_r^2) - 2d_\ell \sqrt{(b^2 - (d_\ell - d_r)^2)(1 - d_r^2)} + d_\ell = d_r - b.
\end{align*}
\]

Solving \( d_r \) from the first equation, we have \( d_{r,\pm} = \frac{b \pm \sqrt{3 - b^2}}{2} \). We plug it into the second equation and obtain

\[
d_{\ell,\pm} := d_{r,\pm} + b(1 - 2d_{r,\pm}^2) = \pm \frac{1}{2}(1 - 2b^2)\sqrt{3 - b^2}.
\]

Plugging \((d_{\ell,\pm}, d_{r,\pm})\) into the third equation, we get, respectively

\[
\begin{align*}
&\frac{1}{2} \left((1 - 2b^2)^3(3 - b^2)^{3/2} + 4(1 - 2b^2)\sqrt{3 - b^2} - 4\sqrt{3 - b^2} - (1 - 2b^2)^3(b^2 - 3)\sqrt{3 - b^2} + b \right) \\
&\quad + (2b^2 - 1)\sqrt{3 - b^2} \sqrt{b^2(4b^6 - 16b^4 + 13b^2 + 1)(4b^4 - 12b^2 - 4\sqrt{3 - b^2}b + 3)} = 0, \\
&\frac{1}{2} \left((1 - 2b^2)^3(3 - b^2)^{3/2} + 4(2b^2 - 1)\sqrt{3 - b^2} - 4\sqrt{3 - b^2} + (b^2 - 3)(1 - 2b^2)^3(\sqrt{3 - b^2} - b) \right) \\
&\quad + (1 - 2b^2)\sqrt{3 - b^2} \sqrt{b^2(4b^6 - 16b^4 + 13b^2 + 1)(4b^4 - 12b^2 + 4\sqrt{3 - b^2}b + 3)} = 0.
\end{align*}
\]

The solutions to the Equations (6.47) and (6.48) are

\[
\begin{align*}
b &= 0, -\sqrt{\frac{3}{2}} \pm \sqrt{3}, \frac{1}{2}(2 - \sqrt{2}), \frac{1}{2}(2 - \sqrt{2}); \\
b &= 0, \sqrt{\frac{3}{2}} \pm \sqrt{3}, \frac{1}{2}(-2 \pm \sqrt{2}), \frac{1}{2}(2 + \sqrt{2}),
\end{align*}
\]

respectively. Since none of the solutions is in \((1.5, 1 + 2^{-1/2})\), the lemma follows. \( \Box \)

7. Heteroclinic intersections and topological entropy

Let \( 1.5 < b \leq b_{\text{max}} \), \( F_b \) be the billiard map on the lemon table \( Q(b) \). As we have seen in Section 2.5, the periodic orbit \( \mathcal{O}_2(b) \) is elliptic and \( D_P F_b^2 \) is conjugate to the rotation matrix \( R_{\rho} \), where \( \theta = \arccos(2(b - 1)^2 - 1) \). It follows that the rotation number \( \rho(F_b^2, P) \) of \( F_b^2 \) at \( P \) satisfies \( \rho(F_{1.5}^2, P) = \frac{1}{3} \) and is decreasing with respect to \( b \) for \( b > 1.5 \). Let \( \rho > \frac{1}{3} \) be a Diophantine number that is sufficiently close to \( \frac{1}{3} \), \( \delta_{\rho} > 0 \) be a positive number given by Proposition 2.7. For any \( 1.5 < b < 1.5 + \delta_{\rho} \), let \( C_{\rho}(b) \) be the union of two \( F_b^2 \)-invariant circles of rotation number \( \rho \) surrounding the elliptic periodic points \( P, F_b(P) \in \mathcal{O}_2(b) \), and \( D_{\rho}(b) \subset M \) be the union of two \( F_b^2 \)-invariant disks enclosed by \( C_{\rho}(b) \). Since \( \rho > \frac{1}{3} \), \( D_{\rho}(b) \) contains both periodic orbits \( \mathcal{O}_2^p(b) \) and \( \mathcal{O}_2^h(b) \) constructed in Section 3. Let \( \delta_M > 0 \) be the number given at the end of Section 3.1. Then the elliptic periodic orbit \( \mathcal{O}_2^p(b) \) is nonlinearly stable for \( 1.5 < b < 1.5 + \delta_M \). In the phase space \( \mathcal{M} \), the elliptic periodic orbit \( \mathcal{O}_2^p(b) \) corresponds to the origin \((0,0)\), \( C_{\rho}(b) \) corresponds to a circle \( C_{\rho}(b) \) surrounding the origin \((0,0)\) that is invariant under both \( \mathcal{L}_b \) and \( \mathcal{R}_b \), and \( D_{\rho}(b) \) corresponds to a disk \( D_{\rho}(b) \) containing the origin.
(0, 0) that is invariant under both $L_b$ and $R_b$. In particular, $D_\rho(b)$ contains the periodic orbits of period 6: the orbit of $P_0$, the orbit of $Q_0$ and the orbit of $E_0$.

7.1. Some characterizations of the system for $b$ close to 1.5. Let $C_{\Phi_b}^{\text{sym}} \subset \mathcal{M}^+$ be the set of points $(d_\ell, d_r) \in \mathcal{M}^+$ such that
\[ d'_\ell - d'_r = d_\ell - d_r, \tag{7.1} \]
where $(d'_\ell, d'_r) = \Phi_b(d_\ell, d_r)$. Geometrically, $C_{\Phi_b}^{\text{sym}}$ consists of points in $\mathcal{M}^+$ whose reflection aims at the point $B_r = (1, 0) \in \Gamma_r$. It follows that $C_{\Phi_b}^{\text{sym}} = L_b\{(−x, (b−1)x) : 0 \leq x \leq 1\}$. Note that $C_{\Phi_b}^{\text{sym}}$ intersects the diagonal exactly once at $E_0$, since any such intersection point leads to a parallel orbit that is symmetric with respect to both axes and must be the periodic point $E_0$. Similarly we can define $C_{\Psi_b}^{\text{sym}}$. It is easy to see that $I(C_{\Phi_b}^{\text{sym}}) = C_{\Psi_b}^{\text{sym}}$. See Fig. 21.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure21.png}
\caption{The curves $C_{\Phi_b}^{\text{sym}}$ (red) and $C_{\Psi_b}^{\text{sym}}$ (red and dashed) and $C_{\Psi_b}^{\text{sym}}$ (blue and dashed), where $b = 1.6$. The cyan curve is an illustration of part of $C_{\rho}(b)$ in the first quadrant.}
\end{figure}

Lemma 7.1. (1) Both $C_{\Phi_b}^{\text{sym}} \cap C_{\Phi_b} = \{E_0(b)\}$ and $C_{\Psi_b}^{\text{sym}} \cap C_{\Psi_b} = \{E_0(b)\}$ hold for any $1.5 < b < b_{\text{crit}}$.

(2) There exists a positive number $\delta_C > 0$ such that $C_{\Phi_b}^{\text{sym}} \cap C_{\Psi_b} = \{E_0(b)\}$ and $C_{\Psi_b}^{\text{sym}} \cap C_{\Phi_b} = \{E_0(b)\}$ for any $1.5 < b \leq 1.5 + \delta_C$.

Proof. (1) It is clear that $E_0(b) \in C_{\Phi_b}^{\text{sym}} \cap C_{\Phi_b}$. It suffices to show that $(d_\ell, d_r) = E_0$ for any $(d_\ell, d_r) \in C_{\Phi_b}^{\text{sym}} \cap C_{\Phi_b}$. Note that such a point satisfies
\[ d'_\ell - d'_r = d_\ell - d_r = d_r - d_\ell, \tag{7.2} \]
where $(d'_\ell, d'_r) = \Phi_b(d_\ell, d_r)$. It follows that $d_\ell = d_r$, and hence $(d_\ell, d_r) \in C_{\Phi_b} \cap \Delta = \{E_0(b)\}$ for all $1.5 < b \leq b_{\text{crit}}$. The second intersection follows in the same way.

(2) Note that the tangent maps of $L_b$, $R_b$ and $\Phi_b$ at the origin for $b_0 = 1.5$:
\[ T_0L_{b_0} = \begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix}, \quad T_0R_{b_0} = \begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix}, \quad T_0\Phi_b = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \]

In this case, the linearized equation of (7.1) at $(0, 0)$ is
\[ d_\ell = d_r. \]
which gives the diagonal. Since for \( b \) close to 1.5 and in a small neighborhood of the origin, Eq. (7.1) is a \( C^\infty \)-perturbation of \( d_t - d_o = 0 \), \( C^{\text{sym}}_{\Phi_b} \) is cut out smoothly as a \( C^\infty \)-perturbation of the diagonal. It follows that for any \( \epsilon > 0 \), there exist a number \( \delta > 0 \) and a neighborhood \( U \subset M \) of the origin \((0,0)\) such that for every \( 1.5 < b < 1.5 + \delta \), \( C^{\text{sym}}_{\Phi_b} \cap U \) is a smooth (properly embedded) curve that is \( \epsilon \)-close to the diagonal under \( C^1 \) norm. Similarly, there exists a number \( \delta' > 0 \) such that for each \( 1.5 < b < 1.5 + \delta' \), the slope of the curve \( C_{\Psi_b} \) is strictly negative. It follows such two curves can intersect only once for \( 1.5 < b < 1.5 + \min\{\delta, \delta'\} \), which corresponds to the point \( E_b \). This completes the proof.

For convenience we introduce the constant
\[
\delta_0 = \min\{\delta, \delta_M, \delta_C\}. \tag{7.3}
\]

**Proposition 7.2.** Let \( b \in (1.5, 1.5 + \delta_0) \). Then there is no other periodic orbit of period 6 that is contained in the domain \( D^+_\rho(b) \) beside the hyperbolic and elliptic periodic orbit found in Section 3.

**Proof.** Let \( b \in (1.5, 1.5 + \delta_0) \). It follows from Eq. (4.4) that any periodic orbit of period 6 in the domain \( D^+_\rho(b) \) must intersect the first quadrant \( D^+_\rho(b) := D^+_\rho(b) \cap M^+ \). Moreover, if \( (d_\ell, d_r) \in D^+_\rho(b) \) is a periodic point of period 6, then \( \Psi_b \Phi_b(d_\ell, d_r) = (d_\ell, d_r) \) and \( \Phi_b(d_\ell, d_r) = \Psi_b(d_\ell, d_r) \in D^+_\rho(b) \). To prove Proposition 7.2, we will divide the domain \( D^+_\rho(b) \) into smaller components and consider them separately. More precisely, the union \( C_{\Phi_b} \cup C_{\Psi_b} \) divides \( D^+_\rho(b) \) into four connected components:

(i) \( D^+_\rho(b, 1) \) – the closed component containing the curve \( C_b(b) \cap M^+ \);
(ii) \( D^+_\rho(b, 2) \) – the closed (modulo the origin) component neighboring to the origin;
(iii) \( D^+_\rho(b, 3) \) – the upper-left open component of the domain enclosed by \( C_{\Phi_b} \cup C_{\Psi_b} \);
(iv) \( D^+_\rho(b, 4) \) – the lower-right open component of the domain enclosed by \( C_{\Phi_b} \cup C_{\Psi_b} \).

(1). Note that any point \( (d_\ell, d_r) \in D^+_\rho(b, 1) \) satisfies \((1, -1) \cdot \Phi_b(d_\ell, d_r) + d_\ell - d_r \geq 0 \) and \((1, -1) \cdot \Psi_b(d_\ell, d_r) + d_\ell - d_r \leq 0 \). Moreover, at most one of the two inequalities is an equality except at \( P_0, Q_0 \) or \( E_0 \). The reversed inequalities hold for points in \( D^+_\rho(b, 2) \). Let \( (d_\ell, d_r) \in D^+_\rho(b, 1) \) be a periodic point of period 6. Then \( \Phi_b(d_\ell, d_r) = \Psi_b(d_\ell, d_r) =: (d'_\ell, d'_r) \). It follows that \( d'_\ell + d'_r + d_\ell - d_r \geq 0 \) and \( d'_\ell + d'_r + d_\ell - d_r \leq 0 \) hold simultaneously. Hence \( d'_\ell + d'_r + d_\ell - d_r = 0 \), and
\[
(d_\ell, d_r) \in C_{\Phi_b} \cap C_{\Psi_b} = \{P_0, Q_0, E_0\}.
\]
The case when \( (d_\ell, d_r) \in D^+_\rho(b, 2) \) can be proved in the same way. Therefore, there is no other periodic point of period 6 in the two components \( D^+_\rho(b, 1) \cup D^+_\rho(b, 2) \) besides \( P_0, Q_0 \) and \( E_0 \).

(2). Consider the remaining open components \( D^+_\rho(b, 3) \) and \( D^+_\rho(b, 4) \). Note that

(a) a point \( (d_\ell, d_r) \in D^+_\rho(b, 3) \) satisfies \((1, -1) \cdot \Phi_b(d_\ell, d_r) + d_\ell - d_r < 0 \) and \((1, -1) \cdot \Psi_b(d_\ell, d_r) + d_\ell - d_r < 0 \);
(b) a point \( (d_\ell, d_r) \in D^+_\rho(b, 4) \) satisfies \((1, -1) \cdot \Phi_b(d_\ell, d_r) + d_\ell - d_r > 0 \) and \((1, -1) \cdot \Psi_b(d_\ell, d_r) + d_\ell - d_r > 0 \).

It follows that any periodic point \( (d_\ell, d_r) \) of period 6 in \( D^+_\rho(b, 3) \) satisfies \( \Phi_b(d_\ell, d_r) \in D^+_\rho(b, 3) \), since \( \Phi(d_\ell, d_r) = \Psi(d_\ell, d_r) \). Therefore, \( (d_\ell, d_r) \in D^+_\rho(b, 3) \cap \Phi_b D^+_\rho(b, 3) \). Similarly, any
periodic point \((d_\ell, d_r)\) of period 6 in \(\mathcal{D}_\rho^+(b, 4)\) satisfies \((d_\ell, d_r) \in \mathcal{D}_\rho^+(b, 4) \cap \Phi_b \mathcal{D}_\rho^+(b, 4)\). Next we will show that there is no periodic point of period 6 in the two open components \(\mathcal{D}_\rho^+(b, 3) \cup \mathcal{D}_\rho^+(b, 4)\).

The curve \(S_{\Phi_b}^{sym}\) divides the domain \(\mathcal{D}_\rho^+(b)\) into two components: one contains \(P_0\) and the other one contains \(Q_0\). Denote them by \(\mathcal{D}_\rho^+(b, P_0)\) and \(\mathcal{D}_\rho^+(b, Q_0)\), respectively. Let \((d'_\ell, d'_r) = \Phi_b(d_\ell, d_r)\). Since \(d'_\ell - d'_r - d_\ell + d_r \neq 0\) on \(\mathcal{D}_\rho^+(b, P_0)\) and it is positive at \(P_0\), it follows that \(d'_\ell - d'_r - d_\ell + d_r > 0\) on \(\mathcal{D}_\rho^+(b, P_0)\). Similarly, we have \(d'_\ell - d'_r - d_\ell + d_r < 0\) on \(\mathcal{D}_\rho^+(b, Q_0)\).

It follows from Lemma 7.1 that \(\mathcal{D}_\rho^+(b, 3) \subset \mathcal{D}_\rho^+(b, P_0)\) and \(\mathcal{D}_\rho^+(b, 4) \subset \mathcal{D}_\rho^+(b, Q_0)\). Any periodic point of period 6 in \(\mathcal{D}_\rho^+(b, 3)\), say \((d_\ell, d_r)\), satisfies that \((d'_\ell, d'_r) = \Phi_b(d_\ell, d_r) \in \mathcal{D}_\rho^+(b, P_0)\). Since \(\Phi_b\) is an involution, \(\Phi_b(d'_\ell, d'_r) = (d_\ell, d_r)\). It follows that \(d_\ell - d_r - d'_\ell + d'_r > 0\) since \((d'_\ell, d'_r) \in \mathcal{D}_\rho^+(b, P_0)\). This contradicts the hypothesis that \((d_\ell, d_r) \in \mathcal{D}_\rho^+(b, P_0)\). In the same way we have that there is no periodic point of period 6 in \(\mathcal{D}_\rho^+(b, 4)\). This completes the proof.

Let \(1.5 < b < 1.5 + \delta_0\). We have that the two hyperbolic periodic orbits of \(P_0\) and \(Q_0\) are contained in the invariant domain \(\mathcal{D}_\rho(b)\) bounded by the invariant circle \(\mathcal{C}_\rho(b)\) of rotation number \(\rho\). Note that there is no singular points in \(\mathcal{D}_\rho(b)\), and hence every point in it alternates bounces between the two arcs \(\Gamma_\ell\) and \(\Gamma_r\) of the boundary \(\partial \mathcal{Q}(b)\). It follows that the iterates of the billiard map \(F_b\) on the domain \(\mathcal{D}_\rho(b)\) agree with the iterates of \(\mathcal{R}_b\) and \(\mathcal{L}_b\) on the domain \(\mathcal{D}_\rho(b)\). Note that the stable and unstable manifolds of the hyperbolic periodic orbit of \(P_0\) are contained in \(\mathcal{D}_\rho(b)\). It follows that iterates of the billiard map \(F_b\) on the stable and unstable manifolds of the hyperbolic periodic orbits of period 6 agree with the iterates of \(\mathcal{R}_b\) and \(\mathcal{L}_b\).

**Lemma 7.3.** The union of the stable and unstable manifolds of the hyperbolic periodic orbit of period 6 of \(P_0\) is invariant under \(\mathcal{L}_b\) and \(\mathcal{R}_b\).

**Proof.** Let \(P \in W^s(P_0)\). That is \(F^{6n}(P) \to P_0\) as \(n \to \infty\). Then \(\mathcal{L}_b(P) \in \mathcal{L}_b W^s(P_0)\) satisfies that \(F^{-6n}(\mathcal{L}_b(P)) = \mathcal{L}_b(F^{6n}(P)) \to \mathcal{L}_b(P_0) = Q_1\) as \(n \to \infty\). It follows that \(\mathcal{L}_b W^s(P_0) = W^s(Q_1)\). Similarly, we have \(\mathcal{R}_b W^u(Q_1) = W^u(P_0)\) and \(\Phi_b(W^u(P_0)) = W^u(Q_0)\). The completes the proof. Alternatively, note that the involution \(I: (\varphi, \theta) \mapsto (\varphi, \pi - \theta)\) on the phase space \(M_b\) of the billiard map \(F_b\) amounts to reversing the direction of the trajectories. In particular, it takes a stable manifold of a hyperbolic periodic point to the unstable manifold of the reversed hyperbolic periodic point and vice versa. It follows that the union of the stable and unstable manifolds of the the hyperbolic periodic orbit of \(P_0\) is invariant under the involution \(I\). Note that \(\mathcal{L}_b\) and \(\mathcal{R}_b\) are the realizations of the involution \(I\) on the left and right arc, respectively. Hence the lemma follows.

Let \(Z \subset (1.5, 1.5 + \delta_0)\) be the set of parameters \(b\) such that there is no homoclinic intersection for the hyperbolic periodic point \(P_0(b)\). It is easy to see that the point \(P_0(b)\) in this definition can be replaced by any of the remaining 5 hyperbolic periodic points given in Remark 6.3.

**Proposition 7.4.** The topological entropy \(h_{top}(F_b)\) of the billiard map \(F_b\) is positive for every \(b \in (1.5, 1.5 + \delta_0)\) \(\backslash Z\).
Proof. Let $b \in (1.5, 1.5 + \delta_0) \setminus Z$ be given. That is, there is a homoclinic intersection for the hyperbolic $\Theta_b$-fixed point $P_0(b)$. We claim that it cannot be a homoclinic loop. Suppose on the contrary that it is a homoclinic loop, say $\eta(P_0)$. Let $D(P_0) \subset M$ be the domain bounded by the loop $\eta(P_0)$. Then by Proposition 2.2, there must be a $\Theta_b$-fixed point contained in the interior of $D(P_0)$, say $T$. Note that the only points fixed by $\Theta_b$ are those given in Remark 6.3 and the origin $(0, 0)$. If $T = (0, 0)$ or $S_0$, then by symmetry, $\mathcal{I}(\eta(P_0))$ is a homoclinic loop for $Q_0$ and also encloses the same point $T$, a contradiction. Similarly, there is a contradiction for any other choice of $T$.

Since both stable and unstable manifolds of $P_0$ are analytic, they can only intersect at discrete points. Moreover, since the system is conservative, there must be an intersection that is a homoclinic intersection of $P_0$ that is topological crossing. See [5, Theorem 3.1]. Then it follows from Proposition 2.1 that $h_{\text{top}}(F_b) > 0$. □

7.2. Local phase portrait. Let $\Delta = \{d_x = d_y\}$ be the diagonal of the phase space, $P_0$ and $Q_0$ be the two hyperbolic periodic points in the first quadrant, $E_0$ be the elliptic periodic point in the first quadrant, see Fig. 17. Let $\Gamma_{b,P_0,+}^s$ and $\Gamma_{b,P_0,+}^u$ be the two branches of the stable and unstable manifolds of $P_0$ that contains the local manifolds that enter the first quadrant, respectively. Similarly, we define $\Gamma_{b,Q_0,+}^s$ and $\Gamma_{b,Q_0,+}^u$. See Fig. 18. Recall that $Z \subset (1.5, 1.5 + \delta_0)$ is the set of parameters $b$ such that there is no homoclinic intersection for the hyperbolic periodic point $P_0(b)$.

Proposition 7.5. Let $b \in Z$. Then $\Gamma_{b,P_0,+}^s = \Gamma_{b,Q_0,+}^u$ is a heteroclinic between $P_0$ and $Q_0$. Moreover, this connection intersects the diagonal exactly once, and the intersection is perpendicular. The same holds for $\Gamma_{b,P_0,+}^u = \Gamma_{b,Q_0,+}^s$. Furthermore, both connections are contained in the first quadrant $M^+$ and their union encloses the elliptic island around the point $E_0$.

Proof. Let $b \in (1.5, 1.5 + \delta_0)$ be fixed. We divide the proof into several steps:

Step 1. We claim that $\Gamma_{b,P_0,+}^s$ can intersect the diagonal $\Delta$ at most once. Moreover, such an intersection point, if exists, implies that $\Gamma_{b,P_0,+}^s = \Gamma_{b,Q_0,+}^u$ is a heteroclinic connection. More precisely, $\Gamma_{b,P_0,+}^s = \Gamma_{b,Q_0,+}^u. \hfill \square$

Proof of Claim. We only need to prove the claim for $\Gamma_{b,P_0,+}^s$. Suppose $P \in \Gamma_{b,P_0,+}^s \cap \Delta$ is the first intersection. Then by symmetry, $\Gamma_{b,Q_0,+}^u$ will intersect $\Delta$ at the same point $P$ for the first time. It follows that $P$ is a heteroclinic point between $P_0$ and $Q_0$. Suppose on the contrary that $\Gamma_{b,P_0,+}^s \neq \Gamma_{b,Q_0,+}^u$. Then there exists a point $A \in \Gamma_{b,P_0,+}^s \cap \{(d_x, d_y) \in M^+ : d_x < d_y\} \setminus \Gamma_{b,Q_0,+}^u$. Then by symmetry, the point $A' = \mathcal{I}(A) \in \Gamma_{b,Q_0,+}^u \setminus \Gamma_{b,P_0,+}^s$. It follows that $\Gamma_{b,Q_0,+}^u$ and $\Gamma_{b,P_0,+}^s$ have a crossing intersection between $A$ and $A'$. Such a crossing leads to a crossing heteroclinic intersection between the corresponding branches of $P_1$ and $Q_1$ (applying $L_b$) and a crossing heteroclinic intersection between the corresponding branches of $P_2$ and $Q_2$ (applying $R_b$). Then by center-symmetry, we get crossing heteroclinic intersections of the corresponding branches of $(P_1, Q_2)$, $(P_0, Q_1)$ and $(P_2, Q_0)$. Together they form a cycle of crossing heteroclinic intersections that lie on the same side. Then the point $P_0$ admits a crossing homoclinic intersection by the lambda-lemma. This contradicts the choice $b \in Z$. □
Step 2. Suppose that $\Gamma^s_{b,P_0} \cap \{d_{\ell} = 0\} = \emptyset$. Let $T = (0, d^+_r) \neq P_0$ be the first intersection of $\Gamma^s_{b,P_0}$ with $\{d_{\ell} = 0\}$. We claim that $0 < d^+_r < (1 - \frac{1}{4b^2 - 8})^{1/2}$.

Proof of the Claim. It follows from Step (1) that $[P_0, T]_b^s \cap \Delta = \emptyset$. In particular, $[P_0, T]_b^s$ is contained in the wedge $\{0 < d_{\ell} < d_r < 1\}$. So there are two possibilities for $d^+_r$: either $d^+_r > (1 - \frac{1}{4b^2 - 8})^{1/2}$ or $0 < d^+_r < (1 - \frac{1}{4b^2 - 8})^{1/2}$. It remains to show that it is impossible to have $d^+_r > (1 - \frac{1}{4b^2 - 8})^{1/2}$.

Suppose on the contrary that $d^+_r > (1 - \frac{1}{4b^2 - 8})^{1/2}$. Consider the following two parametric curves:

$$\eta^+_1 := \mathcal{L}_b(\{0\} \times [0, 1]) = \{(2d_r(1 - d_{\ell}^2 - \sqrt{(1 - d_{\ell}^2)(b^2 - d_{\ell}^2)}), d_r) : d_r \in [0, 1]\};$$

$$\eta^-_2 := \mathcal{R}_b([-1, 0] \times \{0\}) = \{(d_r, 2d_r(1 - d_{\ell}^2 - \sqrt{(1 - d_{\ell}^2)(b^2 - d_{\ell}^2)})) : d_r \in [-1, 0]\}.$$

See Fig. 22. Note that $Q_1 = \mathcal{L}_b(P_0) \in \eta^+_1$ and $Q_1 = \mathcal{R}_b(P_1) \in \eta^-_2$. Moreover, it is easy to check that $\eta^+_1$ is below the line $d_{\ell} + d_r = 0$ in the region $0 < d_r < (1 - \frac{1}{4b^2 - 8})^{1/2}$ and above the line $d_{\ell} + d_r = 0$ in the region $(1 - \frac{1}{4b^2 - 8})^{1/2} < d_r < 1$. The pushforward stable arc $\mathcal{L}_b([P_0, T]_{b_0}^s)$ is an unstable arc of the hyperbolic periodic point $Q_1$, while the reflection of $\mathcal{L}_b([P_0, T]_{b_0}^s)$ along the line about $d_{\ell} + d_r = 0$ is a stable arc of $Q_1$. Since $\mathcal{L}_b$ is orientation-reversing inside the domain $\mathcal{R}_b$, the stable arc $\mathcal{L}_b(P_0, T]_{b_0}^s)$ connects $Q_1$ and $\mathcal{L}_b(T)$ through the left side of $\eta^+_1$ and has to intersect the line $d_r + d_{\ell} = 0$ via a point $T' \neq Q_1$. Then the unstable arc, being the reflection of the stable arc with respect to $d_r + d_{\ell} = 0$, intersect the line $d_r + d_{\ell} = 0$ at the same point $T'$. It turns out that $T'$ is a crossing homoclinic intersection for $Q_1$. It follows that $\mathcal{L}_b(T')$ is a crossing homoclinic intersection for $P_0$, contradicting $b \in Z$. This completes the proof of the claim.

![Figure 22](image-url)

Step 3. Suppose $\Gamma^u_{b_0,P_0} \cap \{d_{\ell} = 0\} = \emptyset$. Let $T = (0, d^+_r) \neq P_0$ be the first intersection of $\Gamma^u_{b_0,P_0}$ with $\{d_{\ell} = 0\}$. We claim that $d^+_r > (1 - \frac{1}{4b^2 - 8})^{1/2}$.

Proof of the Claim. Using the same argument as in Step 2, we have that $[P_0, T]_{b_0}^u$ is contained in the wedge $\{0 < d_{\ell} < d_r < 1\}$. So there are two possibilities for $d^+_r$: either $d^+_r > (1 - \frac{1}{4b^2 - 8})^{1/2}$.
or \(0 < d^*_r < (1 - \frac{1}{4b^2-8})^{1/2}\). It remains to show that it is impossible to have \(0 < d^*_r < (1 - \frac{1}{4b^2-8})^{1/2}\).

Suppose on the contrary that \(0 < d^*_r < (1 - \frac{1}{4b^2-8})^{1/2}\). Let \(\eta^+_s = \{ R_b L_b(a_z,0) : a_z \in [0,1]\}\), and \(\eta^-_s = \{ R_b L_b(a_z,0) : a_z \in [-1,0]\}\), which are symmetric with respect to the line \(d_\ell + d_r = 0\). Moreover, \(P_2 = R_b L_b(Q_0) \cap \eta^-_s\). It is easy to check that \(\eta^+_s\) is above the line \(d_\ell + d_r = 0\) in the region \(-(1 - \frac{1}{4b^2-8})^{1/2} < d_r < 0\) and below the line \(d_\ell + d_r = 0\) in the region \(-1 < d_r < -(1 - \frac{1}{4b^2-8})^{1/2}\). The pushforward arc \(L_b R_b([P_0,T]_{b_0}^u)\) is an unstable arc of the hyperbolic periodic point \(P_2 = L_b R_b(P_0) \in \eta^+_s\), while its reflection is a stable arc of \(P_2\). Moreover, the unstable arc \(L_b R_b([P_0,T]_{b_0}^u)\) connects \(P_2\) to \(L_b R_b(T)\) through the left side of \(\eta^+_s\) and has to intersect the line \(d_\ell + d_r = 0\) at a point \(T' \not\in P_2\). Then the stable arc intersects the line \(d_\ell + d_r = 0\) at the same point \(T'\), which makes \(T'\) a crossing homoclinic intersection of \(P_2\). By center symmetry, \(-T'\) is a crossing homoclinic intersection of \(Q_1\), and hence \(L_b(-T')\) is a crossing homoclinic intersection of \(P_0\), contradicting \(b \in Z\). This completes the proof of the claim.

**Step 4.** The conclusion \(0 < d^*_r < (1 - \frac{1}{4b^2-8})^{1/2}\) in Step 2 is also impossible, since it would force an intersection of the unstable branch \(\Gamma^s_{b_0,P_0,+}\) and the segment \(\{d_\ell = 0\}\) at some point with \(0 < d^*_r < (1 - \frac{1}{4b^2-8})^{1/2}\), which has been excluded in Step 3. If follows that \(\Gamma^s_{b_0,P_0,+}\) and \(\{d_\ell = 0\}\) do not intersect at all.

**Step 5.** Similarly, the conclusion \(d^*_r > (1 - \frac{1}{4b^2-8})^{1/2}\) in Step 3 is also impossible, since it would force an intersection of the stable branch \(\Gamma^s_{b_0,P_0,+}\) and the segment \(\{d_\ell = 0\}\) at some point with \(d^*_r > (1 - \frac{1}{4b^2-8})^{1/2}\), which has been excluded in Step 2. If follows that \(\Gamma^s_{b_0,P_0,+}\) and \(\{d_\ell = 0\}\) do not intersect at all.

**Step 6.** We claim that \(\Gamma^s_{b,P_0,+} \cap \Delta \neq \emptyset\) and \(\Gamma^u_{b,P_0,+} \cap \Delta \neq \emptyset\).

**Proof of the claim.** Suppose on the contrary that the claim is false. Without loss of generality we assume \(\Gamma^s_{b,P_0,+} \cap \Delta = \emptyset\). Then the whole branch \(\Gamma^s_{b,P_0,+}\) is contained in the open wedge \(\{(d_\ell,d_r) : d_r > d_\ell > 0\}\). In the closed wedge \(\{(d_\ell,d_r) : d_r \geq d_\ell \geq 0\}\), there are three points that are fixed by \(\Theta_b\): the only hyperbolic periodic point \(P_0\), two elliptic periodic points: the origin \((0,0)\) and the point \(E_0\), both of which are nonlinearly stable. It follows from Proposition 2.5 that

(i) either \(\Gamma^s_{b,P_0,+}\) is a saddle connection: in this case, it has to be a homoclinic loop for \(P_0\). Let \(D\) be the simply connected domain enclosed by the homoclinic loop \(\Gamma^s_{b,P_0,+}\). Note that \(D\) is \(\Theta_b\)-invariant and is nonwandering. It follows from Proposition 2.2 that there is an additional \(\Theta_b\)-fixed point in \(D\), contradicting Proposition 7.2.

(ii) or \(\Gamma^u_{b,P_0,+}\) is current. In this case, the only hyperbolic branch that is contained in the same wedge is \(\Gamma^u_{b,P_0,+}\). It follows that both branches are recurrent, and accumulate each other through the quadrant between them. Then it follows from Proposition 2.6 that there is a crossing homoclinic intersection for \(P_0\), contradicting \(b \in Z\).

We have derived a contradiction for each possible case under the assumption that the claim is false. This completes the proof of the claim. \(\square\)
Collecting terms, we have that $\Gamma_{b,P_0,+}^s$ intersects the diagonal $\Delta$ at some point $P$, which is also the only intersection point of $\Gamma_{b,Q_0,+}^u$ with the diagonal $\Delta$ due to the symmetry of the system. It follows from Step 1 that $\Gamma_{b,P_0,+}^s = \Gamma_{b,Q_0,+}^u$. That is, it is a heteroclinic connection and is contained in the first quadrant. For convenience, we may denote this connection by either $\Gamma_{b,P_0,Q_0}$ or $\Gamma_{b,Q_0,P_0}^u$. Applying symmetry again, its intersection with the diagonal is perpendicular.

Using a similar argument, we have that $\Gamma_{b,P_0,+}^s = \Gamma_{b,Q_0,+}^u$ is another heteroclinic connection of $P_0$ and $Q_0$ and is contained in the first quadrant. Then the union $\Gamma_{b,P_0,Q_0} := \Gamma_{b,P_0,Q_0}^u \cup \Gamma_{b,Q_0,P_0}$ is a simple closed $\Theta_b$-invariant curve which encloses a simply connected $\Theta_b$-invariant domain $D$ in the first quadrant. It follows from Proposition 2.2 that there is a $\Theta_b$-fixed point $p \in D$. Since there is exactly one such point in the interior of first quadrant— the elliptic periodic point $E_0$, we conclude that the $\Theta_b$-fixed point contained in $D$ is exactly the elliptic periodic point $E_0$. This completes the proof.

**Corollary 7.6.** For every $b \in Z$, the stable and unstable branches of all six hyperbolic periodic points $P_j$ and $Q_j$, $j = 0, 1, 2$, form two invariant cycles of heteroclinic connections with its neighboring points, and each of the six elliptic islands around the elliptic periodic points $S_j$ are contained in a pair of heteroclinic connections.

**Proof.** Let $\Gamma_{b,P_0,Q_0}$ and $\Gamma_{b,P_0,Q_0}^u$ be the pair of heteroclinic connections between $P_0$ and $Q_0$ given by Proposition 7.5. Then we get a pair of heteroclinic connections between $P_1$ and $Q_1$ (applying $L_b$) and a pair of heteroclinic connections between $P_2$ and $Q_2$ (applying $R_b$). By center symmetry, we get a pair of heteroclinic connections between $P_j$ and $Q_{j+1}$, $j = 0, 1, 2$, where the indices are modulo 3. The statements about the elliptic islands can be proved in the same way. This completes the proof.

7.3. **Hypothesis for Contradiction.** Recall that $Z \subset (1.5, 1.5 + \delta_0)$ is the set of parameters $b$ such that $P_0(b)$ has no homoclinic point. Then for each $b \in Z$, $\Gamma_{b,P_0,+}^s = \Gamma_{b,Q_0,+}^u$ is a heteroclinic connection between $P_0$ and $Q_0$ that is contained in the first quadrant, and intersect the diagonal exactly once, say $T_b = (d_b, d_b)$. Let $U \subset (1.5, 1.5 + \delta_0)$ be the set of parameters such that both $\Gamma_{b,P_0,+}^s$ and $\Gamma_{b,Q_0,+}^u$ cross the diagonal. Clearly $U$ is open. It follows from Proposition 7.5 that that $U \supset Z$. Let $U' \subset U$ be the union of the connection components of $U$ that intersect $Z$. For each $b \in U'$, let $T_{b,P_0} = (d_{b,P_0}, d_{b,P_0})$ be the first intersection of $\Gamma_{b,P_0,+}^s$ with the diagonal and $T_{b,Q_0} = (d_{b,Q_0}, d_{b,Q_0})$ for $\Gamma_{b,Q_0,+}^u$. Let $\eta_{b,P_0} : (d_{b,P_0} - \epsilon, d_{b,P_0} + \epsilon) \to \mathbb{R}$ be the function with $\eta(d_{b,P_0}) = d_{b,P_0}$ and whose graph is a short segment of the stable manifold $\Gamma_{b,P_0,+}^s$ around $T_{b,P_0}$. Similarly, define $\eta_{b,Q_0} : (d_{b,Q_0} - \epsilon, d_{b,Q_0} + \epsilon) \to \mathbb{R}$ for $\Gamma_{b,Q_0,+}^u$. Since both invariant manifolds $\Gamma_{b,P_0,+}^s$ and $\Gamma_{b,Q_0,+}^u$ depend analytically on $b$, the difference $\eta_{b,P_0} - \eta_{b,Q_0}$ also depends analytically for $b \in U'$.

**(HC) Hypothesis for contradiction.** The set $Z$ has a limit point in $(1.5, 1.5 + \delta_0)$.

Note that $T_{b,P_0} = T_{b,Q_0} = T_b$ and $\eta_{b,P_0} - \eta_{b,Q_0} = 0$ for every $b \in Z$. By analytical dependence on the parameter $b$, $\eta_{b,P_0} - \eta_{b,Q_0}$ must be identically zero over $b \in U'$. The crossing condition between the branches of the invariant manifolds of $P_0$ and $Q_0$ and the diagonal is an open condition while the equation $\eta_{b,P_0} - \eta_{b,Q_0} = 0$ defines a closed condition. It follows that $U' = U = (1.5, 1.5 + \delta_0)$, and hence $\Gamma_{b,P_0,+}^s = \Gamma_{b,Q_0,+}^u$ for every $b \in (1.5, 1.5 + \delta_0)$. Applying the same argument to the branches $\Gamma_{b,P_0,+}^u$ and $\Gamma_{b,Q_0,+}^s$ and then the argument in
Corollary 7.6, we have that there are two cycles of six heteroclinic connections between the six hyperbolic periodic points for \( b \in (1.5, 1.5 + \delta_0) \): one exterior cycle (denote by \( C_b^{\text{ext}} \)) containing \( \Gamma_{b,P_{b},Q_0}^s \) and one interior cycle (denote by \( C_b^{\text{int}} \)) containing \( \Gamma_{b,P_{b},Q_0}^u \). See Fig. 18 for an illustration with the phase portrait for \( b = 1.56 \). See [9] for methods of plotting stable and unstable manifolds of hyperbolic manifolds using normal forms.

Now we study what happens for \( b \in [1.5 + \delta_0, b_{\text{max}}) \), where \( b_{\text{max}} = 1 + 2^{-1/2} \). By the analytic dependence of the invariant manifolds on the parameter \( b \), the two cycles \( C_b^{\text{ext}} \) and \( C_b^{\text{int}} \) will persist as we increase \( b \) from \( b = 1.5 + \delta_0 \) till it is broken. If no breakdown happens for every \( b \in [1.5 + \delta_0, b_{\text{max}}) \), then the connection cycles will exist for all such tables. We will deal with this case later. For now we assume that a breakdown of some of the connections does happen. Let \( b_0 \in [1.5 + \delta_0, b_{\text{max}}) \) be the first parameter for which there is a breakdown for some heteroclinic connection of the two cycles (maybe both). Because of the analytic dependence, a breakdown of a connection on the interior cycle can only happen when the connection oscillates indefinitely\(^1\), while a breakdown of a connection on the exterior cycle can happen either when the connection oscillates indefinitely or when a connection falls onto one singularity curve of the domain. The indefinite oscillation breakdown for both interior and exterior cycles are similar. We will focus on the case that the cycle \( C_b^{\text{ext}} \) breaks down at \( b_0 \), and show necessary changes to the case when only the cycle \( C_b^{\text{int}} \) breaks down at \( b_0 \). We first show that all of conclusions in Proposition 7.5 continue to hold up to the breakpoint.

**Proposition 7.7.** Under the hypothesis (HC), both connections \( \Gamma_{b,P_{b},Q_0}^s \) and \( \Gamma_{b,P_{b},Q_0}^u \) are contained in the first quadrant \( \mathcal{M}^+ \), intersect the diagonal exactly once and the intersection is perpendicular, and their union encloses the point \( E_0 \) for every \( 1.5 < b < b_0 \).

*Proof.* We just need to examine the steps in the proof of Proposition 7.5. All of the conclusions in Step 1 and Step 6 hold automatically for the analytic continuation of the heteroclinic connections for \( 1.5 < b < b_0 \). The arguments in Step 2 and Step 3 work the same. The conclusions in Step 4 and Step 5 follow directly from Step 2 and Step 3. Since both the point \( E_0(b) \) and the curve \( \Gamma_{b,P_{b},Q_0}^s \cup \Gamma_{b,P_{b},Q_0}^u \) depend analytically on \( b \), and the point \( E_0(b) \) is enclosed by the curve for \( 1.5 < b < 1.5 + \delta_0 \), it continues to hold for \( 1.5 + \delta_0 \leq b < b_0 \). Therefore, all conclusions in Proposition 7.5 continue to hold for \( 1.5 < b < b_0 \). \( \square \)

Note that a breakdown between any connection of the cycle \( C_b^{\text{ext}} \) leads to the total breakdown of all six connections simultaneously. For example, if the heteroclinic connection \( \Gamma_{b,P_{1},Q_1}^s \) intersects the singular curve \( S_{2,-} \) at a point \( T \), then \( \mathcal{R}_b(T) \) is a point of intersection of the same connection \( \Gamma_{b,Q_{1'},P_1}^u = \mathcal{L}_b \Gamma_{b,P_{1},Q_1}^s \) with \( L_b \), and \( L_b \mathcal{R}_b(T) \) is a point of intersection of the connection \( \Gamma_{b,P_{b},Q_0}^s = L_b \Gamma_{b,Q_{1'},P_1}^u \) with \( S_{1,+} \). The same argument works for other connections and other intersections. Therefore, \( b_0 \) is the first parameter that either the heteroclinic connection \( \Gamma_{b,P_{b},Q_0}^s \) or the connection \( \Gamma_{b,P_{b},Q_0}^u \) breaks down (maybe both). In the following we will discuss possible scenarios when the breakdown of the heteroclinic connection \( \Gamma_{b,P_{b},Q_0}^s \) or \( \Gamma_{b,P_{b},Q_0}^u \) happens. We will exclude these possibilities and show that the two cycles of heteroclinic connections continues to exist for all \( b < b_{\text{max}} = 1 + 2^{-1/2} \). Then we will show that the heteroclinic connection \( \Gamma_{b,P_{b},Q_0}^s \) must have been broken for any \( b \) sufficiently close to \( b_{\text{max}} \) due to geometric reasons. This contradiction shows that the hypothesis (HC) is wrong.

\(^1\)Indefinite oscillations are possible since there might be new \( \Theta_b \)-fixed points that are either degenerate or non-stable elliptic, in which case the prime-end analysis is no longer applicable.
that $Z$ has a limit point in $(1.5, 1.5 + \delta_0)$. Combining with Proposition 7.4, we complete the proof of Theorem 1.2.

In the case that $\Gamma^s_{b,P_0,Q_0}$ breaks at $b = b_0$, we will denote $\Gamma^s_{b_0,P_0,Q_0} = \lim_{b \to b_0} \Gamma^s_{b,P_0,Q_0}$. Note that $\Gamma^s_{b_0,P_0,Q_0}$ is a continuum containing both points $P_0$ and $Q_0$, is contained in the first quadrant, is symmetric with respect to the diagonal, and is invariant under both $\Phi_b$ and $\Psi_b$. For convenience, we will call $\Gamma^s_{b_0,P_0,Q_0}$ a continuum connection. The continuum connection for $\Gamma^u_{b_0,P_0,Q_0}$ (if it breaks down) can be defined in the same way and it has the same property.

**Definition 7.8.** Let $b \in (1.5, b_0)$, $\rho_b : t \in [-1, 1] \mapsto \Gamma^s_{b,P_0,Q_0}$ be a parametrization for the connection $\Gamma^s_{b,P_0,Q_0}$ with $\rho_b(-1) = Q_0$, $\rho_b(0) = P_0$ and $\rho_b(t) = I(\rho_b(t))$.

Note that $\rho_b(0) = T_b$ is the unique intersection of $\Gamma^s_{b,P_0,Q_0}$ with the diagonal. Similar parametrization exists for the connection $\Gamma^u_{b,P_0,Q_0}$ with $\rho_b(-1) = P_0$, $\rho_b(t) = I(\rho_b(t))$.

**Lemma 7.9.** Let $1.5 < b < b_0$. Then $\rho_b^{-1}(\Phi_b(\rho_b(t))) < -t$ and $\rho_b^{-1}(\Psi_b(\rho_b(t))) > t$ for any $t \in [0, 1]$. Similarly, $\rho_b^{-1}(\Phi_b(\rho_b(t))) < -t$ and $\rho_b^{-1}(\Psi_b(\rho_b(t))) > t$ for any $t \in [0, 1]$.

**Proof.** Let $1.5 < b < b_0$ be fixed. We only need to prove the statement about $\rho_b$, since the same argument works for $\rho_b$. Moreover, we only need to prove the inequality about $\Phi_b$ since $\Gamma^s_{b,P_0,Q_0}$ is invariant under $I$ and the parametrization is symmetric. Let $A_b = \{t \in [0, 1] : \rho_b^{-1}(\Phi_b(\rho_b(t))) < -t\}$, which is an open subset of $[0, 1]$ for each $1.5 < b < b_0$. It suffices to show that $A_b$ is nonempty and closed. Then we have $A_b = \{0, 1\}$.

**Step 1.** $0 \in A_b$: We start with the case $b \in (1.5, 1.5 + \delta_0)$. It follows from Lemma 7.1 that a point $(d, d) \in \mathcal{M}^+$ satisfies $\Phi_b(d, d) \in \{d_\ell = d_r\}$ if and only if $(d, d) = E_0$. Applying Lemma 6.7, we have

$$
\Phi_b(d, d) \begin{cases} 
\in \{d_\ell < d_r\}, & \text{for } d < \sqrt{1 - \frac{1}{4(b-1)^2}}, \\
= (d, d), & \text{for } d = \sqrt{1 - \frac{1}{4(b-1)^2}}, \\
\in \{d_\ell > d_r\}, & \text{for } d > \sqrt{1 - \frac{1}{4(b-1)^2}}.
\end{cases}
$$

It is proved in Proposition 7.5 that $\Gamma^s_{b,P_0,Q_0}$ goes over the elliptic island through the northeastern corner. This implies that $\rho_b^{-1}(\Phi_b(\rho_b(0))) < 0$, or equally, $0 \in A_b$ for each $b \in (1.5, 1.5 + \delta_0)$.

Now assume that there exists $\tilde{b} \in [1.5 + \delta_0, b_0)$ such that $\rho_b^{-1}(\Phi_b(\rho_b(0))) = 0$, then $\rho_b(0) = T_{\tilde{b}}$ satisfies $\Phi_b(T_{\tilde{b}}) = T_{\tilde{b}}$. In particular, $T_{\tilde{b}}$ satisfies (6.20) and hence $T_{\tilde{b}} \in C_{\Phi_{\tilde{b}}} \cap \{d_\ell = d_r\}$, which is $\{S_b\}$ for $b \leq b_{\text{crit}}$ and $\{E_0(\tilde{b}), E_1(\tilde{b}), E_2(\tilde{b})\}$ for $b > b_{\text{crit}}$. In either case, only $E_0(\tilde{b})$ is fixed by $\Phi_{\tilde{b}}$. It follows that $T_{\tilde{b}} = E_0(\tilde{b})$. However, there cannot be a fixed/periodic point on any stable/unstable manifold, a contradiction. It follows that $0 \in A_b$ holds for each $b \in (1.5, \min\{b_{\text{crit}}, b_0\})$.

**Step 2.** $A_b$ is a closed subset of $[0, 1)$: Let $t_n \in A_b \to t_* \in (0, 1)$. Denote $(d_\ell^*, d_r^*) = \rho_b(t_*)$. Then $\rho_b^{-1}(\Phi_b(\rho_b(t_*))) \leq -t_*$. If $\rho_b^{-1}(\Phi_b(\rho_b(t_*))) = -t_*$, then $\Phi_b(\rho_b(t_*)) = \Phi_b(-t_*)$ is symmetric with respect to the diagonal. That is, $\Phi_b(d_\ell^*, d_r^*) = (d_r^*, d_\ell^*)$. It follows that $\rho_b(t_*) \in C_{\Phi_b}$. Combining with Lemma 6.7, we have that either $\rho_b(t_*) = P_0$ or $\rho_b(t_*) = E_0$, contradicting the hypothesis that $t_* \in (0, 1)$.

Therefore, $A_b$ is nonempty and closed, and hence $A_b = [0, 1)$. This completes the proof. □
In the following we will consider the union of the two curves $C_{Ψ_b}$ and $C_{Ψ_δ}$. We need to chop off some parts of these curves since there are multi-intersections of both curves with the diagonal for $b_{crit} < b \leq b_{max}$. More precisely, let $γ_1(b) \subset C_{Ψ_b}$ be the part from $P_0$ to $E_1(b) = (sin β, sin β)$, $γ_2(b) \subset C_{Ψ_δ}$ the part from $Q_0$ to $E_2(b) = (sin α, sin α)$, where $(α, β) \in D' \cap J$ is the unique pair with $b = b(α, β)$. Note that this construction also works for $1.5 < b \leq b_{crit}$, just with both $E_1(b)$ and $E_2(b)$ collapsing onto $E_0(b)$. Define $γ_3(b) := I(γ_1(b)) \subset C_{Ψ_b}$ and $γ_4(b) := I(γ_2(b)) \subset C_{Ψ_δ}$. It follows from Lemma 6.5 that $Φ_b(γ_1(b)) = γ_2(b)$ and $Ψ_b(γ_3(b)) = γ_4(b)$. Next we set

$$C_{prl,b} = \bigcup_{1 \leq j \leq 4} γ_j(b).$$  

(7.4)

See Fig. 19. Note that $C_{prl,b} = C_{Ψ_b} \cup C_{Ψ_δ}$ for $1.5 < b \leq b_{crit}$. It follows from the definition that $C_{prl,b}$ is symmetric with respect to the diagonal.

It follows from Lemma 5.6 that the curve $γ_1(b)$ is transverse to the lines $L_c := \{d_x - d_η = c\}$, $c \in \mathbb{R}$. So $γ_1(b)$ intersects each line $L_c$ exactly once and is a simple curve. Therefore, $C_{prl,b}$ is

1. a closed curve with one simple self-intersection at $E_0$ for $1.5 < b \leq b_{crit}$;
2. a simple closed curve enclosing a domain containing $E_0$ for $b_{crit} < b \leq b_{max}$.

**Lemma 7.10.** For each $1.5 < b \leq b_0$, both connections $Γ^s_{b,P_b,Q_0}$ and $Γ^u_{b,P_b,Q_0}$ do not intersect the curve $C_{prl,b}$.

**Proof.** We only need to give the proof for the connection $Γ^s_{b,P_b,Q_0}$, since the same argument works for the connection $Γ^u_{b,P_b,Q_0}$. See Fig. 19 for plots of the curve $C_{prl,b}$.

**Step 1.** Suppose $Γ^s_{b,P_b,Q_0} \cap C_{prl,b} \neq \emptyset$ for some $1.5 < b < b_0$. It follows from Lemma 6.6 that $Γ^s_{b,P_b,Q_0}$ starts at the exterior component of the complement of $C_{prl,b}$. Let $A_0$ be the first intersection of $Γ^s_{b,P_b,Q_0}$ with $C_{prl,b}$ after leaving $P_0$. Then $A_0 \in γ_1$. Since there is no periodic point on $Γ^s_{b,P_b,Q_0}$, we have $A_0 \neq E_0$ for $1.5 < b \leq b_{crit}$ and $A_0 \notin \{E_1(b), E_2(b)\}$ for $b_{crit} < b \leq b_0$. Since the point $A_0$ is not on the diagonal, $A_0 = ρ_b(t_a)$ for some $0 < t_a \leq 1$, where $ρ_b : [-1, 1] \rightarrow Γ^s_{b,P_b,Q_0}$ is a symmetric parametrization given in Definition 7.8. Let $B_0 = Φ_b(A_0) \in γ_2$. Since $Φ_b(Γ^s_{b,P_b,Q_0}) = Γ^s_{b,P_b,Q_0}$, $B_0$ is also an intersection point of $Γ^s_{b,P_b,Q_0}$ with $C_{prl,b}$. It follows from Lemma 7.9 that $B_0 = ρ_b(t_b)$ for some $-1 < t_b < -t_a$. So $B_0$ is an intersection point of $Γ^s_{b,P_b,Q_0}$ with $C_{prl,b}$ after leaving $Q_0$ that happens before the symmetric point $I(A_0)$. This contradicts that fact that both $Γ^s_{b,P_b,Q_0}$ and $C_{prl,b}$ are symmetric with respect to the diagonal. Therefore, $Γ^s_{b,P_b,Q_0} \cap C_{prl,b} = \emptyset$ for every $1.5 < b < b_0$.

**Step 2.** Suppose $Γ^s_{b_0,P_b,Q_0} \cap C_{prl,b_0} \neq \emptyset$. Note that $b_0 > 1.5 + δ_0$. By continuity, it follows from Step 1 that the continuum connection $Γ^s_{b_0,P_b,Q_0}$ lies at the exterior component of the complement of $C_{prl,b_0}$. We will divide the remaining discussion into three subcases:

1. $b_0 > b_{crit}$. Denote by $A_0$ the first intersection of $Γ^s_{b_0,P_b,Q_0}$ and $C_{prl,b_0}$. Note that $A_0 \in γ_1$, and $A_0 \neq E_0$, since $E_0$ is enclosed in the interior of $C_{prl,b_0}$. It follows that $Γ^s_{b_0,P_b,Q_0}$ and $C_{prl,b_0}$ also intersect at $B_0 = Φ_{b_0}(A_0) \in γ_2$, which means $Γ^s_{b_0,P_b,Q_0}$ have crossed $γ_3$. It follows that the heteroclinic connection $Γ^s_{b_0,P_b,Q_0}$ has to cross both curves $C_{Ψ_b}$ and $C_{Ψ_δ}$ for $b < b_0$ sufficiently close to $b_0$, contradicting Step 1.

2. $b_0 < b_{crit}$. As in Case (2a), it suffices to show that $E_0 \notin Γ^s_{b_0,P_b,Q_0}$. Suppose on the contrary that $E_0 \in Γ^s_{b_0,P_b,Q_0}$. Recall that the point $E_0$ is an elliptic $Θ_{b_0}$-fixed point. Since $Γ^s_{b_0,P_b,Q_0}$ is $Θ_{b_0}$-invariant, it has to accumulate on $E_0$ through both sides of any smooth curve
passing through $E_0$. This again implies that the continuum connection $\Gamma^s_{b_0, P_0, Q_0}$ has crossed both curves $C_{\Phi_{b_0}}$ and $C_{\Psi_{b_0}}$, contradicting Step 1 as in Case (2a).

$(2c) \ b_0 = \crit$. It suffices to show that $E_0 \not\in \Gamma^s_{b_0, P_0, Q_0}$. Suppose on the contrary that $E_0 \in \Gamma^s_{b_0, P_0, Q_0}$. Since $\Gamma^s_{b_0, P_0, Q_0}$ is a continuum, there are points $T_n \in \Gamma^s_{b_0, P_0, Q_0} \to E_0$ through the cusp between $\gamma_1$ and $\gamma_3$, see the middle plot in Fig. 19. Note that $E_0$ is a parabolic fixed point for $\Phi_{b_0}$ with eigenvalues $\lambda_1 = -1$ and $\lambda_2 = 1$, and the eigenvector for $\lambda_1 = -1$ is $v_1 = (1, 1)$. This also follows from the fact that $\Phi_{b_0}(\gamma_1) = \gamma_2$. Then the points $\Phi_{b_0}(T_n) \in \Phi_{b_0}(\Gamma^s_{b_0, P_0, Q_0}) = \Gamma^s_{b_0, P_0, Q_0}$ get flipped to somewhere near the cusp between $\gamma_2 = \Phi_{b_0}(\gamma_1)$ and $\gamma_4 = \Phi_{b_0}(\gamma_3)$. This again contradicts Step 1 as in Case (2a).

Collecting terms, we have $\Gamma^s_{b_0, P_0, Q_0} \cap C_{pr1, b_0} = \emptyset$. This completes the proof of the lemma.

**Proposition 7.11.** If $\Gamma^s_{b_0, P_0, Q_0}$ is a continuum connection, then one of the following holds:

1. $\Gamma^s_{b_0, P_0, Q_0} \cap \partial D_{b_0} \neq \emptyset$;
2. there are $\Theta_{b_0}$-fixed points on $\Gamma^s_{b_0, P_0, Q_0}$ other than $P_0$ and $Q_0$.

Similarly, if $\Gamma^u_{b_0, P_0, Q_0}$ is a continuum connection, then there are $\Theta_{b_0}$-fixed points on $\Gamma^u_{b_0, P_0, Q_0}$ other than $P_0$ and $Q_0$.

**Proof.** We only need to consider the case that $\Gamma^s_{b_0, P_0, Q_0}$ is a continuum connection. We argue by contradiction. Suppose on the contrary that $\Gamma^s_{b_0, P_0, Q_0} \cap \partial D_{b_0} = \emptyset$ and there is no other $\Theta_{b_0}$-fixed point on $\Gamma^s_{b_0, P_0, Q_0}$ besides $P_0$ and $Q_0$. Recall that the point $(0, 0)$ is nonlinearly stable and hence cannot be accumulated by any heteroclinic connection. Let $U$ be the connected component of $M \setminus \Gamma^s_{b_0, P_0, Q_0}$ containing the origin $(0, 0)$. Pick a simply connected neighborhood $V \supset \Gamma^s_{b_0, P_0, Q_0}$ such that $\partial V \subset U$ and there is no $\Theta_{b_0}$-fixed point in $V$: if such a neighborhood does not exist, then there will be a sequence of $\Theta_{b_0}$-fixed points $T_n \in U$ converging to $\partial U \subset \Gamma^s_{b_0, P_0, Q_0}$. It follows that any limiting point of $p_n$ will be a fixed point on $\Gamma^s_{b_0, P_0, Q_0}$, say $T$. Moreover, $T \neq P_0, Q_0$ since both $P_0$ and $Q_0$ are non-degenerate. This contradicts the hypothesis that there is no other $\Theta_{b_0}$-fixed point on $\Gamma^s_{b_0, P_0, Q_0}$ besides $P_0$ and $Q_0$.

For convenience, we collapse $U \setminus V$ into a point. The quotient space $M/\sim$ can be identified with $S^2$, in which the quotient open set $\mathcal{U}$ of $U$ is a simply connected open subset of $S^2$ with boundary $\partial \mathcal{U} \subset \Gamma^s_{b_0, P_0, Q_0}$. Let $\hat{\mathcal{U}} = \mathcal{U} \sqcup S^1$ be the prime-end compactification of $\mathcal{U}$. Note that $P_0, Q_0 \in \partial \mathcal{U}$, since $P_0$ and $Q_0$ are on the boundary of the first quadrant, and $\Gamma^s_{b_0, P_0, Q_0}$ is contained in the first quadrant. Let $\hat{P}_0$ and $\hat{Q}_0$ be the prime ends on $\hat{\mathcal{U}}$ corresponding to $P_0$ and $Q_0$, respectively. These two points are those accessible points of $S^1 \subset \hat{\mathcal{U}}$.

Note that $\Theta_{b_0}$ may be not well defined on the quotient space $S^2$, since the kernel $U \setminus V$ is not necessarily $\Theta_{b_0}$-invariant. Since $\Gamma^s_{b_0, P_0, Q_0}$ is $\Theta_{b_0}$-invariant and is strictly contained in $\mathcal{R}_{b_0}$, there exists a smaller neighborhood $V_1 \subset V \cap \mathcal{U}$ on which $\Theta_{b_0}$ is well defined. Then we can define the prime-end extension $\hat{\Theta}_{b_0}$ of $\Theta_{b_0}$ from $V_1$ to $\hat{V}_1 \sqcup S^1$. Note that $\hat{P}_0, \hat{Q}_0 \in S^1$ are fixed by $\hat{\Theta}_{b_0}$, since $P_0$ and $Q_0$ are fixed by $\Theta_{b_0}$ and are accessible through $V_1$. In particular, the rotation number $\rho(\hat{\Theta}_{b_0}, S^1) = 0$, or equally, $\rho(\hat{\Theta}_{b_0}, V_1, \partial V_1) = 0$. Then it follows from Proposition 2.3 that there must be some heteroclinic connection contained in $V_1$ between the two hyperbolic periodic point $P_0$ and $Q_0$ contained in $\partial V_1$, which contradicts our choice of the parameter $b_0$. This completes the proof. □
Step 1. At the point $S_0 = (2^{-1/2}, 2^{-1/2})$, denote $L_b(S_0) = (d_{t,1}, 2^{-1/2})$ and $R_b L_b S_0 = (d_{t,1}, d_{r,1})$, where $d_{t,1} = 2^{-1/2} - b$ and $d_{r,1} = d_{t,1} - b (1 - 2d_{t,1}^2)$. Let $d_t(b) = \frac{\sqrt{8b^2 + 1} - 1}{4b}$ be the solution of equation $d_t - b (1 - 2d_t^2) = 0$, and $b_2 \approx 1.58885$ be the positive solution of $d_t(b) = 2^{-1/2} - b$, or equally, $b - \frac{\sqrt{8b^2 + 1}}{4b} = 2^{-1/2}$, see Fig. 24. Then we have $d_{r,1} = d_{t,1} - b (1 - 2d_{t,1}^2) < 0$ for $1.5 < b < b_2$.

Next we will show that none of the two alternatives for the continuum connection $\Gamma^s_{b_0,P_0,Q_0}$ in Proposition 7.11 can happen. Similarly, there is no new fixed point on $\Gamma^u_{b_0,P_0,Q_0}$. This leads to a contradiction and hence the working hypothesis (HC) must be wrong.

**Lemma 7.12.** The continuum connection $\Gamma^s_{b_0,P_0,Q_0}$ does not intersect $\partial D_{b_0}$.

**Proof.** Suppose on the contrary that $\Gamma^s_{b_0,P_0,Q_0} \cap \partial D_{b_0} \neq \emptyset$. Then one of the following holds:

1. either the stable manifold $\Gamma_{b_0,P_0,+}^s$ is a finite curve that ends at a point on $\partial D_{b_0}$; denote such a point by $R_0(d_{t,0}, d_{r,0})$. Then $R_0 \in S_{1,+}$, and $\Gamma^s_{b_0,P_0,R_0}$ is contained in $D_{b_0}$ and intersects $\partial D_{b_0}$ tangentially at $R_0$;

2. or $\Gamma_{b_0,P_0,+}^s$ is an infinite curve with some accumulation points on $\partial D_{b_0}$; pick the accumulation point with smallest $d_t$-coordinate and denote it by $R_0(d_{t,0}, d_{r,0})$. Then $R_0 \in S_{1,+}$.

Note that $R_0 \in S_{1,+}$ implies $d_{t,0} \leq 2^{-1/2} \leq d_{r,0}$. Let $T_0 = (d_{t,0}, d_{r,0})$, which is a tangential intersection point of $\Gamma^u_{b_0,Q_0,+}$ with $S_{2,+}$ in Case (1), and is an accumulation point of $\Gamma^u_{b_0,Q_0,+}$ on $S_{2,+}$ in Case (2). Since our arguments below work the same for both cases, we will not distinguish these two cases, denote the corresponding branches by $\Gamma^s_{b_0,P_0,*}$ and $\Gamma^u_{b_0,Q_0,*}$ and call $R_0$ an intersection from now on. Denote

1. $R_1(d_{t,1}, d_{r,0}) := L_b(R_0)$, which is the intersection of $\Gamma^u_{b_0,Q_1,*} = L_b \Gamma^s_{b_0,P_0,*}$ with $L_b S_{1,+} \subset L_b$;

2. $R_2(d_{t,1}, d_{r,1}) := R_b(R_1)$, which is the intersection of $\Gamma^s_{b_0,P_1,*} = R_b \Gamma^u_{b_0,Q_1,*}$ with $S_{2,-} \subset R_b(L_b)$;

3. $T_1(d_{t,2}, d_{r,0}) := L_b(T_0)$, which is the intersection of $\Gamma^u_{b_0,Q_0,*} = L_b \Gamma^s_{b_0,P_1,*} \subset L_b S_{2,+}$.

We will derive a contradiction in three steps, each step excluding a possibility for $b_0$. Putting them together, we conclude that the hypothesis $\Gamma^s_{b_0,P_0,Q_0} \cap \partial D_{b_0} \neq \emptyset$ does not hold.

**Step 1.** At the point $S_0 = (2^{-1/2}, 2^{-1/2})$, denote $L_b(S_0) = (d_{t,1}, 2^{-1/2})$ and $R_b L_b S_0 = (d_{t,1}, d_{r,1})$, where $d_{t,1} = 2^{-1/2} - b$ and $d_{r,1} = d_{t,1} - b (1 - 2d_{t,1}^2)$. Let $d_t(b) = \frac{\sqrt{8b^2 + 1} - 1}{4b}$ be the solution of equation $d_t - b (1 - 2d_t^2) = 0$, and $b_2 \approx 1.58885$ be the positive solution of $d_t(b) = 2^{-1/2} - b$, or equally, $b - \frac{\sqrt{8b^2 + 1}}{4b} = 2^{-1/2}$, see Fig. 24. Then we have $d_{r,1} = d_{t,1} - b (1 - 2d_{t,1}^2) < 0$ for $1.5 < b < b_2$. 

![Figure 23](image-url)
It follows from Proposition 7.7 that the heteroclinic connection \( \Gamma_{b,P_0,Q_0}^s \) is contained in the first quadrant for \( 1.5 < b < b_0 \). By taking limit \( b \to b_0 \), we conclude that the continuum connection \( \Gamma_{b_0,P_0,Q_0}^s \) is contained in the first quadrant, which implies \( L_b(\Gamma_{b_0,P_0,Q_0}^s) = \Gamma_{b_0,Q_1,P_1}^u \) is contained in the second quadrant. Since \( d_{e,1} = d_{e,1} - b(1 - 2d_{e,1}^2) < 0 \) for \( 1.5 < b < b_2 \), it is impossible for \( \Gamma_{b_0,Q_1,P_1}^u \) to have an intersection at \( R_0 \in S_{1,+} \) if \( b_0 < b_2 \). This completes Step 1.

**Step 2.** Let \((d_{e,1}, d_{e,1}) = R_bL_b(S_0)\), and \(b_3 \approx 1.62326\) be the parameter \(b\) with \(d_{e,3} := (\Phi_b(d_{e,1}, d_{e,1}))_1 = 0\), see Fig. 24. Then \(d_{e,3} < 0\) for \(b_2 < b < b_3\). Let \(\hat{d}_r := b - \frac{1 + \sqrt{8b^2 + 1}}{4b} \in (2^{-1/2}, 1)\), which satisfies \(R_b(\hat{d}_r - b, \hat{d}_r) = (\hat{d}_r - b, 0)\). Denote \(\eta_{1,} = \{(d_r + b(1 - 2d_r^2), d_r) : 2^{-1/2} < d_r < \hat{d}_r\} \subset S_{1,+}\), which is the part satisfying that \(\eta_{2,-} := R_bL_b(\eta_{1,+})\) is contained in the second quadrant. In the following we will work with a slightly smaller bound \(b_3 = 1.622 < b_3\).

![Figure 24. Some plots of the orbit of \(S_0\). Left: \(b = 1.58885\). Middle: \(b = 1.62326\). Right: \(b = 1.622\).](image)

**Claim.** For \(b_2 < b < \hat{b}_3 = 1.622\), the curve \(\Phi_b(\eta_{2,-})\) is contained in the third quadrant.

**Proof of the Claim.** Note that two endpoints of \(\eta_{2,-} = R_bL_b(\eta_{1,+})\) are given by

\[
\begin{align*}
R_bL_b(S_0) &= R_b(2^{-1/2} - b, 2^{-1/2}) = (2^{-1/2} - b, 2^{-1/2} - b - 2\sqrt{2b^2 + 2b^3}); \\
R_bL_b(\hat{d}_r + b(1 - 2d_r^2), \hat{d}_r) &= R_b(\hat{d}_r - b, \hat{d}_r) = (\hat{d}_r - b, 0).
\end{align*}
\]

Note that the quantity \(2^{-1/2} - b - 2\sqrt{2b^2 + 2b^3}\) is monotone increasing with respect to \(b\) on the interval \(b_2 \leq b \leq \hat{b}_3\), on which it satisfies \(0 \leq 2^{-1/2} - b - 2\sqrt{2b^2 + 2b^3} < 0.18\). It is clear that \(L_b(\hat{d}_r - b, 0) = (b - \hat{d}_r, 0)\) and \(R_b(b - \hat{d}_r, 0) = (b - \hat{d}_r, -\hat{d}_r)\). Moreover, the point \(L_b(b - \hat{d}_r, -\hat{d}_r)\) is in the third quadrant. By the choice of \(b_3\) and the fact that \(b_3 < b_3\), we have that \(\Phi_b(2^{-1/2} - b, 2^{-1/2} - b - 2\sqrt{2b^2 + 2b^3})\) is in the third quadrant when \(b_2 \leq b \leq b_3\). So to prove the claim, it suffices to show that \(\Phi_b(\eta_{2,-})\) does not intersect the line \(d_{e,1} = 0\), which is equivalent to that \(R_bL_b(\eta_{2,-})\) does not intersect the curve \(\{L_b(0, d_r) : -\frac{1 + \sqrt{8b^2 + 1}}{4b} < d_r < 0\}\) (see the gray curve in the four quadrant of the plot \(b = 1.622\) in Fig. 24). We just need to compare the first coordinates of these two curves. Since the map \(R_b\) does not change the first coordinate, we can even use \(L_b(\eta_{2,-})\) instead of \(R_bL_b(\eta_{2,-})\).
(1) Let \((d_\ell, d_r) \in \eta_{2,-}\), where \(d_\ell = -1 + \sqrt{8b^2 + 8bd_r + 1}/4b\) and \(0 \leq d_r \leq 2^{-1/2} - b - 2\sqrt{2b^2 + 2b^3}\). Then the component \(L_b(d_\ell, d_r)_1\) consists of two parts \(I + II\), where

\[
I = d_r + (1 - 2d_r^2) \left(\frac{(8b^2 + 8bd_r + 1)^{1/2} + 1}{4b} + d_r\right),
\]

\[
II = -2d_r(1 - d_r^2)^{1/2} \left(b^2 - \left(\frac{(8b^2 + 8bd_r + 1)^{1/2} + 1}{4b} + d_r\right)^2\right)^{1/2}.
\]

Note that for fixed \(d_r\), it is easy to see that \(I\) is monotone decreasing with respect to \(b\) (as long as \(0 < d_r < 2^{-1/2}\)). For \(II\), it is also monotone decreasing since \(b^2 - \left(\frac{(8b^2 + 8bd_r + 1)^{1/2} + 1}{4b} + d_r\right)^2\) is monotone increasing. Therefore, the minimum of \(L_b(\eta_{2,-})\) is achieved when \(b = \hat{b}_3\). Plugging \(b = \hat{b}_3\), we have that \(L_b(d_\ell, d_r)_1 \geq 0.78143\) for every \((d_\ell, d_r) \in \eta_{2,-}\).

(2) It is easy to see that the component \(L_b(0, d_r)_1\) is smaller than the minimum of \(L_b(d_\ell, d_\ell - b(1 - 2d_r^2))_1\) for \(b_3 \leq b \leq \hat{b}_3\). It follows that these two curves do not intersect. This completes the proof of the claim.

Collecting terms, we conclude that the max of \(L_b(\{0\} \times [-1, 0])_1\) is smaller than the minimum of \(L_b(d_\ell, d_\ell - b(1 - 2d_r^2))_1\) for \(b_3 \leq b \leq \hat{b}_3\). It follows that these two curves do not intersect. This completes the proof of the claim.

Note that an intersection of \(\Gamma^s_{b_0,P_0,Q_0}\) with \(S_{1,\pm}\), if happens, lies in the curve \(\eta_{1,+}\). Moreover, \(\Phi_b R_b L_b(\Gamma^s_{b_0,P_0,Q_0})\) is a continuum connection from \(Q_2\) to \(P_2\) and hence contained in the fourth quadrant. Since \(\Gamma^s_{b_0,P_0,Q_0}\) and \(S_{1,\pm}\) do not intersect for \(b_2 < b \leq \hat{b}_3\), it is impossible to have \(b_2 \leq b_0 \leq \hat{b}_3\). This completes Step 2.

**Step 3.** Next we show that it is impossible to have \(\hat{b}_3 \leq b_0 \leq b_{\text{max}}\). We claim that the curve \(L_b(S_{2,-}(b))\) intersects the curve \(C_{\text{prl,b}}\) for \(\hat{b}_3 \leq b \leq b_{\text{max}}\). See the plot \(b = 1.622\) in Fig. 24, in which the dashed curves are \(L_b(S_{2,\pm}(b))\).

**Proof of the claim.** Consider the point \((d_\ell, d_r)\) on \(S_{2,-}(b)\) with \(0 \leq d_r \leq b_{\text{max}} - 1\). As we have seen in Eq. (7.5) and (7.6) in Step 2, \(L_b(d_\ell, d_r)_1\) is monotone decreasing with respect to \(b\) for each fixed \(d_\ell\). To prove the claim, it suffices to consider the case that \(b = \hat{b}_3\).

The part \(\gamma_3 \subset C_{\text{prl,b}}\) is given by the reflected version of Eq. (6.22) (since Eq. (6.22) itself corresponds to \(\gamma_1\)), in which the coordinate \((\alpha, \beta)\) satisfies

\[
b^2 = 1 + \frac{\sin \alpha + \sin \beta}{\sin(\alpha + \beta)} + \frac{\sin \alpha \sin \beta}{\sin^2(\alpha + \beta)} + \frac{(\sin \alpha - \sin \beta)^2}{\sin^2(2\alpha + 2\beta)} = 1.622^2.
\]

See (5.4). For concreteness, we set \(d_r = 0.45\). Then it follows from (6.22) and (7.7) that \(\sin \alpha = 0.37315\) and \(d_\ell = \sin \beta = 0.679152\).

On the other hand, for \((d_\ell, d_r) \in S_{2,-}\) with \(d_r = 0.45\), we have \(d_\ell = -1 + \sqrt{8b^2 + 8bd_r + 1}/4b \approx -0.968056\) and \(L_b(d_\ell, d_r)_1 \approx 0.660888 < d_\ell = 0.679152\). It follows that \(L_b(S_{2,-}(b))\) already crosses the curve \(C_{\text{prl,b}}\) before reaching \(d_r = 0.45\) for all \(\hat{b}_3 \leq b \leq b_{\text{max}}\). This completes the proof of the claim.
Note that the unstable manifold $\Gamma_{b_0,Q_0}^u$ cannot cross the curve $\mathcal{L}_b(S_{2,-}(b_0))$, since such a crossing would imply that $\Gamma_{b,P_1,Q_1}^u$ crosses the boundary $S_{2,-}(b)$ for any $b$ sufficiently close to $b_0$, contradicting our choice of the parameter $b_0$. If $b_s \leq b_0 \leq b_{\text{max}}$, then $\Gamma_{b_0,Q_0}^u$ has to follow $\mathcal{L}_b(S_{2,-})$ and intersect the curve $\mathcal{C}_{prl,b_0}$ in order for it to reach the boundary $\partial \mathcal{D}_{b_0}$, which contradicts Lemma 7.10. Therefore, it is impossible to have $b_3 \leq b_0 \leq b_{\text{max}}$. This completes Step 3.

Collecting the conclusions from these three steps, we have exhausted all possibilities for $1.5 < b_0 \leq b_{\text{max}}$ under the hypothesis that $\Gamma_{b_0,P_0,Q_0}^s \cap \partial \mathcal{D}_{b_0} \neq \emptyset$. This completes the proof of the lemma. \hfill \Box

**Lemma 7.13.** There is no other $\Theta_{b_0}$-fixed point on $\Gamma_{b_0,P_0,Q_0}^s$ or $\Gamma_{b_0,P_0,Q_0}^u$ besides $P_0$ and $Q_0$.

**Proof.** We only need to consider the case $\Gamma_{b_0,P_0,Q_0}^s$ is a continuum connection. Suppose on the contrary that there were other $\Theta_{b_0}$-fixed points on $\Gamma_{b_0,P_0,Q_0}^s$. Pick one of them, say $T_0(d_{\ell,0},d_{r,0})$. Since $\Gamma_{b_0,P_0,Q_0}^s$ is symmetric with respect to the diagonal, we assume $d_{\ell,0} \leq d_{r,0}$ without losing any generality. Let $T_1(d_{\ell,1},d_{r,1}) := \Phi_b(T_0)$. Since $\Gamma_{b_0,P_0,Q_0}^s$ lies at the exterior of $\mathcal{C}_{prl,b_0}$, it follows from Lemma 6.7 that
\[
d_{\ell,1} - d_{r,1} > d_{r,0} - d_{\ell,0} \geq 0. \tag{7.8}
\]
Note that $T_1 = \Phi_b(T_0) \in \Phi_b(\Gamma_{b_0,P_0,Q_0}^s) = \Gamma_{b_0,P_0,Q_0}^s$, which also lies at the exterior of $\mathcal{C}_{prl,b_0}$. Applying Lemma 6.8 to $T_2(d_{\ell,2},d_{r,2}) := \Psi_b(T_1)$, we have
\[
d_{r,2} - d_{\ell,2} > d_{\ell,1} - d_{r,1}. \tag{7.9}
\]
However, since $T_0$ is a periodic point of period 6, we have $T_2 = \Psi_b(T_1) = \Psi_b \Phi_b(T_0) = T_0$, and hence $d_{r,2} - d_{\ell,2} = d_{r,0} - d_{\ell,0}$, contradicting (7.8) and (7.9). \hfill \Box

**Proof of Theorem 1.2.** Let $Z$ be the set of parameters $b \in (1.5, 1.5 + \delta_0)$ such that $P_0(b)$ has no homoclinic point. Suppose (HC) holds that the set $Z$ has a limiting point in $(1.5, 1.5 + \delta_0)$. Collecting the results from Proposition 7.11, Lemma 7.12 and Lemma 7.13, we conclude that none of the heteroclinic connections $\Gamma_{b_0,P_0,Q_0}^s$ or $\Gamma_{b_0,P_0,Q_0}^u$ breaks down at any $1.5 + \delta < b < b_{\text{max}} = 1 + 2^{-1/2}$. So $b_0 \geq b_{\text{max}}$. On the other hand, as $b$ increases from $b_{\text{crit}}$ to $b_{\text{max}}$, the exterior corner point $E_1(b) \in \mathcal{C}_{prl,b}$ slides along the diagonal and already passes through the point $S_0$ at $b = b_{\text{sing}} \approx 1.67892$. It follows that the exterior part of the domain $\mathcal{D}_{b} \cap \mathcal{M}^+ \setminus \mathcal{C}_{prl,b}$ breaks into two connected components containing $P_0$ and $Q_0$, respectively for every $b \geq b_{\text{sing}}$. Since $\Gamma_{b_0,P_0,Q_0}^s$ connects $P_0$ and $Q_0$ through the exterior part of the domain $\mathcal{D}_{b} \cap \mathcal{M}^+ \setminus \mathcal{C}_{prl,b}$, we must have $b_0 \leq b_{\text{sing}}$, contradicting the fact that $b_0 \geq b_{\text{max}}$. Therefore, the hypothesis (HC) does not hold. Combining with Proposition 7.4, we complete the proof. \hfill \Box

**An alternative proof of Theorem 1.2.** It suffices to show that it is impossible to have $b_0 \geq b_{\text{max}}$. Note that $\Gamma_{b_0,P_0,Q_0}^s$ does not intersect the set $\mathcal{C}_{prl,b} \cup \partial \mathcal{D}_b$ for every $1.5 < b < b_{\text{max}}$. It follows that the whole segment $I = \{(x,x) : 0 \leq x \leq 2^{-1/2}\}$ is contained in the domain $U$ that is invariant under both $\mathcal{L}_b$ and $\mathcal{R}_b$. In particular, $\mathcal{L}_b(I) \subset U$. On the other hand, $\mathcal{L}_b(x,x) = (x-2bx(1-x^2)^{1/2},x)$, and $\mathcal{L}_{\text{max}}(0.6, 0.6) = (-1.03882, 0.6)$ already exceeds the domain $U$, contradicting the fact that $\mathcal{L}_{\text{max}}(I) \subset U$. \hfill \Box
REFERENCES

[1] P. Berger, D. Turaev. *On Herman’s positive entropy conjecture*. Adv. Math. 349 (2019), 1234–1288.
[2] A. Blumenthal, J. Xue, L.-S. Young. *Lyapunov exponents for random perturbations of some area-preserving maps including the standard map*. Ann. of Math. (2) 185 (2017), no. 1, 285–310.
[3] L. E. J. Brouwer. *Beweiss des ebenen Translationssatzes*. Math. Ann. 72(1912), 37–54.
[4] L. Bunimovich, H.-K. Zhang, P. Zhang. *On another edge of defocusing: hyperbolicity of asymmetric lemon billiards*. Comm. Math. Phys. 341 (2016), 781–803.
[5] K. Burns, H. Weiss. *A geometric criterion for positive topological entropy*. Comm. Math. Phys. 172 (1995), no. 1, 95–118.
[6] J. Chen, L. Mohr, H.-K. Zhang, P. Zhang. *Ergodicity and coexistence of elliptic islands in a family of convex billiards*. Chaos 23 (2013), 043137.
[7] N. Chernov, R. Markarian. *Chaotic billiards*. Math. Surv. Monogr. 127, AMS, Providence, RI, 2006.
[8] A. Cima, A. Gasull, V. Manosa. *Non-integrability of measure preserving maps via Lie symmetries*. J. Differential Equations 259 (2015), no. 10, 5115–5136.
[9] G. L. da Silva Ritter, A. M. Ozorio de Almeida, R. Douady. *Analytical determination of unstable periodic orbits in area preserving maps*. Phys. D 29 (1987), no. 1-2, 181–190.
[10] J. Franks. *A new proof of the Brouwer plane translation theorem*. Ergodic Theory Dynam. Systems 12 (1992), no. 2, 217–226.
[11] E. Heller and S. Tomsovic. *Postmodern quantum mechanics*. Physics Today 46 (1993), 38–46.
[12] X. Jin, P. Zhang. *Hyperbolicity of asymmetric lemon billiards*. Nonlinearity 34 (1) (2021), 92–117.
[13] X. Jin, P. Zhang. *Birkhoff Normal Form and Twist coefficients of periodic orbits of billiards*. arXiv:2108.12098.
[14] S. O. Kamphorst, S. Pinto-de-Carvalho. *The first Birkhoff coefficient and the stability of 2-periodic orbits on billiards*. Experiment. Math. 14 (2005), no. 3, 299–306.
[15] A. Katok. *Five most resistant problems in dynamics*. MSRI-Evans Lecture, Berkeley, September 2004.
[16] A. Katok. *Fifty years of entropy in dynamics: 1958–2007*. J. Mod. Dyn. 1 (2007), no. 4, 545–596.
[17] A. Katok, B. Hasselblatt. *Introduction to the modern theory of dynamical systems*. Encyclopedia of Mathematics and its Applications, 54. Cambridge University Press, Cambridge, 1995.
[18] V. F. Lazutkin. *Splitting of complex separatrices*. Funct. Anal. Appl. 22 (1988), no. 2, 154–156.
[19] C. Liverani, M. Wojtkowski. *Ergodicity in Hamiltonian systems*, in “Dynamics reported”, Expositions Dynam. Systems (N.S.), 4, Springer, Berlin, 130–202, 1995.
[20] G. Lowther. *Integrability of the Cohen map*. https://mathoverflow.net/q/94321, 2012.
[21] J. Mather. *Invariant subsets of area-preserving homeomorphisms of surfaces*. Advances in Math. Suppl. Studies 7B (1981), 531–561.
[22] J. Mather, K. Moser. *Topological proofs of some purely topological consequences of Caratheodory’s theory of prime ends*. In: Selected Studies, Eds. Th. M. Rassias, G. M. Rassias, 1982, 225–255.
[23] K. R. Meyer. *Generic bifurcation of periodic points*. Trans. Amer. Math. Soc. 149 (1970), 95–107.
[24] K. R. Meyer, D. C. Offin. *Introduction to Hamiltonian dynamical systems and the N-body problem*. Third edition. Applied Mathematical Sciences, 90. Springer, Cham, 2017.
[25] J. K. Moser. *Stable and Random Motions in Dynamical Systems*. With special emphasis on celestial mechanics. Reprint of the 1973 original. Princeton, NJ, 2001.
[26] M. Rychlik, M. Torgerson. *Algebraic non-integrability of the Cohen map*. New York J. Math. 4 (1998), 57–74.
[27] C. L. Siegel, J. K. Moser. *Lectures on Celestial Mechanics*. Berlin: Springer-Verlag, 1971.
[28] Y. G. Sinai. *Topics in ergodic theory*. Princeton Mathematical Series, 44, Princeton Univ. Press, 1994.
[29] S. Smale. *Diffeomorphisms with many periodic points*. Differential and Combinatorial Topology, Princeton Univ. Press, 63–80, 1965.
[30] S. Tabachnikov. *Geometry and billiards*. Student Mathematical Library, 30. American Mathematical Society, Providence, RI; Mathematics Advanced Study Semesters, University Park, PA, 2005.
[31] M. Wojtkowski. *Invariant families of cones and Lyapunov exponents*. Ergodic Theory Dynam. Systems 5 (1985), no. 1, 145–161.
[32] Z. Xia, P. Zhang. *Homoclinic points for convex billiards*. Nonlinearity 27 (2014), 1181–1192.
Math Department, Boston College, Chestnut Hill, MA 02467.

Email address: xin.jin@bc.edu

Department of Mathematics, University of Oklahoma, Norman, OK 73019.

Email address: pengfei.zhang@ou.edu