Ergod. Th. & Dynam. Sys. (1997), **17**, 957–976
Printed in the United Kingdom © *1997 Cambridge University Press*

Introducing projective billiards

SERGE TABACHNIKOV

Department of Mathematics, UARK, Fayetteville AR 72701, USA
and Wolfson College and I. Newton Institute, University of Cambridge, UK

(*Received* 30 *March* 1995 *and revised* 2 *March* 1996)

Abstract. We introduce and study a new class of dynamical systems, the projective billiards, associated with a smooth closed convex plane curve equipped with a smooth field of transverse directions. Projective billiards include the usual billiards along with the dual, or outer, billiards.

1. *Introduction and main results*

Given a closed smooth convex plane curve γ, the (usual) billiard transformation is a transformation of the set of oriented lines (or rays) intersecting γ which is defined by the law of geometrical optics: the angle of incidence equals that of reflection—see Figure 1.

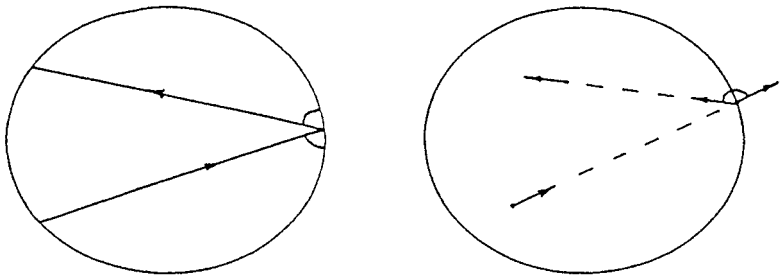


FIGURE 1.

One also defines the billiard flow on the set of tangent vectors to the plane with footpoints inside γ: a vector moves freely along a straight line until it hits the curve; then it instantaneously changes according to the same law of equal angles, and the motion continues—see Figure 1 again. The billiard dynamics is a much studied subject, rich with beautiful results—see, for example, the surveys [**ChGZ**, **T1**].

The billiard transformation is defined in metric terms (equal angles), and, in particular, it is not equivariant under projective transformations of the plane. Here we start the study of a wider class of billiards called *projective billiards*, free from this shortcoming. The

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

way we see the relation between the projective billiards and the usual ones is expressed by the following 'equation':

$$\frac{\textit{Usual Billiards}}{\textit{Projective Billiards}} = \frac{\textit{Euclidean Geometry}}{\textit{Projective Geometry}}.$$

To define projective billiards we need an extra structure on the billiard curve. Let γ be a closed smooth strictly convex plane curve equipped with a smooth field of transverse directions n. Define the projective billiard transformation T of the set of rays intersecting γ. Let an incoming ray r meet γ at point X; then r, the tangent line to γ at X, the outgoing ray $T(r)$ and the transverse line n at X constitute a harmonic quadruple of lines—see Figure 2.

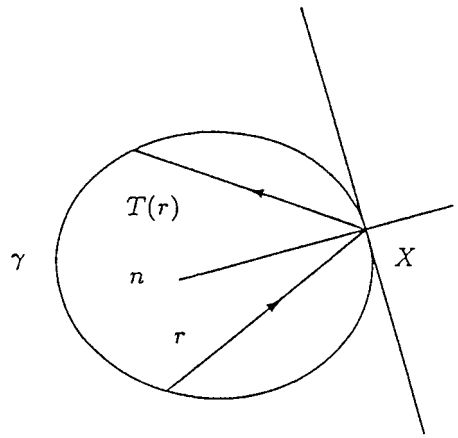


FIGURE 2.

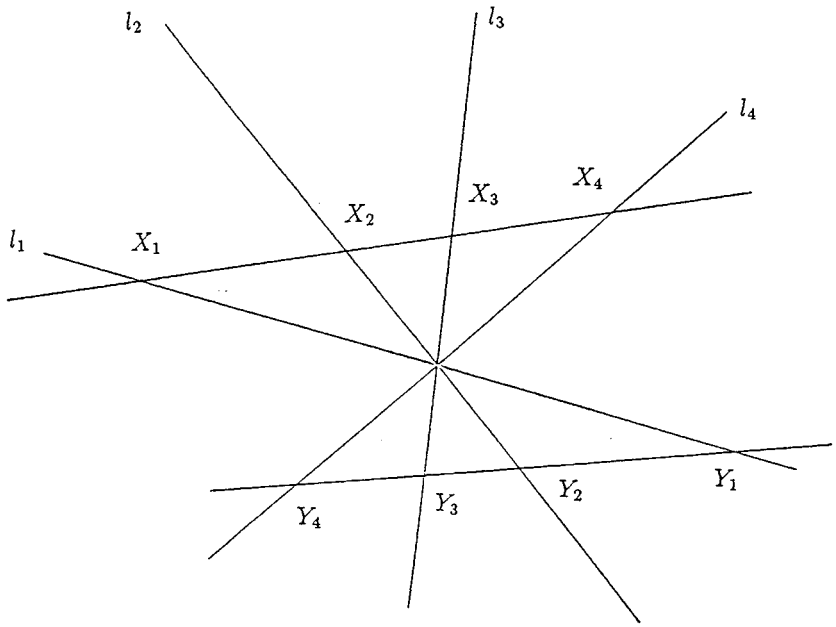
A harmonic quadruple of lines are four lines whose cross-ratio equals -1. The cross-ratio of four lines through a point is the only projective invariant of such lines; it is defined as the cross-ratio of their four intersection points with a fifth line (and does not depend on the choice of this auxiliary line)—see Figure 3. In particular, a ray tangent to γ remains intact, while a ray, whose direction is that of the transversal at its intersection point with γ, changes its direction under the projective billiard transformation.

By the very definition, the projective billiard transformation T is equivariant under projective transformations of the plane. Namely, if A is such a transformation and T' is the projective billiard transformation, associated with the curve $A(\gamma)$ equipped with the field $A(n)$, then $A \circ T = T' \circ A$.

If the transverse field consists of normals to γ then the incoming and outgoing rays make equal angles with γ (to see this, let the auxiliary line be parallel to the tangent line of γ; then the fourth point of intersection with this line goes to infinity, and the second one must bisect the segment made by the first and the third—see Figure 4). Thus the billiard transformation is a particular case of the projective billiard one.

We also define the projective billiard flow on the set of tangent vectors to the plane with footpoints inside γ. Again, a vector u moves freely until it hits γ at point X. There

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press



$$[l_1,\ l_2,\ l_3,\ l_4] = [X_1,\ X_2,\ X_3,\ X_4] = [Y_1,\ Y_2,\ Y_3,\ Y_4]$$

FIGURE 3.

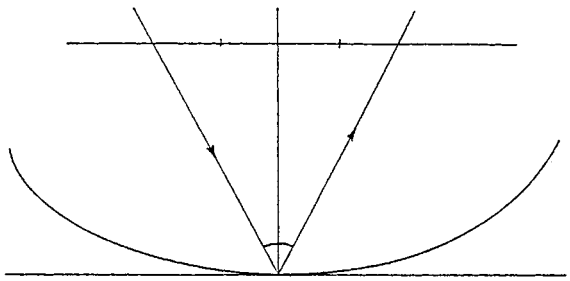


FIGURE 4.

u is decomposed into the tangential and transverse components (using the transversal to γ at X), and u is instantaneously replaced by a new vector v with the same tangential and the opposite transverse components. After this the motion continues as before—see Figure 5.

This definition agrees with the previous one: if r and R are the rays generated by u and v, then $T(r) = R$ (to see this, consider the auxiliary line parallel to the transverse direction in Figure 5). Let L^2 be the set of rays intersecting γ, and let M^3 be the set of inward vectors with footpoints on γ. Then L is diffeomorphic to an annulus, and M has a natural projection π to L: a vector goes to the ray it generates. We thus have defined

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

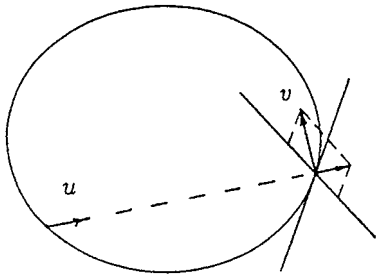


FIGURE 5.

a transformation F of M, namely $F(u) = v$ in Figure 5; and we have $\pi \circ F = T \circ \pi$.

What makes the study of (usual) billiards fruitful and interesting is the existence of an area form in the set of rays L invariant under the billiard transformation. This is the unique, up to a constant factor, form in the set of rays invariant under the motions of the Euclidean plane. It is convenient to introduce coordinates in L: the angle parameter α is the angle made by a ray with a fixed direction, and the real parameter p is the signed distance from an origin to the ray—see Figure 6. Then the invariant form is $dp \wedge d\alpha$.

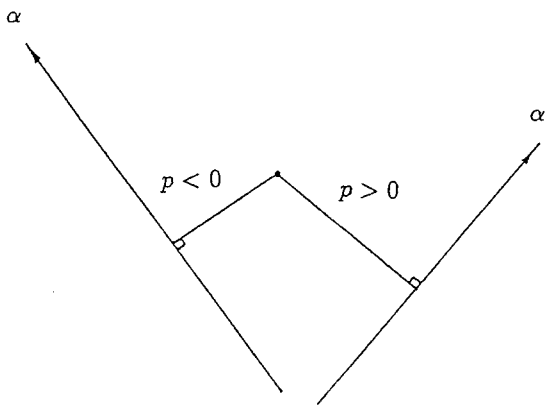


FIGURE 6.

One of the most interesting problems on projective billiards is when the projective billiard transformation $T : L \to L$ has an invariant form. We want this form to exist at least in a vicinity of the boundary, i.e. on the set of rays that make small angles with γ, and to be smooth up to the boundary. Our first result provides a criterion for the existence of such a form.

THEOREM A. *T has an invariant area form if and only if there exists a section $i : L \to M$ of the projection $\pi : M \to L$ which commutes with the projective billiard maps: $F \circ i = i \circ T$. If the section exists then the invariant form is given by the formula $|i(\alpha, p)|^3 \, d\alpha \wedge dp$.*

An obvious example in which such a section exists is that of the usual billiard

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

transformation: i sends an oriented line to its unit tangent vector at its first point of intersection with the curve.

Note that the invariant form for the usual billiard transformation can be obtained from the canonical symplectic form in the cotangent bundle of the billiard table by symplectic reduction, i.e. by first restricting the symplectic form to the hypersurface of unit covectors, and then descending to its quotient space of oriented lines (see, for example, [**A**, **T1**]). It would be very interesting to understand the symplectic meaning of Theorem A.

Call a transverse field along γ *conservative* if there exists a parametrization $\gamma(t)$ such that the transverse direction at every point $\gamma(t)$ is that of the acceleration vector $\gamma''(t)$ (primes denote derivatives with respect to t). The normals to γ is a particular case, the parametrization being by arc length.

The next result gives a necessary condition for the existence of a T-invariant area form.

THEOREM B. *If there exists the infinite jet of an invariant area form for the projective billiard transformation along a boundary component of the phase annulus, then the transverse field is conservative.*

We do not know how far this condition is from being sufficient for the existence of an invariant form at a vicinity of a boundary component of the phase annulus. Our conjecture is that if the transverse field is conservative then an invariant form exists on the formal level, that is, there exists a formal series for it along γ. However this series may diverge.

We also make two observations concerning conservative transversals—see Figure 7.

THEOREM C.

(a) *Let γ be equipped with a conservative transverse field. Then, given two points A and B inside γ, there exist at least two rays through A which, after the projective billiard reflection in γ, pass through B.*

(b) *Let A and B be two points inside γ (but not on it). Then there exists a conservative transverse field along γ such that γ enjoys the optical property of an ellipse with foci A and B: every ray through A projectively reflects to a ray through B.*

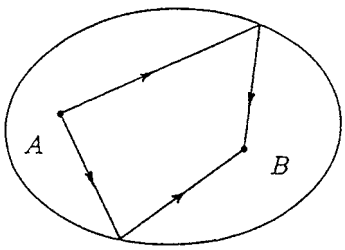


FIGURE 7.

As another indication that conservative transverse fields are a reasonable object of study, we mention a generalization of the classical 4-vertex theorem found in [**T2**]: if the lines of a conservative transverse field along a convex curve γ always revolve in the

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

same sense then the envelope of these lines has at least four cusp singularities (the usual 4-vertex theorem being the case of the normals). It is also worth mentioning that the property of a transverse field to be conservative is projectively-invariant: a projective transformation A takes a conservative field n along γ to the conservative field $A(n)$ along $A(\gamma)$—see [**T6**].

Next we study a particular case of projective billiards, the *centrally-projective* ones. In centrally-projective billiards the transverse field consists of lines through a point inside the billiard curve γ; this point is chosen as the origin—see Figure 8. Such a field is conservative.

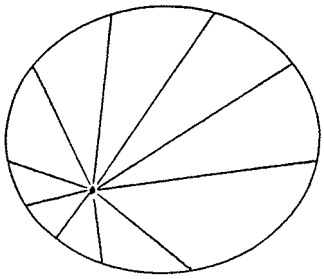


FIGURE 8.

THEOREM D. *The centrally-projective billiard transformation has an invariant area form* $p^{-3}\,d\alpha \wedge dp$ *that blows up at the set of rays through the origin.*

We find a generating function for the centrally-projective billiard transformation in Theorem 5.1 and discuss some unusual properties of the geometrical optics in the 'centrally-projective world' in Theorem 5.2. We also show that this transformation is conjugate to the dual (or outer) billiard transformation by polar duality (see [**B**, **GK**, **T1**, **T3**, **T4**] concerning dual billiards).

A particular class of projective billiards which possess invariant area forms was suggested to the author by E. Gutkin. Consider a metric in a plane domain, that contains the curve γ, whose non-parametrized geodesics are straight lines (such a metric is called projectively equivalent to the Euclidean one). Then the (usual) billiard flow in this metric, i.e. the geodesic flow inside γ, induces the billiard transformation of the set of rays, intersecting γ. This transformation has an invariant form that comes from the symplectic form in the cotangent bundle of the billiard table. The transformation in question is a projective billiard transformation, the transversals being the normals to γ in the metric under consideration.

Here are two examples of this construction. Let S be the upper unit hemisphere in Euclidean 3-space with Cartesian coordinates x, y, z, and let H be the upper sheet of the hyperboloid $z^2 - x^2 - y^2 = 1$. Consider S in the metric induced by the Euclidean one and H in the one induced by $dx^2 + dy^2 - dz^2$. Then S and H become surfaces of constant curvature, positive in the former and negative in the latter cases. The geodesics in S and H are the curves of intersection with the planes through the origin.

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

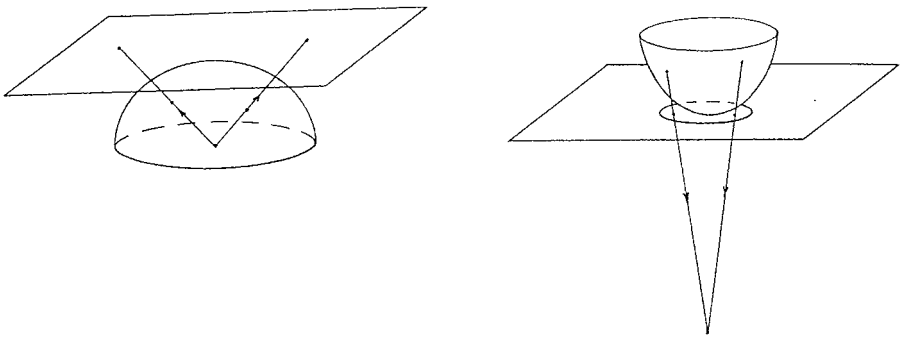


FIGURE 9.

Consider the plane $z = 1$ and project S or H on it from the origin—see Figure 9. This projection takes geodesics to straight lines, so it defines the metrics of constant curvature in the plane projectively equivalent to the Euclidean one (the hyperboloid is projected into the unit circle). One may further apply projective transformations of the plane to these metrics to keep the geodesics straight, but that is as far as one can get. According to a Beltrami theorem a metric, projectively equivalent to a metric of constant curvature, has itself constant curvature (see [**D**]). If another mapping from S or H is given to the plane that takes geodesics to straight lines then the composition of this mapping and the inverse of the above projection is a mapping of the plane which takes lines to lines. Such a mapping is a projective transformation.

Next we turn our attention to projective billiards in a circle. Let γ be a circle equipped with an arbitrary transverse field. We choose the origin at the center of the circle.

THEOREM E. *The form* $(1 - p^2)^{-3/2}\, d\alpha \wedge dp$ *is invariant under every projective billiard transformation* T *in the unit circle.*

This does not contradict Theorem B because this form blows up at the boundary $p = \pm 1$ of the phase annulus.

Suppose however that T has another invariant form, smooth up to the boundary. Then the quotient of the two is a T-invariant function whose level curves foliate a neighbourhood of the boundary of the phase annulus.

COROLLARY F. *If a projective billiard transformation* T *in a circle has an invariant area form, smooth up to the boundary of the phase annulus, then* T *is integrable.*

It is a challenging problem to describe all such cases.

An example of this situation is given by the usual billiard in an ellipse; the billiard transformation has an invariant area form. Apply a projective transformation that takes the ellipse to a circle to obtain a projective billiard therein with an invariant measure smooth up to the boundary of the phase annulus. Likewise, one considers an ellipse in the plane with the spherical or the hyperbolic metric projectively equivalent to the Euclidean one (the corresponding curves in S and H are their intersections with quadratic cones). We obtain a new proof of a known result (see [**V**]).

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

COROLLARY G. *The billiard transformation in an ellipse in the Euclidean, spherical and hyperbolic plane is integrable.*

A consequence of this result is the classical Poncelet theorem of projective geometry—see [**T5**].

Let us mention that projective billiards can be defined in the multi-dimensional setting as well: a closed convex hypersurface is equipped with a field of transverse directions; an incoming ray, the transversal and the outgoing ray lie in the same 2-plane and constitute a harmonic quadruple of lines therein with the tangent line to the hypersurface. We finish our discussion with a problem: are there multi-dimensional analogues of Theorem E and its corollaries†?

The contents of this paper are as follows. §2 provides some formulas to be used later. §3 contains the proofs of Theorems E and A, §4 those of Theorems C and B. The last section is a discussion of centrally-projective billiards, and Theorem D is proved there.

2. *A few formulas*

Consider the lines of a harmonic quadruple, and choose vectors along these lines—see Figure 10. Denote by $[a, b]$ the skew-product of two vectors, i.e. the determinant whose columns are the vectors.

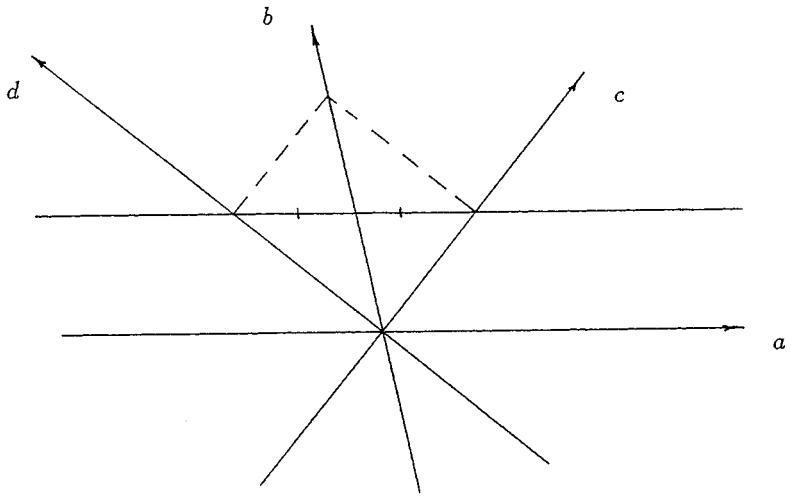


FIGURE 10.

LEMMA 2.1. $[a, c][b, d] + [a, d][b, c] = 0$.

Proof. The expression in question is linear in each vector, so we may choose their lengths arbitrarily. Draw a line parallel to a; its intersection point with the b-line bisects the segment made by the intersection points with the c- and d-lines. Thus the lines

† Added in proof: see the author's paper 'Exact transverse line fields and projective billiards in a ball', to appear in *GAFA*.

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

through c and d are the sides of a parallelogram whose diagonals are parallel to a and b. Therefore we may set: $a = c - d$, $b = c + d$. Substitute to the above expression and use skew-symmetricity of the skew-product to conclude. □

Let α, β and ϕ be the angles made by the lines of a harmonic quadruple—see Figure 11.

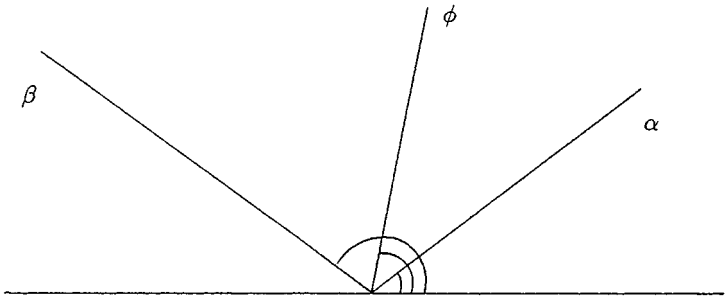


FIGURE 11.

LEMMA 2.2. $\cot\alpha + \cot\beta = 2\cot\phi$.

Proof. Set in the previous lemma:

$$a = (1, 0), \quad b = (\cot\phi, 1), \quad c = (\cot\alpha, 1), \quad d = (\cot\beta, 1). \qquad \square$$

For later use we also formulate a property of the skew-product.

LEMMA 2.3. *For every four vectors a, b, c, d one has*

$$[a, b][c, d] = [a, c][b, d] + [a, d][c, b].$$

Consider now a parametrized projective billiard curve γ; we write $\gamma(t)$ for the position vectors of its points and $\gamma'(t)$ for the tangent velocity vector at $\gamma(t)$. Let the transverse directions at $\gamma(t)$ be given by a smooth vector field $n(t)$. A ray intersecting γ is parametrized by (t_1, t_2) where $\gamma(t_1)$ and $\gamma(t_2)$ are its first and second intersection points with γ. Thus (t_1, t_2) are coordinates in L. Consider an instance of the projective billiard reflection—see Figure 12. The projective billiard transformation acts as follows: $T(t_1, t_2) = (t_2, t_3)$.

According to Lemma 2.1

$$\frac{[\gamma(t_2) - \gamma(t_1), n(t_2)]}{[\gamma(t_2) - \gamma(t_1), \gamma'(t_2)]} = -\frac{[\gamma(t_3) - \gamma(t_2), n(t_2)]}{[\gamma(t_3) - \gamma(t_2), \gamma'(t_2)]}.$$

Take the exterior differential and wedge-multiply by dt_2 to obtain

$$\frac{[\gamma(t_2) - \gamma(t_1), n(t_2)][\gamma'(t_1), \gamma'(t_2)] - [\gamma'(t_1), n(t_2)][\gamma(t_2) - \gamma(t_1), \gamma'(t_2)]}{[\gamma(t_2) - \gamma(t_1), \gamma'(t_2)]^2} dt_1 \wedge dt_2$$
$$= \frac{[\gamma(t_3) - \gamma(t_2), \gamma'(t_2)][\gamma'(t_3), n(t_2)] - [\gamma'(t_3), \gamma'(t_2)][\gamma(t_3) - \gamma(t_2), n(t_2)]}{[\gamma(t_3) - \gamma(t_2), \gamma'(t_2)]^2} dt_2 \wedge dt_3.$$

Simplify the numerators using Lemma 2.3 and cancel the common nonzero factor $[\gamma'(t_2), n(t_2)]$ to arrive at the following.

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

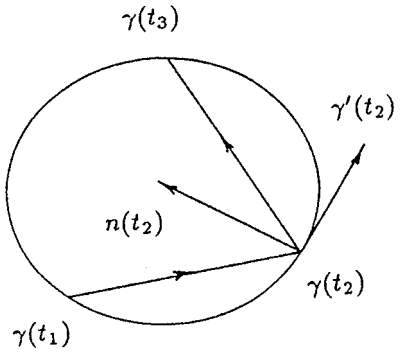


FIGURE 12.

LEMMA 2.4.

$$\frac{[\gamma'(t_1), \gamma(t_2) - \gamma(t_1)]}{[\gamma(t_2) - \gamma(t_1), \gamma'(t_2)]^2}\, dt_1 \wedge dt_2 = \frac{[\gamma(t_3) - \gamma(t_2), \gamma'(t_3)]}{[\gamma(t_3) - \gamma(t_2), \gamma'(t_2)]^2}\, dt_2 \wedge dt_3.$$

3. *An invariant area form*

Denote the 2-form from Lemma 2.4 by ω. Since $T(t_1, t_2) = (t_2, t_3)$ we have

$$T^*\omega = \frac{[\gamma'(t_2), \gamma(t_3) - \gamma(t_2)]}{[\gamma(t_3) - \gamma(t_2), \gamma'(t_3)]^2}\, dt_2 \wedge dt_3.$$

Hence

$$\frac{T^*\omega}{\omega} = \frac{[\gamma'(t_2), \gamma(t_3) - \gamma(t_2)]^3}{[\gamma(t_3) - \gamma(t_2), \gamma'(t_3)]^3}.$$

We look for a T-invariant form $g(t_1, t_2)\omega$ where g is a function. Thus we have the following.

LEMMA 3.1. *A necessary and sufficient condition for T-invariance of the form $g\omega$ is that the function g satisfies the equality*

$$\frac{g \circ T}{g} = \frac{[\gamma(t_3) - \gamma(t_2), \gamma'(t_3)]^3}{[\gamma'(t_2), \gamma(t_3) - \gamma(t_2)]^3}.$$

Consider, as an illustration, the case of the usual billiard. Parametrize γ by arc length. Then the right-hand side expression in Lemma 3.1 equals $\sin^3\alpha_3/\sin^3\alpha_2$—see Figure 13. Thus one may take $g(t_1, t_2) = \sin^3\alpha_2$. In view of Lemma 3.1 we obtain an invariant form

$$|\gamma(t_2) - \gamma(t_1)|^{-1} \sin\alpha_1 \sin\alpha_2\, dt_1 \wedge dt_2,$$

which can be rewritten as $\sin\alpha_1\, dt_1 \wedge d\alpha_1$. This is the well-known invariant form of the billiard transformation.

Consider next the case when γ is a unit circle. Parametrize it by the standard angle parameter t. Then the vector $\gamma(t_3) - \gamma(t_2)$ makes equal angles with the vectors $\gamma'(t_2)$

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

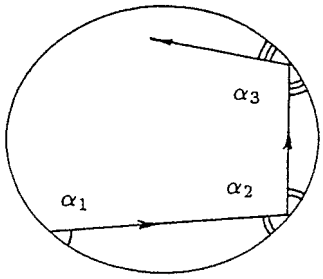


FIGURE 13.

and $\gamma'(t_3)$, and the right-hand side of the formula in Lemma 3.1 is equal to 1. Hence one may take $g = 1$. This gives an invariant form

$$\frac{[\gamma'(t_1), \gamma(t_2) - \gamma(t_1)]}{[\gamma(t_2) - \gamma(t_1), \gamma'(t_2)]^2}\, dt_1 \wedge dt_2 = [\gamma(t_2) - \gamma(t_1), \gamma'(t_2)]^{-1}\, dt_1 \wedge dt_2.$$

Since $\gamma(t) = (\cos t, \sin t)$ this form is equal to

$$(1 - \cos(t_2 - t_1))^{-1}\, dt_1 \wedge dt_2.$$

To obtain Theorem E one rewrites this form in coordinates α, p related to t_1, t_2 by the formulas

$$p = \cos((t_2 - t_1)/2), \quad \alpha = (t_1 + t_2)/2 + \pi/2.$$

See Figure 14.

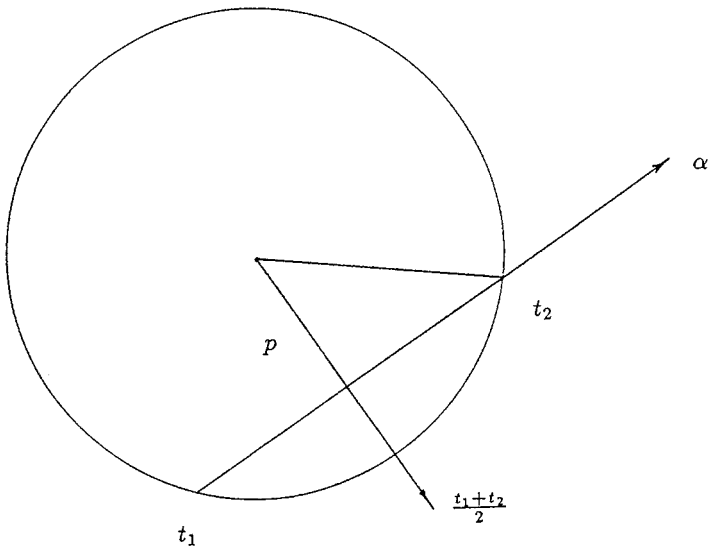


FIGURE 14.

Now we prove Theorem A. Assume first that there exists a section $i : L \to M$ such that $F \circ i = i \circ T$. Then $i(t_1, t_2)$ is a vector proportional to $\gamma(t_2) - \gamma(t_1)$. Let

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

$i(t_1, t_2) = h(t_1, t_2)(\gamma(t_2) - \gamma(t_1))$. Since $F \circ i = i \circ T$ the vector

$$h(t_1, t_2)(\gamma(t_2) - \gamma(t_1)) + h(t_2, t_3)(\gamma(t_3) - \gamma(t_2))$$

is proportional to $\gamma'(t_2)$, i.e.

$$h(t_1, t_2)[\gamma(t_2) - \gamma(t_1), \gamma'(t_2)] + h(t_2, t_3)[\gamma(t_3) - \gamma(t_2), \gamma'(t_2)] = 0.$$

Thus

$$[\gamma'(t_2), \gamma(t_3) - \gamma(t_2)] = \frac{h(t_1, t_2)}{h(t_2, t_3)}[\gamma(t_2) - \gamma(t_1), \gamma'(t_2)].$$

It follows that the right-hand side of the expression in Lemma 3.1 equals

$$\left(\frac{h(t_2, t_3)[\gamma(t_3) - \gamma(t_2), \gamma'(t_3)]}{h(t_1, t_2)[\gamma(t_2) - \gamma(t_1), \gamma'(t_2)]}\right)^3.$$

One may, therefore, set

$$g(t_1, t_2) = (h(t_1, t_2)[\gamma(t_2) - \gamma(t_1), \gamma'(t_2)])^3.$$

This yields an invariant form

$$g\omega = h^3(t_1, t_2)[\gamma'(t_1), \gamma(t_2) - \gamma(t_1)][\gamma(t_2) - \gamma(t_1), \gamma'(t_2)]\, dt_1 \wedge dt_2.$$

It is a routine but rather tedious exercise to rewrite this form in coordinates α, p, and we omit it. The result, as stated in Theorem A, is

$$|i(\alpha, p)|^3\, d\alpha \wedge dp,$$

where (α, p) are the coordinates of the ray from $\gamma(t_1)$ to $\gamma(t_2)$.

Conversely, if an invariant form exists one has a function $g(t_1, t_2)$ satisfying the equality from Lemma 3.1. Define the section $i : L \to M$ by the formula

$$i(t_1, t_2) = \frac{g(t_1, t_2)^{1/3}}{[\gamma(t_2) - \gamma(t_1), \gamma'(t_2)]}(\gamma(t_2) - \gamma(t_1)).$$

The equation from Lemma 3.1 implies then that $F \circ i = i \circ T$.

4. *Conservative transverse fields*

We start with a criterion for a smooth transverse field to be conservative. Let $\gamma(t)$ be a parametrized curve and let $n(t)$ be a transverse vector field.

LEMMA 4.1. *The field of directions generated by $n(t)$ is conservative if and only if*

$$\int \frac{[\gamma''(t), n(t)]}{[\gamma'(t), n(t)]}\, dt = 0$$

(integral along the curve γ).

Proof. Let τ be another parametrization of γ. Then

$$\gamma_t = \gamma_\tau \tau_t, \quad \gamma_{tt} = \gamma_{\tau\tau}\tau_t^2 + \gamma_\tau \tau_{tt},$$

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

where the lower index denotes the derivative with respect to the corresponding variable.

If n is conservative then it is proportional to $\gamma_{\tau\tau}$ for some parameter τ. Thus

$$\frac{[\gamma_{tt}, n]}{[\gamma_t, n]} = \frac{[\gamma_{\tau\tau}\tau_t^2 + \gamma_\tau \tau_{tt}, \gamma_{\tau\tau}]}{[\gamma_\tau \tau_t, \gamma_{\tau\tau}]} = \frac{\tau_{tt}}{\tau_t} = (\log \tau_t)_t,$$

and the integral of a derivative vanishes.

Conversely, if the integral vanishes then

$$-\frac{[\gamma_{tt}, n]}{[\gamma_t, n]}\, dt = df$$

for some function f. Define a new parameter τ by setting $t_\tau = \exp f$. Then

$$df = \frac{t_{\tau\tau}}{t_\tau}\, d\tau = \frac{t_{\tau\tau}}{t_\tau^2}\, dt,$$

hence

$$\frac{t_{\tau\tau}}{t_\tau^2} = -\frac{[\gamma_{tt}, n]}{[\gamma_t, n]}.$$

It follows that $[\gamma_{tt}t_\tau^2 + \gamma_t t_{\tau\tau}, n] = 0$, or $[\gamma_{\tau\tau}, n] = 0$. Thus n is proportional to $\gamma_{\tau\tau}$, and therefore, conservative. □

Next we prove Theorem C(b). Let $\gamma(t)$ be a parametrized curve. If every ray through A reflects to a ray through B then the transverse field $n(t)$ is uniquely determined, up to a factor, by the condition from Lemma 2.1:

$$[\gamma(t) - A, \gamma'(t)][\gamma(t) - B, n(t)] + [\gamma(t) - B, \gamma'(t)][\gamma(t) - A, n(t)] = 0.$$

If we decompose

$$n(t) = u(t)(\gamma(t) - A) + v(t)(\gamma(t) - B)$$

then

$$u(t)[\gamma(t) - B, \gamma'(t)][\gamma(t) - B, \gamma(t) - A] + v(t)[\gamma(t) - B, \gamma'(t)][\gamma(t) - A, \gamma(t) - B] = 0.$$

Thus one may take

$$u(t) = [\gamma(t) - B, \gamma'(t)], \quad v(t) = [\gamma(t) - A, \gamma'(t)].$$

To apply Lemma 4.1 one computes

$$[n(t), \gamma''(t)] = [\gamma(t) - B, \gamma'(t)][\gamma(t) - A, \gamma''(t)] + [\gamma(t) - B, \gamma''(t)][\gamma(t) - A, \gamma'(t)],$$
$$[n(t), \gamma'(t)] = 2[\gamma(t) - B, \gamma'(t)][\gamma(t) - A, \gamma'(t)].$$

Hence

$$[n(t), \gamma''(t)] = [n(t), \gamma'(t)]'/2,$$

and the integrand is a logarithmic derivative. Therefore $n(t)$ is a conservative field.

The parameter τ for which n is proportional to $\gamma_{\tau\tau}$ is given by the condition

$$[\gamma - A, \gamma_\tau][\gamma - B, \gamma_\tau] = 1$$

identically in τ.

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

Now we prove Theorem C(a). Let $\gamma(t)$ be a parametrized curve with the transverse field $\gamma''(t)$. Then a ray from A reflects at point $\gamma(t)$ to pass through B if and only if

$$[\gamma(t) - A, \gamma'(t)][\gamma(t) - B, \gamma''(t)] + [\gamma(t) - B, \gamma'(t)][\gamma(t) - A, \gamma''(t)] = 0,$$

or

$$([\gamma(t) - A, \gamma'(t)][\gamma(t) - B, \gamma'(t)])' = 0.$$

Since a function has at least two critical points on a circle the result follows.

Finally we turn our attention to Theorem B. Suppose that there exists a formal invariant form along a boundary component of the phase annulus. According to Theorem A there exists a section $i : L \to M$, where L is the formal neighbourhood of the boundary component of the phase annulus, such that $F \circ i = i \circ T$. The mapping i assigns a vector to a line.

In particular, a vector is assigned to every positive tangent line to γ. Consider the parametrization $\gamma(t)$ such that these vectors are equal to the velocity vectors $\gamma'(t)$ at every point of the curve. This parametrization is canonically, up to an affine reparametrization, determined by the invariant form. We will show that this parametrization is the one in which the transversals are generated by the acceleration vectors $\gamma''(t)$.

We use the coordinates t, ϵ in L, related to t_1, t_2 by the formulae: $t = t_1$, $\epsilon = t_2 - t_1$ (ϵ is thought of as a small non-negative number). The section i is then a vector-function $i(t, \epsilon)$ satisfying $i(t, 0) = \gamma'(t)$.

Consider the Taylor expansion of i in powers of ϵ:

$$i(t, \epsilon) = \gamma'(t) + \epsilon u(t) + O(\epsilon^2).$$

Since $i(t, \epsilon)$ is collinear with $\gamma(t + \epsilon) - \gamma(t)$ one has

$$[i(t, \epsilon), \gamma(t + \epsilon) - \gamma(t)] = 0,$$

hence

$$[u(t) - \gamma''(t)/2, \gamma'(t)] = 0.$$

Therefore

$$u(t) = \gamma''(t)/2 + f(t)\gamma'(t),$$

where f is a function, and

$$i(t, \epsilon) = \gamma'(t) + \epsilon(\gamma''(t)/2 + f(t)\gamma'(t)) + O(\epsilon^2).$$

Let $T(t - \epsilon, \epsilon) = (t, \delta)$—see Figure 15. By the law of projective billiard reflection the vector $i(t - \epsilon, \epsilon) + i(t, \delta)$ is collinear with $\gamma'(t)$. Let

$$\delta(t, \epsilon) = \epsilon a(t) + O(\epsilon^2).$$

Then

$$i(t - \epsilon, \epsilon) + i(t, \delta) = 2\gamma'(t) + \epsilon((a(t) + 1)(\gamma''(t)/2 + f(t)\gamma'(t))\gamma''(t)) + O(\epsilon^2).$$

This vector is collinear with $\gamma'(t)$, thus

$$(a(t) - 1)[\gamma''(t), \gamma'(t)] = 0,$$

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

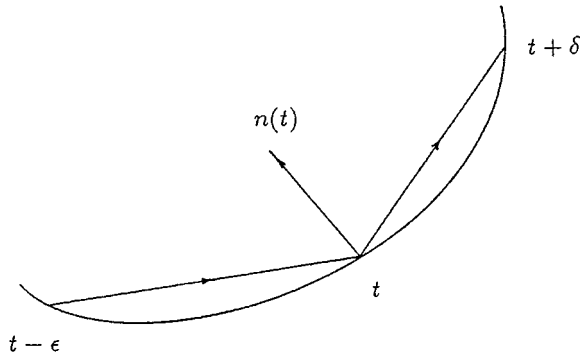


FIGURE 15.

and therefore $a(t) = 1$ identically. Hence

$$\delta(t, \epsilon) = \epsilon + O(\epsilon^2).$$

Finally one finds the transverse directions $n(t)$. By the law of reflection, $n(t)$ is collinear with $i(t, \delta) - i(t - \epsilon, \epsilon)$. Using the above computation one obtains

$$i(t, \delta) - i(t - \epsilon, \epsilon) = \epsilon\gamma''(t) + O(\epsilon^3).$$

Hence $n(t)$ is collinear to $\gamma''(t)$, and the transverse field is conservative.

To conclude this section we indicate another approach to the proof of Theorem B. If an invariant form exists, one may apply the theory of interpolating Hamiltonians to the map T—see, for example, [**MM**]. According to this theory there exists a smooth function $f(t, \epsilon)$ in a neighbourhood of the boundary component of the phase annulus which is T-invariant to infinite order in ϵ, and which vanishes on the boundary. Expanding the equation $f(t - \epsilon, \epsilon) = f(t, \delta)$ in powers of ϵ one arrives at the same conclusions as before. We do not believe, however, that this method will yield any further necessary conditions for the existence of an invariant form along the boundary of the phase annulus.

5. *Centrally-projective billiards: outer billiards seen from within*

To start with, we prove Theorem D. Consider the section $i : L \to M$ which assigns the vector of length p^{-1} to a line at distance p from the origin. Let l be an oriented line and C a point on it. Then $|[C, i(l)]| = 1$.

Consider now a centrally-projective reflection of a ray l to the ray m—see Figure 16. We have: $[\gamma(t), i(l)] = 1$ and $[\gamma(t), i(m)] = 1$. Hence $i(m) - i(l)$ is collinear with $\gamma(t)$. The law of reflection reads

$$[i(l), \gamma(t)][i(m), \gamma'(t)] + [i(l), \gamma'(t)][i(m), \gamma(t)] = 0,$$

and it follows that

$$[i(l), \gamma'(t)] = -[i(m), \gamma'(t)].$$

Hence $i(l) + i(m)$ is collinear with $\gamma'(t)$. One concludes that the vector $i(l)$ reflects to the vector $i(m)$ according to the law of centrally-projective reflection, that is, $i \circ T = F \circ i$.

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

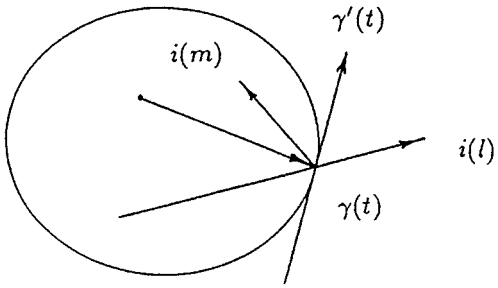


FIGURE 16.

By Theorem A the map T has an invariant form

$$\omega = p^{-3} d\alpha \wedge dp.$$

Parametrize γ so that $[\gamma(t), \gamma'(t)] = 1$ for all t. Then $[\gamma(t), \gamma''(t)] = 0$, so $\gamma''(t)$ is collinear with $\gamma(t)$.

Since $[\gamma(t_1), i(t_1, t_2)] = 1$ and $[\gamma(t_1), \gamma'(t_1)] = 1$ the vector $i(t_1, t_2) - \gamma'(t_1)$ is proportional to $\gamma(t_1)$. Introduce a new variable q by setting:

$$i(t_1, t_2) - \gamma'(t_1) = -q(t_1, t_2)\gamma(t_1).$$

We use $t = t_1$ and $q = q(t_1, t_2)$ as new coordinates in L.

Since $i(t_1, t_2)$ is collinear with $\gamma(t_2) - \gamma(t_1)$ one has

$$[\gamma(t_2) - \gamma(t_1), \gamma'(t_1) - q\gamma(t_1)] = 0.$$

Hence

$$q = \frac{1 + [\gamma'(t_1), \gamma(t_2)]}{[\gamma(t_1), \gamma(t_2)]}.$$

A direct computation, which we omit, makes it possible to rewrite the invariant form in coordinates t, q:

$$\omega = q\, dq \wedge dt = d\tfrac{1}{2}q^2\, dt.$$

Let $T(t_1, q_1) = (t_2, q_2)$. Since $T^*\omega = \omega$ the one-form $q_2^2\, dt_2 - q_1^2\, dt_1$ is closed. Our next goal is to find a *generating function* $H(t_1, t_2)$ such that

$$q_2^2\, dt_2 - q_1^2\, dt_1 = dH(t_1, t_2).$$

We know that $\gamma''(t)$ is proportional to $\gamma(t)$ for all t. Introduce a function $k(t)$ by $\gamma''(t) = k(t)\gamma(t)$, and let $K(t)$ be its inverse derivative.

THEOREM 5.1. *The generating function is given by the formula*

$$H(t_1, t_2) = \frac{[\gamma(t_2) - \gamma(t_1), \gamma'(t_2) - \gamma'(t_1)]}{[\gamma(t_1), \gamma(t_2)]} + K(t_2) - K(t_1).$$

Proof. It is convenient to introduce a function

$$F(t_1, t_2) = [\gamma(t_1), \gamma(t_2)].$$

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

Denote by F_1 and F_2 the partial derivatives with respect to the first and the second variables. Since $[\gamma(t), \gamma'(t)] = 1$, one concludes from Lemma 2.3 that $F_{12}F - F_1F_2 = 1$. Also $F_{11} = k(t_1)F$ and $F_{22} = k(t_2)F$.

We know that $q_1 = (1 + F_1)/F$. To find q_2 use the fact that, according to the law of reflection, the vector

$$\gamma'(t_1) - q_1\gamma(t_1) + \gamma'(t_2) - q_2\gamma(t_2)$$

is collinear with $\gamma'(t_2)$. One finds

$$q_2 = (1 - F_2)/F.$$

One needs to check that $H_1 = -q_1^2$ and $H_2 = q_2^2$. We check the first identity, the second one being analogous. One has

$$H = (F_1 - F_2 + 2)/F + K(t_2) - K(t_1).$$

Therefore

$$\begin{aligned} H_1 &= \frac{F_{11} - F_{12}}{F} - \frac{(F_1 - F_2 + 2)F_1}{F^2} - k(t_1) \\ &= k(t_1) - \frac{F_{12}F - F_1F_2 + F_1^2 + 2F_1}{F^2} - k(t_1) = -\frac{(1 + F_1)^2}{F^2} = -q_1^2. \end{aligned}$$ □

Note that $H(t_1, t_2) \to 0$ as $t_1 \to t_2$.

Now we discuss the relation between the centrally-projective billiards and the dual, or outer, billiards. Given a strictly convex closed smooth oriented curve γ, the dual billiard map T is a transformation of its exterior defined as follows. Let x be a point outside of γ; draw the tangent line to γ from x, and let $T(x)$ be the reflection of x in the tangency point y—see Figure 17.

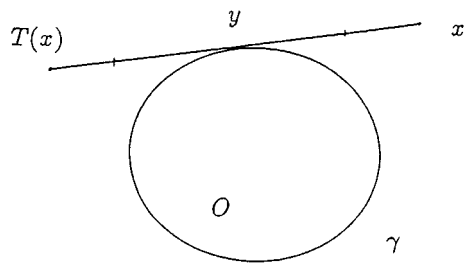


FIGURE 17.

Choose an origin O inside γ. The polar duality is a correspondence between points distinct from the origin and lines not through the origin: to a line l with coordinates (α, p), $p > 0$, there corresponds the point l^* whose polar coordinates are $(\alpha - \pi/2, 1/p)$—see Figure 18. Polar duality preserves the incidence relation: if a point x lies on a line l then the dual line x^* contains the point l^*. Given a smooth curve γ, that is, a one-parameter family of points $\gamma(t)$, the dual curve γ^* is the envelope of the lines $\gamma^*(t)$. This relation is indeed a duality: $\gamma^{**} = \gamma$.

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

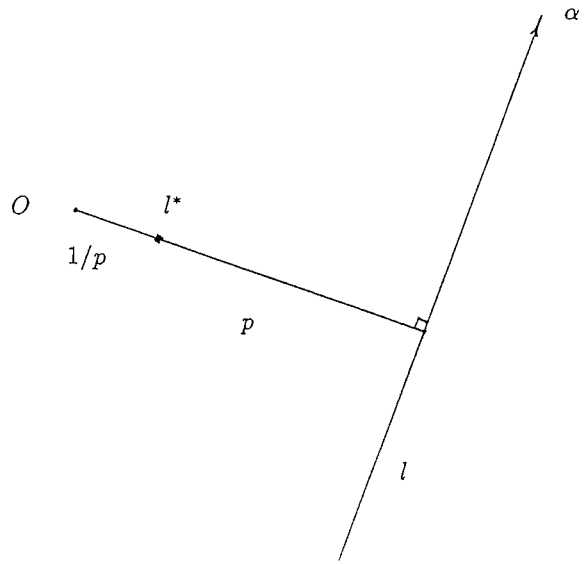


FIGURE 18.

Extend the plane to the projective one, i.e. add the line l at infinity to it. Consider Figure 17 again. Denote by ∞ the intersection point of the line xy with the line l. Then the points $x, y, T(x)$ and ∞ constitute a harmonic quadruple. Apply the polar duality to this configuration. The points $x, y, T(x)$ and ∞ are transformed to lines, and these lines constitute a harmonic quadruple—see Figure 19. One recognizes the centrally-projective billiard transformation.

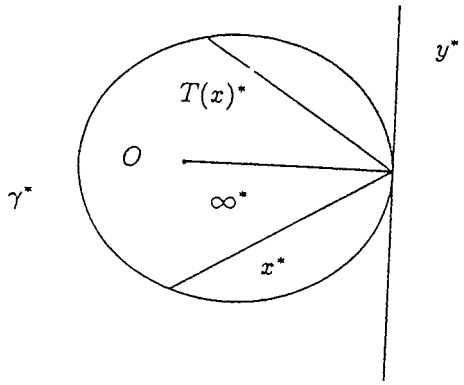


FIGURE 19.

The dual billiard transformation preserves the standard area form in the plane given in polar coordinates by the formula $r\,dr \wedge d\alpha$. Since $r = 1/p$ one concludes that the centrally-projective billiard transformation preserves the form $p^{-3}\,d\alpha \wedge dp$—a fact the reader is well familiar with by now.

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

We conclude with an observation on a strange feature of 'geometrical optics' in the perverse 'centrally-projective world'.

Given a point-like source x consider the propagation of light along rays from x with the velocity equal to the inverse signed distance from the line to the origin O (this unusual law of propagation agrees with the discussion at the beginning of this section). Denote by $\Gamma_t(x)$ the time-t front, that is, the set of points reached by light at time t. By elementary geometry, $\Gamma_t(x)$ is the line parallel to Ox at distance $t/|Ox|$ from the origin—see Figure 20. Therefore $\Gamma_t(x)$ is obtained from x by the polar duality composed with the $\pi/2$ rotation about the origin and the dilation with the coefficient t.

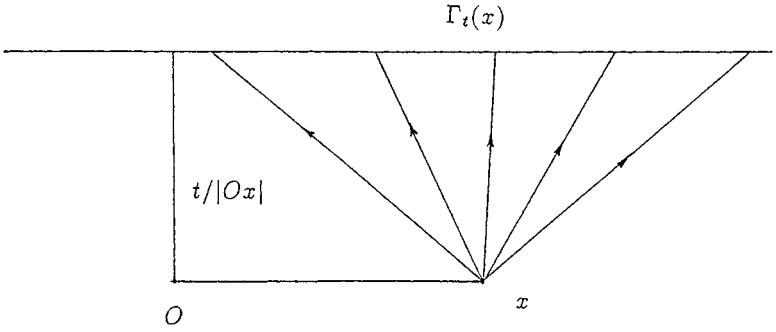


FIGURE 20.

More generally, given a curve γ, define its time-t front $\Gamma_t(\gamma)$ as the envelope of the t-fronts of its points. $\Gamma_t(\gamma)$ is thus obtained from the polar dual curve γ^* by the $\pi/2$ rotation and the t-dilation. Therefore the fronts satisfy the following evolution property.

THEOREM 5.2. *$\Gamma_s(\Gamma_t(\gamma))$ is the curve, homothetic to γ with respect to the origin with the negative coefficient $-s/t$.*

In particular, given a point x, all the lines $\Gamma_s(y)$ for $y \in \Gamma_t(x)$ pass through one point. The centrally-projective geometrical optics is indeed worlds apart from the usual one!

Acknowledgements. It is a pleasure to acknowledge the hospitality of the Wolfson College and the I. Newton Institute at the University of Cambridge. The author is grateful to E. Gutkin for stimulating discussions. This work was supported in part by an NSF Grant DMS-9402732.

REFERENCES

[A] V. Arnold. *Mathematical Methods of Classical Mechanics*. Springer, 1978.

[B] P. Boyland. Dual billiards, twist maps and impact oscillators. *Nonlinearity* **9** (1996), 1411–1438.

[ChGZ] N. Chernov, G. Galperin and A. Zemlyakov. *Mathematics of Billiards*. Cambridge University Press, to appear.

[D] G. Darboux. *Lecon sur le Theorie Generale des Surface*. Chelsea, NY, 1972.

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press

[GK] E. Gutkin and A. Katok. Caustics for inner and outer billiards. *Comm. Math. Phys.* **173** (1995), 101–134.

[MM] S. Marvizi and R. Melrose. Spectral invariants of convex planar regions. *J. Diff. Geom.* **17** (1982), 475–502.

[T1] S. Tabachnikov. Billiards. *SMF 'Panoramas et Syntheses'*, N1, 1995.

[T2] S. Tabachnikov. The four-vertex theorem revisited—two variations on the old theme. *Amer. Math. Monthly* **102** (1995), 912–916.

[T3] S. Tabachnikov. Dual billiards. *Russ. Math. Surv.* **48** (1993), 75–102.

[T4] S. Tabachnikov. On the dual billiard problem. *Adv. in Math.* **115** (2) (1995), 221–249.

[T5] S. Tabachnikov. Poncelet's theorem and dual billiards. *L'Enseign. Math.* **39** (1993), 189–194.

[T6] S. Tabachnikov. Parametrized curves, Minkowski caustics, Minkowski vertices and conservative line fields. *L'Enseign. Math.* to appear.

[V] A. Veselov. Confocal surfaces and integrable billiards on the sphere and in the Lobachevsky space. *J. Geom. Phys.* **7** (1990), 81–107.

https://doi.org/10.1017/S0143385797086239 Published online by Cambridge University Press