Journal of Statistical Physics, Vol. 83, Nos. 1/2, 1996

Asymptotic Dynamics of the Dual Billiard Transformation

Serge Tabachnikov[1]

Received December 22, 1994; final May 31, 1995

Given a strictly convex plane curve, the dual billiard transformation is the transformation of its exterior defined as follows: given a point x outside the curve, draw a support line to it from the point and reflect x at the support point. We show that the dual billiard transformation far from the curve is well approximated by the time 1 transformation of a Hamiltonian flow associated with the curve.

KEY WORDS: Dual billiards; Hamiltonian flow; KAM theory; invariant curves.

1. INTRODUCTION

The dual billiard map T is a transformation of the exterior of a strictly convex closed plane curve γ. Given a point x outside of γ, there are two supporting lines to γ through x; choose one of them (say, the right one from x's viewpoint) and define $T(x)$ to be the reflection of x at the point of support; see Fig. 1. The curve γ is called the dual billiard curve.

The dual billiard map is an outer counterpart of the usual billiard ball map (whence another term, outer billiard, also used by some authors). See refs. 2–8, 10–13 for various aspects of the dual billiard problem, such as existence and nonexistence of invariant curves, polygonal dual billiards, multidimensional dual billiards, etc.

The present paper concerns the following phenomenon observed in the numerical study of dual billiards. Given a dual billiard curve γ, one wants to study the dynamics of the dual billiard transformation very far from the curve. To this end one rescales the plane by choosing an origin O inside γ

[1] University of Arkansas, Fayetteville, Arkansas 72701, and Wolfson College and I. Newton Institute, University of Cambridge, Cambridge, England.

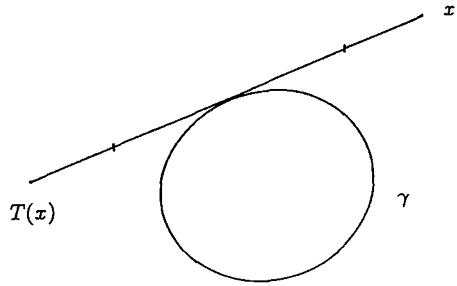

Figure 1

and applying a dilation centered at O with a small coefficient ε. Then one picks a point x at distance of order 1 from the origin (i.e., comparable with the size of the computer screen), and iterates the square of the dual billiard transformation starting at x. The experimental fact is that for the number of iterations of order $1/\varepsilon$ the orbit of x traces a closed, centrally symmetric convex curve, and these curves for different starting points are obtained from one another by dilations centered at O. Moreover, the motion of iterates of x appears continuous (assuming each iteration takes the same small amount of time), and this continuous motion satisfies the law of areas: the area swept by the position vector of a point grows linearly in time.

Moser[7] introduced dual billiards as a crude model for planetary motion. The above-described law of areas plays the role of Kepler's law for dual billiards. It means that dual billiard dynamics is, in the first approximation, a motion subject to a central force whose potential depends on the dual billiard curve.

To explain these phenomena we associate to a dual billiard curve a homogeneous Hamilton function in the plane and show that its Hamiltonian flow approximates, in a certain sense, the dual billiard dynamics (the centrally symmetric curves traced in computer experiments by orbits of the dual billiard transformation are level curves of this Hamiltonian). We emphasize that this picture is only an approximation of a more complicated dynamics of dual billiards, valid for small values of ε and for the number of iterations of order $1/\varepsilon$.

An analysis of dual billiard dynamic "at infinity" for smooth enough dual billiard curves was undertaken in ref. 3. The result, similar to Lazutkin's for usual billiards near the boundary, is that the dual billiard transformation is a small perturbation of an integrable Hamiltonian system, and the KAM theory yields an abundance of invariant curves sufficiently far from the table. The homogeneous Hamiltonian constructed in the present paper provides a normal form for an application of the KAM methods.

The content of the paper is as follows. In Section 1 we construct the Hamiltonian and study its properties; in Section 2 we show how the Hamiltonian flow approximates the dual billiard dynamics. The results of this paper were announced in refs. 12 and 13.

1. HAMILTONIAN FLOW

As a motivation for what follows, consider the square of the dual billiard transformation; see Fig. 2. The vector $T^2(x) - x = 2(b - a)$, where a and b are the tangency points. Observe that if the point x is very far from γ the lines $(x, T(x))$ and $(T(x), T^2(x))$ are almost parallel.

Fix a polar coordinate system (α, r) in the plane whose origin O lies inside the dual billiard curve γ. Given an angle α, consider the two parallel oriented supporting lines to γ in the direction α; order them so that γ lies to the left of the first one. Let $v(\alpha)$ be the vector joining the points of support of these lines (see Fig. 3); this vector may be thought of as the limit case of the vector $b - a$ in Fig. 2 as $x \to \infty$. Consider the homogeneous vector field in the punctured plane whose value at point (α, r) is $v(\alpha)$. Abusing notation, we call it v.

Given an angle α, consider the oriented supporting line l to γ in the direction α that lies to the right from γ. Let $p(\alpha)$ be the supporting function of γ, that is, the distance from the origin to l; see Fig. 4.

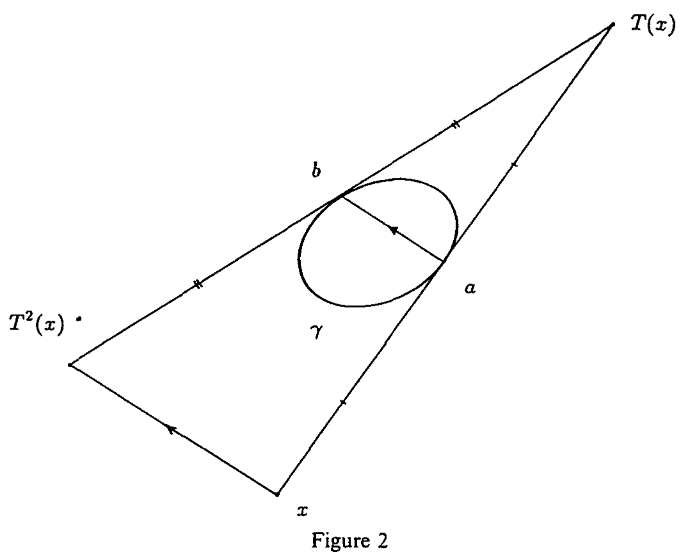

Figure 2

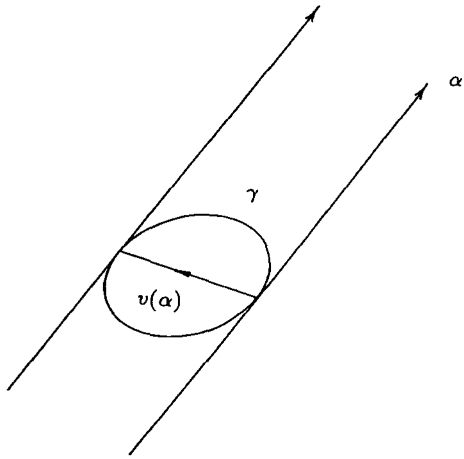

Figure 3

Let $q(\alpha) = p(\alpha) + p(\alpha + \pi)$ be the width of γ in the direction perpendicular to α. If γ is smooth and strictly convex, which we assume, the function $q(\alpha)$ is also smooth [at the other extreme, if γ happens to contain a straight segment, then the derivative $q(\alpha)'$ has a jump discontinuity at the point corresponding to the direction of this segment].

Theorem. The vector field v is Hamiltonian with the Hamilton function $H = rq(\alpha)$. Its trajectories are closed, centrally symmetric curves, and its flow satisfies the law of areas.

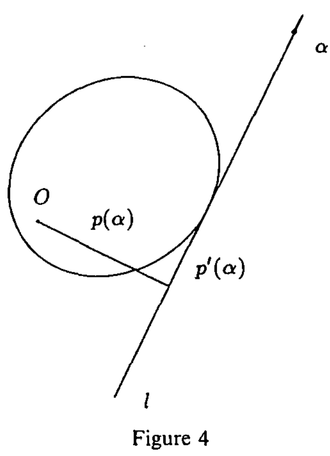

Figure 4

Proof. The Cartesian coordinates of the tangency point of the line l in Fig. 4 are given by

$$x(\alpha) = p(\alpha)\sin\alpha + p'(\alpha)\cos\alpha$$
$$y(\alpha) = -p(\alpha)\cos\alpha + p'(\alpha)\sin\alpha$$

(see, e.g., ref. 9, which is also a good reference for the simple facts from convex geometry we use in the paper). Hence the vector $v(\alpha)$ is

$$\begin{aligned}&(x(\alpha+\pi) - x(\alpha), y(\alpha+\pi) - y(\alpha))\\ &\quad = (q(\alpha)\sin\alpha + q'(\alpha)\cos\alpha, -q(\alpha)\cos\alpha + q'(\alpha)\sin\alpha)\end{aligned}$$

The Hamiltonian vector field *sgrad* H of a Hamilton function H is $(H_y, -H_x)$. Using the chain rule, one finds for $H = rq(\alpha)$

$$H_y = q(\alpha)\sin\alpha + q'(\alpha)\cos\alpha, \qquad H_x = q(\alpha)\cos\alpha - q'(\alpha)\sin\alpha$$

Thus $v = sgrad\, H$.

The trajectories of *sgrad* H are level curves of H which are centrally symmetric and homothetic to each other because H is a homogeneous function.

The law of areas holds for any homogeneous Hamiltonian (not necessarily of degree 1). Indeed, let $z(t) = (x(t), y(t))$ be the trajectory of a point as a function of time, and let $A(t)$ be the sectorial area. Then $A'(t) = 0.5[z'(t), z(t)]$ where [,] is the skew-product of two vectors, i.e., the determinant of the matrix whose columns are the vectors. Since $z'(t) = (H_y, -H_x)$ one has $A'(t) = 0.5(xH_x + yH_y)$. The Euler formula for a homogeneous function of degree k reads $xH_x + yH_y = kH$. Thus $A'(t) = 0.5kH$. Since H is invariant under *sgrad* H this is constant along a trajectory. Q.E.D.

Level curves of the Hamiltonian $rq(\alpha)$ can be described geometrically using the notion of polar duality. Polar duality is the correspondence between points of the punctured plane and lines not through the origin: a line l corresponds to the point X such that OX is perpendicular to l and the distance from O to l equals $1/\|OX\|$. Given a convex plane curve γ that contains the origin in its interior, one defines the dual curve γ^* as follows: tangent lines to γ form a one-parameter family of lines, and γ^* consists of points polar dual to these lines. If $p(\alpha)$ is the supporting function of γ, then γ^* has the equation in polar coordinates $r = 1/p(\alpha)$.

Note that $\gamma^{**} = \gamma$. Also note that polar duality interchanges two types of singularities of curves: to a corner of γ a line segment of γ^* corresponds, and vice versa.

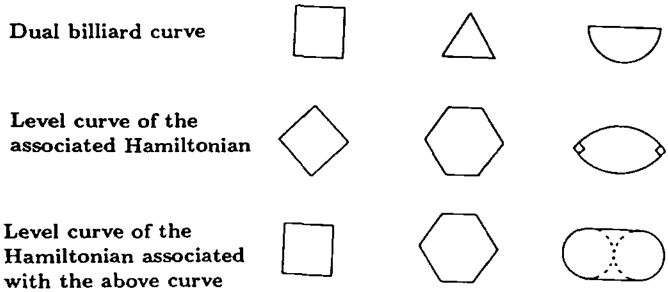

Figure 5

Thus trajectories of the Hamiltonian vector field under consideration are obtained from the dual billiard curve by central symmetrization, i.e., replacing the supporting function $p(\alpha)$ by $q(\alpha) = p(\alpha) + p(\alpha + \pi)$, polar dualization, and homothety. In particular, if the dual billiard curve γ is centrally symmetric, then the second iteration of this construction yields the curves homothetic to γ.

Example. If γ is a triangle or a square, then the corresponding centrally symmetric curve is a hexagon or a square. A somewhat more sophisticated example is that of a semicircle. In this case the Hamiltonian is $|y| + (x^2 + y^2)^{1/2}$ and its level curves consist of two symmetric arcs of parabolas with the common focus at the origin; the parabolas intersect at right angles; see Fig. 5, in which we also show the curves obtained by two iterations of the construction.

Remark. Centrally symmetric polygons approximating trajectories of the dual billiard map for a polygonal dual billiard curve were studied in refs. 6, 10, and 5 with regard to the stability property of the dual billiard transformation.

2. HAMILTONIAN FLOW APPROXIMATING DUAL BILLIARD DYNAMICS

We will estimate the deviation of the T^2-orbit of a point from its trajectory under the Hamiltonian vector field v with the Hamilton function $rq(\alpha)$. Denote the time 2 map of the flow of v by f and set $g = T^2$.

Theorem. Given $\omega > 0$, there exists a constant $C(\gamma, \omega)$ such that $dist(f^n(x), g^n(x)) < C(\gamma, \omega)$ for every point x and all $n \leqslant \omega \|x\|$.

In less technical terms, the theorem states that the deviation of $T^{2n}(x)$ from the time $2n$ image of x in the Hamiltonian flow is uniformly bounded for the number of iterations of order $\|x\|$. The number of iterations it takes the T^2-orbit of x to make a complete turn about γ is of this order. Rescaling by the factor $\varepsilon = 1/\|x\|$, one makes the deviation negligible. Thus, on time scale $[0, 1/\varepsilon]$, the rescaled dual billiard dynamics is well approximated by the Hamiltonian one, which explains the features of the dual billiard dynamics described in the Introduction.

Remark. Berry[1] obtained similar estimates for the dual billiard dynamics far from the dual billiard curve.

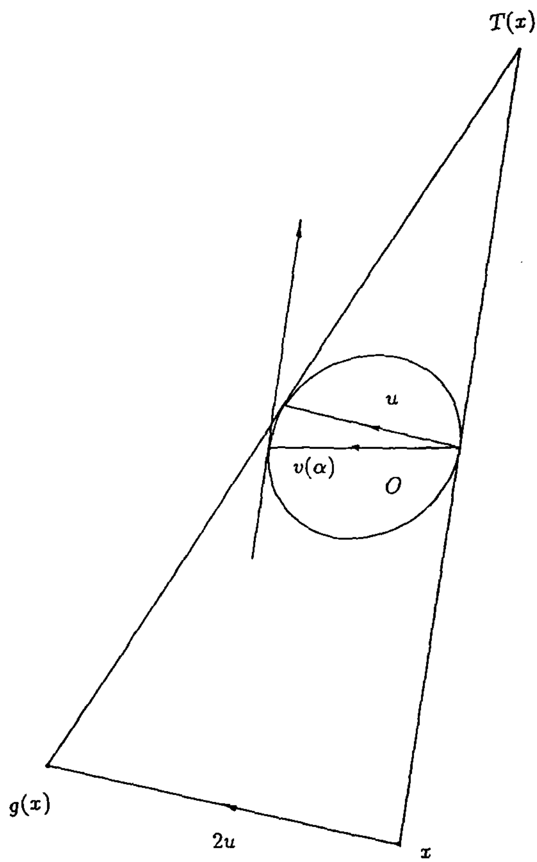

Figure 6

The proof of the theorem consists in somewhat notorious estimates, and we will outline its main steps. The idea is, however, quite simple: the deviation of $g(x)$ from $f(x)$ is of order $1/\|x\|$ for all x, therefore the deviation of $g^n(x)$ from $f^n(x)$ is of order 1 for n of order $\|x\|$.

Outline of Proof. We use the "big O" notation; various constants denoted by C, C_1, etc., depend on ω and γ.

Step 1. The first estimate needed is

$$dist(f(x), g(x)) \leqslant \frac{C_1}{H(x)+H(g(x))} \tag{1}$$

for all x.

To obtain this inequality, introduce the vector $u=0.5(g(x)-x)$, so that $g(x)=x+2u$; see Fig. 6. The angle $(x, T(x), T^2(x))$ is $O((H(x)+H(g(x)))^{-1})$, hence

$$\|u-v(\alpha)\| = O\left(\frac{1}{H(x)+H(g(x))}\right) \tag{2}$$

Compare the orbits of x under the flows ϕ^t and ϕ_0^t of the fields v and the constant field v_0 whose value at all points is $v(\alpha)$, respectively. The deviation $\|\phi^t(x)-\phi_0^t(x)\|$ is $O((H(x)+H(g(x)))^{-1})$ for $t \leqslant 2$. Hence

$$dist(f(x), x+2v(\alpha)) = O\left(\frac{1}{H(x)+H(g(x))}\right) \tag{3}$$

The estimates (2) and (3) imply (1).

Step *2*. Next, one estimates the deviation of $H(g^n(x))$ from $H(f^n(x))=H(x)$ (the equality holds because H is $f-$invariant). Setting $h_n=H(g^n(x))$, one obtains for all i the inequality

$$|h_{n+1}-h_n| \leqslant \frac{C_2}{h_n+h_{n+1}} \tag{4}$$

This follows from (1) and the inequality $|H(y+z)-H(y)| \leqslant C_3\|z\|$ uniformly in y, which is due to the fact that the norm $H(y)$ is equivalent to the norm $\|y\|$.

The summation of the inequalities (4) for $i=0,..., k$ yields

$$|h_k^2-h_0^2| \leqslant C_2 k \tag{5}$$

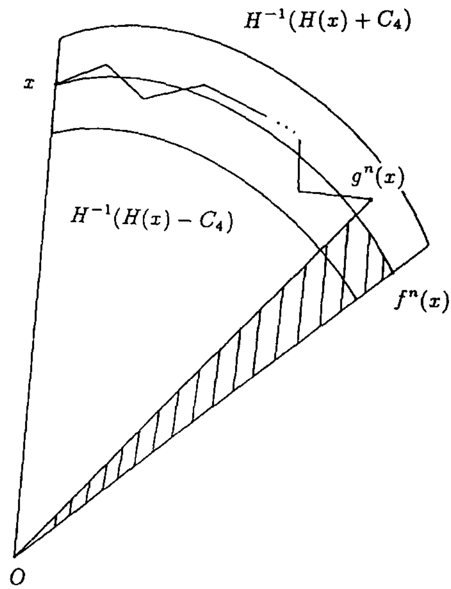

Figure 7

for all k; hence for $k \leqslant \omega\|x\|$ one has

$$|h_k - h_0| \leqslant C_2 k/H(x) \leqslant C_4 \tag{6}$$

This implies that the orbit $x, g(x), \ldots, g^n(x)$ lies in the strip between the level curves $H^{-1}(H(x)-C_4)$ and $H^{-1}(H(x)+C_4)$; see Fig. 7.

Step 3. Finally one estimates the area of the sector bounded by the broken line connecting consecutive images of x under the map g. Let $A(x)$ be the area of the triangle $(O, x, g(x))$. Then $|A(x)-H(x)| \leqslant C_5$ and, in view of (6),

$$\sum_{k=0}^{n-1} |A(g^k(x)) - nH(x)| \leqslant \sum_{k=0}^{n-1} |A(g^k(x)) - H(g^k(x))| + \sum_{k=0}^{n-1} |h_k - h_0| \leqslant C_6 H(x) \tag{7}$$

for $n \leqslant \omega\,\|x\|$.

The area bounded by the level curves $H^{-1}(H(x)-C_4)$ and $H^{-1}(H(x)+C_4)$, and the rays (O, x) and $(O, f^n(x))$, is bounded above by $C_7 H(x)$. This inequality along with (7) implies that the area of the sector bounded by the curve $H^{-1}(H(x))$ and the rays $(O, g^n(x))$ and $(O, f^n(x))$ (shaded in Fig. 7) is bounded above by $C_8 H(x)$. Hence the angle of this sector is $O(H(x)^{-1})$, and finally, $dist(g^n, (x)), f^n(x)) \leqslant C$. Q.E.D.

We would like to emphasize one consequence of the proof, namely of the inequality (5): $H^2(g^n(x)) \leqslant H^2(x) + Cn$ for all n. Since the norm $H(x)$ is equivalent to $\|x\|$, and $\|x\|$ is bounded away from zero, one obtains the following result.

Corollary. There exist constants C_1 and C_2 depending on the dual billiard curve γ such that

$$\|T^n(x)\| \leqslant (C_1 + C_2 n)^{1/2} \|x\|$$

for every point x and all n.

If the dual billiard curve is sufficiently smooth $(C^{6+\varepsilon})$ then the dual billiard map has invariant curves arbitrarily far from the curve, and therefore all orbits are bounded.(7, 8, 2, 3, 11–13) This is a much stronger result than the above corollary. However, if the curve is less smooth, it is not known whether orbits may spiral off to infinity. If such a diffusion is possible, the corollary gives an upper bound for its rate. Whether this "escape to infinity" may actually happen remains unknown at the present writing.

ACKNOWLEDGMENTS

This research would have been impossible without I. Monroe's help with the numerical study of dual billiards. This work was supported in part by NSF grant DMS-9402732. It is a pleasure to recognize the hospitality of the Wolfson College and the Newton Institute at the University of Cambridge. The author is grateful to the referees for their suggestions and comments.

REFERENCES

1. M. Berry, Private communication
2. Ph. Boyland, Dual billiards, twist maps and impact oscillators, Preprint (1994).
3. R. Douady, Thèse de Troisième Cycle, Université Paris 7 (1982).
4. E. Gutkin and A. Katok, Caustics for inner and outer billiards, *Commun. Math. Phys.* **173**:101–133 (1995).
5. E. Gutkin and N. Simanyi, Dual polygonal billiards and necklace dynamics, *Commun. Math. Phys.* **143**:431–450 (1991).
6. R. Kolodziej, The antibilliard outside a polygon, *Bull. Pol. Acad. Sci.*, **37**:163–168 (1989).
7. J. Moser, Is the solar system stable? *Math. Intell.* **1**:65–71 (1978).
8. J. Moser, Stable and random motions in dynamical systems, *Ann. Math. Stud. Princeton* (1973).

9. L. Santalo, *Integral Geometry and Geometric Probability* (Addison-Wesley, Reading, Massachusetts, 1976).
10. A. Shaidenko and F. Vivaldi, Global stability of a class of discontinuous dual billiards, *Commun. Math. Phys.* **110**:625–640 (1987).
11. S. Tabachnikov, On the dual billiard problem, *Adv. Math.* **115**:2, 221–249 (1995).
12. S. Tabachnikov, Dual billiards, *Russ. Math. Surv.* **48**:75–102 (1993).
13. S. Tabachnikov, *Billiards* (SMF "Panoramas et Syntheses," 1995).