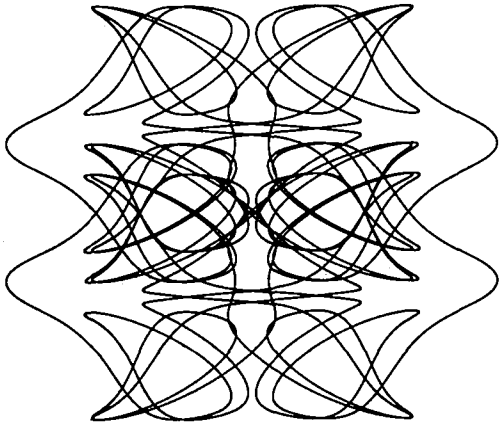
The Dynamics of Quadrilateral Folding

Kevin Charter and Thomas Rogers

CONTENTS



Research supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

When a quadrilateral undergoes a certain infinite folding process, an intricate curve is traced out by its transformed vertices. The curve possesses attractive symmetry properties. We study the folding as a dynamical system in $\mathbf{C}^4$, which is partially described by a billiard-like dynamical system on the bounded component of an elliptic curve associated with the squared diagonal lengths.

INTRODUCTION

Consider a plane polygon. Fold it along one of its diagonals to produce a new polygon: that is, reflect the portion of the polygon on one side of the diagonal over to the other side, and leave the rest unchanged. In general, one can imagine folding forever, according to some (possibly random) scheme for selecting the diagonals. The result is an infinite sequence of polygons, or "walk" in the space of polygons: we call the associated sequence of polygon vertices a *polypath*. This paper studies the dynamics of polypaths arising from quadrilaterals.

Formal definitions and elementary properties of the folding transformation are given in Section 1. If we start things off with an isosceles trapezoid and proceed to fold by choosing the diagonals cyclically (Section 2), the polypath frequently appears to be dense or to recur with high period on a complicated curve whose geometric details depend in some unpredictable manner on the initial trapezoid. Thus the system manifests sensitive dependence on initial conditions. Despite the complexity of the curves, they possess certain symmetry properties that make them attractive objects of study. Figure 1 hints at the wild diversity of polypaths we have encountered for such initial quadrilaterals.

1058-6458/93 $0.50 per page

Although all polypaths in Figure 1 are bounded, it is easy to generate unbounded polypaths, and it seems to be a difficult problem to predict the outcome for a particular initial quadrilateral. Consider, for example, the two-parameter family of initial quadrilaterals that are isosceles trapezoids having a specified perimeter. Figure 4 (left) shows the pattern in parameter space produced by coloring points according to whether their associated polypaths appear to be bounded or unbounded.

To help explain polypath behavior we study in Section 3 an associated sequence of pairs (x, y) of squared diagonal lengths. We show that generically the points (x, y) lie on an irreducible, nonsingular cubic curve—in other words, an elliptic curve. In fact, they are contained in the bounded component of the curve, which we call the *shoe*. The polypath dynamics are related to a billiards-like dynamical system on the shoe, characterized algebraically as translation along the curve (with respect to the standard operation of addition on elliptic curves). When the translational constant is irrational, the billiard trajectories on the shoe are dense, which is consistent with a polypath producing a closed curve. When the constant is rational, the trajectories are periodic, likely with high period.

The connection between periodicity on the shoe and polypath dynamics is examined in Section 4. We show that an isosceles trapezoid is taken under iteration to a (possibly displaced) copy of itself if and only if the associated sequence of squared diagonal pairs is periodic on the shoe, and, consequently, if and only if the sequence contains one of five special points on the shoe. These results, together with some numerical exploration, lead to some conjectures about the topological structure of the subset of parameter space on which the corresponding shoe dynamics is variously periodic or dense. The related subset associated with polypaths that appear dense on closed curves (as in Figure 1) has also proved difficult to characterize completely, but we verify some simple examples.

We mention a few papers that have some connection, at least in spirit, to the present one. G. A. Galperin [1987, p. 197] describes the process of straightening a billiard trajectory on a triangular table. This involves the successive reflections of the triangle about its edges, as they are hit by the ball; of course, all the triangles are congruent and pairwise contiguous, but it takes 45 of them before the original triangle reappears in the original orientation. The periodic properties of a sequence of pedal triangles was examined by J. Kingston and J. Synge, and, in a later paper, Peter Lax proved related ergodic properties (see [Lax 1990] for all references). M. Mendès-France [1983] describes the beautifully intricate curves associated with exponential sums and related paper-folding dynamics.

ACKNOWLEDGEMENTS

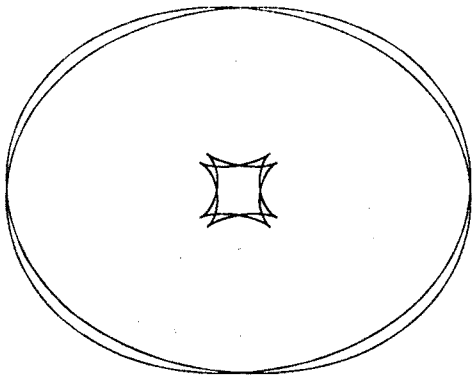
We thank Ted Lewis for introducing us to the remarkable phenomena associated with polygon folding, and for providing us with a sample computer program.

We are greatly obliged to the two referees for many helpful suggestions. In particular, one referee made the crucial observation that the shoe is a cubic curve, while the other made numerous suggestions for improvement both in style and substance.

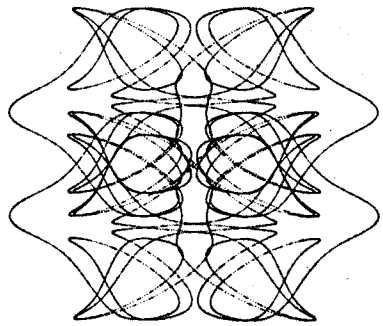
We thank Tom Harke for his collaboration in the early stages of the paper and for programming the computer graphics of the first three figures. We are grateful to Ted Lewis, Jim Pounder and Al Weiss for useful conversations, and particularly to Jim Lewis for his patient advice on elliptic curves.

1. THE CYCLIC FOLDING OF QUADRILATERALS

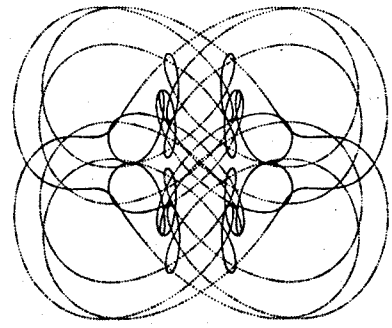
Consider a quadrilateral $PQRS$ (Although our figures will show the edges, we will formally think of a quadrilateral simply as an ordered set of four points in the plane, which may or may not be in general position. We will also use the term *directed quadrilateral* when we want to emphasize the order of the vertices.)



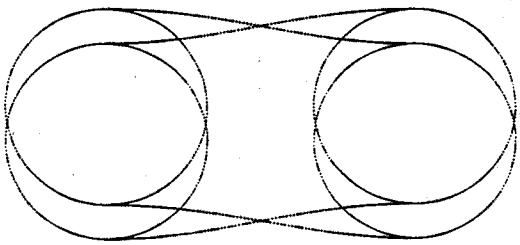
$\mu = 0.05 \quad \nu = 0.1501$



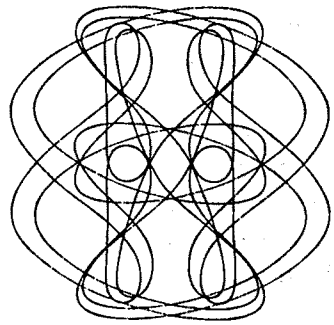
$\mu = 0.45 \quad \nu = 0.4501$



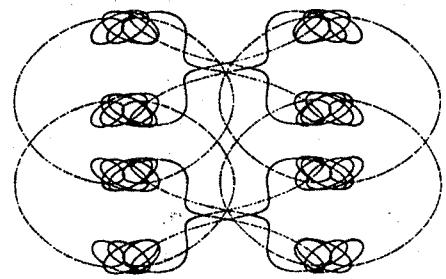
$\mu = 0.3399 \quad \nu = 0.6599$



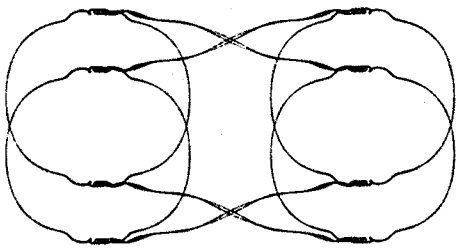
$\mu = 0.15 \quad \nu = 0.1501$



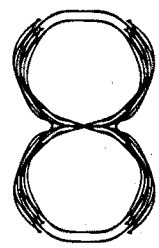
$\mu = 0.4596 \quad \nu = 0.53158$



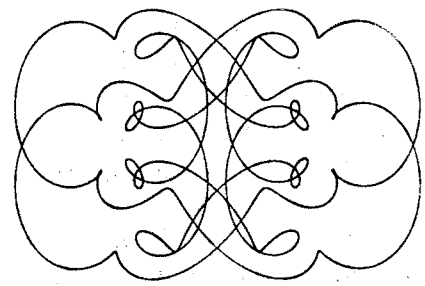
$\mu = 0.35985 \quad \nu = 0.35915$



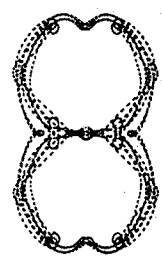
$\mu = 0.25 \quad \nu = 0.2501$



$\mu = 0.4596 \quad \nu = 0.5396$

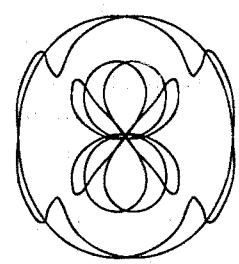


$\mu = 0.31995 \quad \nu = 0.35915$



$\mu = 0.5001 \quad \nu = 0.499899$

$\mu = 0.47955 \quad \nu = 0.51955$



$\mu = 0.31995 \quad \nu = 0.63985$

FIGURE 1. Typical bounded polypaths arising from isosceles trapezoids. For each example, the values of μ and ν for the initial trapezoid are given (see Section 2). Except for the one in the bottom left, all polypaths are drawn to the same scale and represent 10000 iterations of the folding map. The remaining picture is drawn twenty times smaller, and shows 40000 iterations.

If $Q \neq S$, we can define a new quadrilateral by *folding P over the diagonal QS*, that is, by replacing P with its image under reflection in QS. In general, we can fold any vertex over the diagonal determined by the adjacent vertices, provided the adjacent vertices do not coincide.

Given a (finite or infinite) sequence made up of the letters $\{P, Q, R, S\}$ we get a sequence of quadrilaterals by iterating the folding process: fold the vertex labeled by the first letter, then take the result and fold the vertex labeled by the second letter, and so on. For instance, if the sequence is PP, we fold the vertex P and then its image, so we end up with the original quadrilateral.

The only barrier to endless folding is that eventually the images of P and R or of Q and S may coincide. These degenerate situations are easily characterized. Define a *dart* to be a quadrilateral with two pairs of adjacent sides equal: in particular, any quadrilateral with $P = R$ or $Q = S$ is a dart. Folding preserves edge lengths, so it transforms darts into darts and nondarts into nondarts. Therefore any nondart can be folded forever, following any sequence of letters.

From now on we will focus on *cyclic folding*, that is, folding according to the cyclic sequence $PQRSPQRS\ldots$ Alternatively, we can think of the basic step in cyclic folding as being the correspondence $PQRS \to QRSP'$, where P' is the image of P under reflection in QS. Although this operation is not, strictly speaking, a folding (since it involves a reordering of the vertices), this point of view has the advantage of uniformity: every step is the same. Graphically, the directedness of a quadrilateral can be elegantly encoded in the form of an arrow, pointing from the last vertex to the first, so the rule for cyclic folding is: fold the vertex at the tip of the arrow, then advance the arrow one vertex (Figure 2).

Cyclic folding can thus be regarded as a map π from the space of nondarts to itself, and its iteration defines a dynamical system. We describe some elementary properties of π. If $P, Q, S \in \mathbf{R}^2$ with $Q \neq S$, we denote by $P|_Q^S$ the reflection of P in the

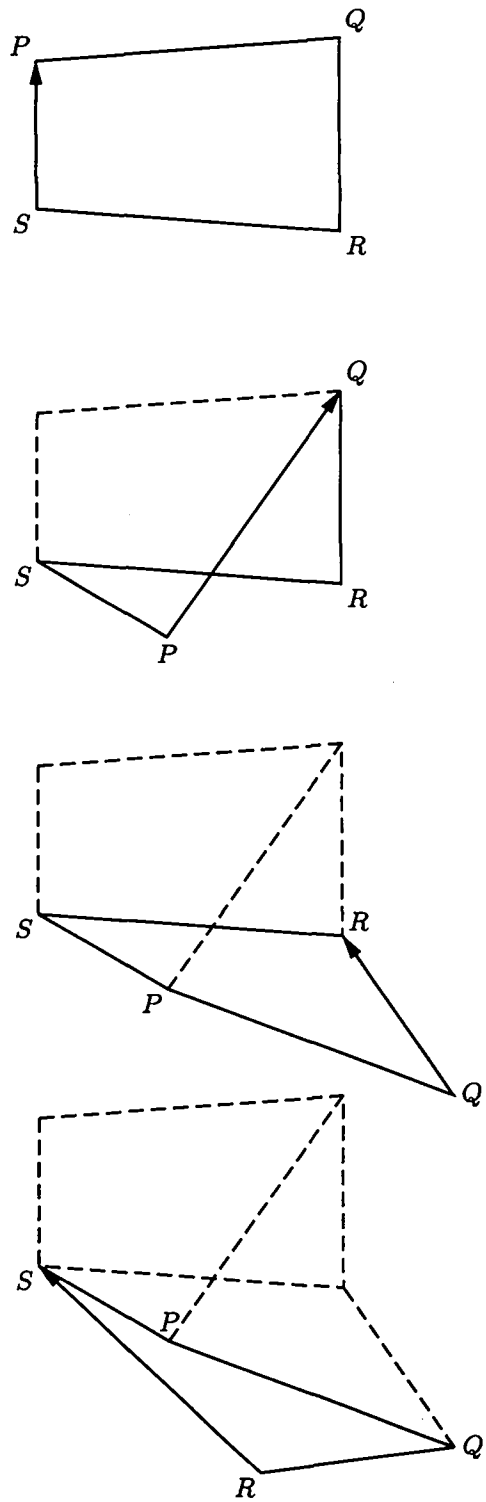


FIGURE 2. An initial quadrilateral $PQRS$ (top) and its first three iterates under cyclic folding.

line QS. Under the usual identification of $\mathbf{R}^2$ with $\mathbf{C}$, we have

$$P|_Q^S = \frac{\overline{(P-Q)}(S-Q)}{\overline{S-Q}} + Q.$$

This follows from the geometrically clear fact that $(P|_Q^S - Q)/(S-Q)$ is the complex conjugate of $(P-Q)/(S-Q)$. The map π is defined by

$$\pi(PQRS) = QRSP|_Q^S,$$

where $PQRS \in \mathbf{C}^4$ is a nondart. We denote the set of such quadruples $PQRS$ by Ω, and remark that it is an open subset of $\mathbf{C}^4 = (\mathbf{R}^2)^4$. The following properties are immediate:

Proposition 1.1. *π is a real analytic diffeomorphism of Ω, with inverse $r \circ \pi \circ r$, where r is the involution $r(PQRS) = SRQP$. Its Jacobian has determinant 1, so π preserves Lebesgue measure.*

Remark. That π preserves Lebesgue measure in Ω suggests a possible application of the Poincaré recurrence theorem: if there exists a subset of Ω with positive, finite Lebesgue measure that maps into itself under π, then almost every point in that subset is recurrent under π. Unfortunately, while we have ideas about how to construct the required subset (for instance, a good candidate might be the union of the closures of all orbits that stay bounded within some ball of large radius about the origin), we have not been able to prove that any such set has positive measure. Section 4 contains a more detailed description of some of these problems.

Since π is a diffeomorphism, we can iterate it backward as well as forward. We define a *polypath* as a (forward and backward) orbit of an element of Ω under π. We also call a polypath the image of such an orbit under the projection $PQRS \mapsto P$, that is, the sequence of first vertices of the iterates of an element of Ω. (Note that a polypath is *not* a path in the usual sense of a map from the interval into a space.)

If we look at every fourth element of a polypath, we get the forward and backward images of a single vertex of the starting polygon. There are four such subpolypaths: sometimes they are easy to tell apart, and sometimes not. Both cases occur in Figure 1.

2. POLYPATHS OF ISOSCELES TRAPEZOIDS

Any isometric image of a polypath is a polypath, and likewise any homothetic (scaled) image. Thus it is reasonable to consider similar quadrilaterals as equivalent. This reduces the dimension of the space of possible starting quadrilaterals from eight to four. This space is still somewhat too big, so we now restrict our study to the orbits of initial quadrilaterals that are isosceles trapezoids. As we see from Figure 1, this still leaves lots of interesting polypaths.

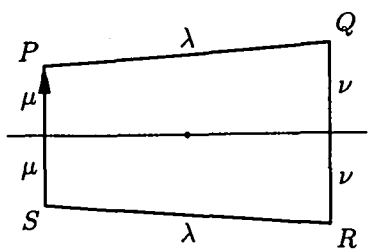


FIGURE 3. Every isosceles trapezoid is similar to a standard one, that is, one that is arranged as shown here and that satisfies $\mu + \nu + \lambda = 2$. (The central dot is the origin and is equidistant from the two parallel sides.)

As representatives of each similarity class of isosceles trapezoids, we take those of the form shown in Figure 3, where $\mu + \nu + \lambda = 2$. Such initial trapezoids are called *standard*. Standard initial trapezoids are uniquely specified by the parameters $(\mu, \nu) \in [0,1] \times [0,1]$. However, we have to omit the point $(\mu, \nu) = (\frac{1}{2}, \frac{1}{2})$, which gives a dart. The case $\mu = 0$ or $\nu = 0$ gives an isosceles triangle; the case $\mu = 1$ or $\nu = 1$ gives four collinear points. These special cases are discussed further in the next subsection and in Example 4.5.

For a standard initial trapezoid, complex conjugation is the same as reversing the order of the vertices. By Proposition 1.1, this implies that the

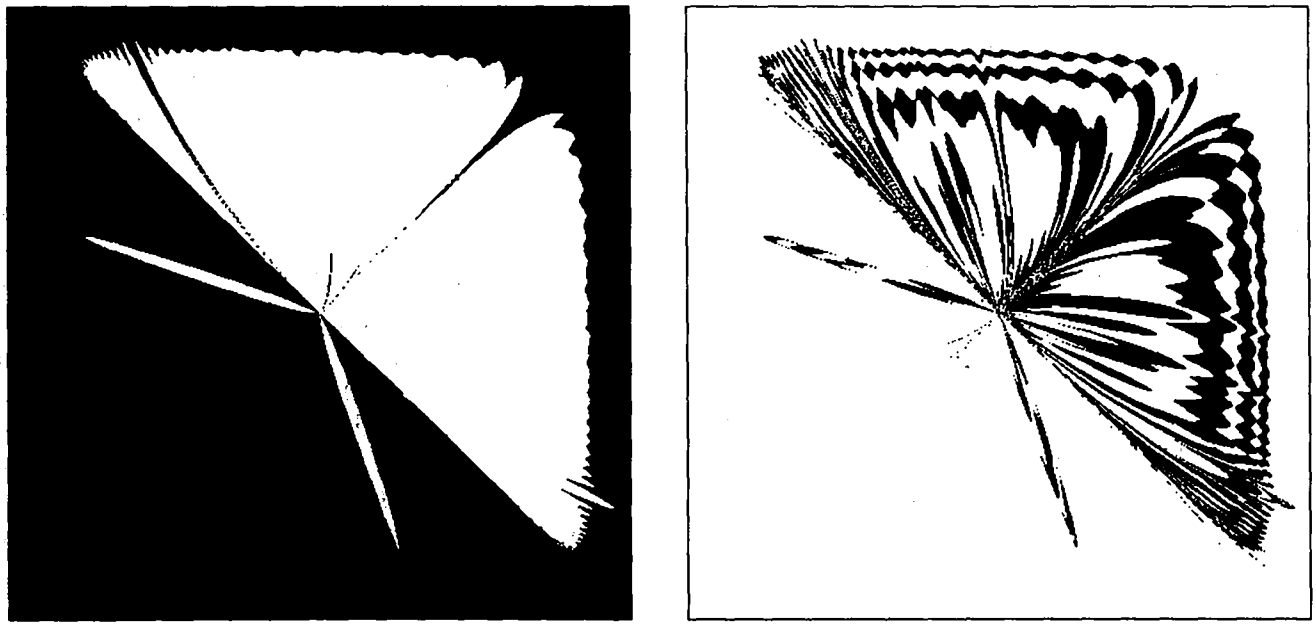


FIGURE 4. The butterfly lives in the space of standard initial trapezoids, $(\mu, \nu) \in [0,1] \times [0,1]$. Left: a pixel is colored white if and only if the polypath for the corresponding initial trapezoid goes out of a ball or radius 50 about the origin sometime in the first 100 iterations. Right: Colors inside the butterfly are chosen by looking at the number n of iterations needed to get out of the ball of radius 50, then dividing by 10 and taking the integer part of the quotient. Black means the result is even, white means odd. The resolution is 640 × 480.

backward orbit under π is the same as the conjugate under the forward orbit; in particular, the two-sided orbit is symmetric with respect to the real axis.

Remark 2.1. The space of isosceles trapezoids is not invariant under π. In this sense it would be more natural to consider instead the space of all *isosceles quadrilaterals*, that is, those that have two opposite sides of the same length. This space is three-dimensional (after modding out by similarities) and invariant under π. However, we will see in Section 3 that we don't gain significantly in generality by considering this bigger space.

Bounded and Unbounded Polypaths

All polypaths selected for Figure 1 are bounded, but there are also unbounded ones. Figure 4 is an attempt to map the region of $\mu\nu$-space that corresponds to bounded polypaths; roughly speaking, the white region (for the figure on the left) corresponds to unbound polypaths. From this figure it appears that the topology of the two regions may be quite complicated. However, one must be careful in drawing conclusions, for at least two reasons:

- The butterfly is not quite symmetric with respect to interchange of $\mu = \nu$, but it is easy to see that the locus of (μ, ν) giving rise to bounded polypaths is symmetric. The asymmetry in the figure is an artifact of the discretization (the resolution being different in the μ and ν directions) and of our cutoff criteria.
- The edges of the square, given by $\mu = 0, 1$ or $\nu = 0, 1$, give polypaths easily seen to be bounded, and so are correctly colored black on the left. However, closer investigation suggests that the region near the edges $\mu = 1$ and $\nu = 1$ contains unbounded polypaths, and so should be colored white (although the rate of divergence slows as we approach the edges). Similarly, there are unbounded polypaths arbitrarily close to the edges $\mu = 0$ and $\nu = 0$ (see the discussion at the end of Section 4).

Roughly, then, what we can say is that polypaths generally appear to be bounded for starting trapezoids with $\mu+\nu > 1$, and unbounded for those with $\mu+\nu < 1$. The most conspicuous exception is the butterfly's antennae, along the lines $3\mu+\nu = 2$ and $\mu+3\nu = 2$ (see also Remark 4.2). We will return to this point at the end of Section 4.

3. THE SHOE

To keep track of the orbit of a single quadrilateral $PQRS$ under π, it is enough to record the lengths of the diagonals PR and QS, because π preserves side lengths, and knowledge of the sides and diagonals is sufficient to determine a quadrilateral up to congruence. Moreover, these six lengths satisfy

$$xy^2+yx^2-(a+b+c+d)xy+(a-d)(b-c)x \\ +(a-b)(d-c)y+(bd-ac)(b+d-a-c)=0,$$

where $x = PR^2$, $y = QS^2$, $a = PQ^2$, $b = QR^2$, $c = RS^2$ and $d = SP^2$. (We are very grateful to a referee, who brought this relation and its proof to our attention.) One way to show this is to write down the formula for the volume of the parallelepiped spanned by the vectors $Q-P$, $R-P$ and $S-P$, for $P,Q,R,S \in \mathbf{R}^3$, and then apply some standard linear algebra; if the vertices are coplanar, the volume is zero. Details can be found in [Berger 1987, p. 237–238], for example.

Thus the pairs (x,y) of squared diagonals lie on a cubic curve in $\mathbf{R}^2$. For our standard initial trapezoids (Figure 3), the coefficients become $a = c = (2-\mu-\nu)^2$, $b = 4\mu^2$ and $d = 4\nu^2$, where $(\mu,\nu) \in [0,1]^2$. The cubic is therefore given by

$$xy^2 + yx^2 - Qxy - L(x+y) + C = 0, \qquad (3.1)$$

where

$$Q = 4(\mu^2+\nu^2) + 2(\mu+\nu-2)^2,$$
$$L = (4\mu^2 - (\mu+\nu-2)^2)(4\nu^2 - (\mu+\nu-2)^2),$$
$$C = (16\mu^2\nu^2 - (\mu+\nu-2)^4) \\ \times (4(\mu^2+\nu^2) - 2(\mu+\nu-2)^2).$$

We denote this curve—or rather its projective completion—by $X_{\mu\nu}$. We will show that, except for (μ,ν) on certain lines, $X_{\mu\nu}$ is nonsingular and irreducible, hence smooth. It is symmetric with respect to the line $x = y$, and always has three points at infinity, with homogeneous coordinates $[1,0,0]$, $[0,1,0]$ and $[1,-1,0]$; the latter is an inflection point. Each coordinate axis intersects $X_{\mu\nu}$ twice at infinity (a tangency) and once at $[0,C,L]$ (for the x-axis) or $[C,0,L]$ (for the y-axis); when $L = 0$ this third intersection is at infinity as well, and therefore a point of inflection. The asymptotes $x = 0$, $y = 0$ and $x+y = Q$ form a triangle that cages the bounded component $\Sigma_{\mu\nu}$, which we call the *shoe*. Since $Q > 0$, the shoe lies in the first quadrant. Figure 5 summarizes the situation.

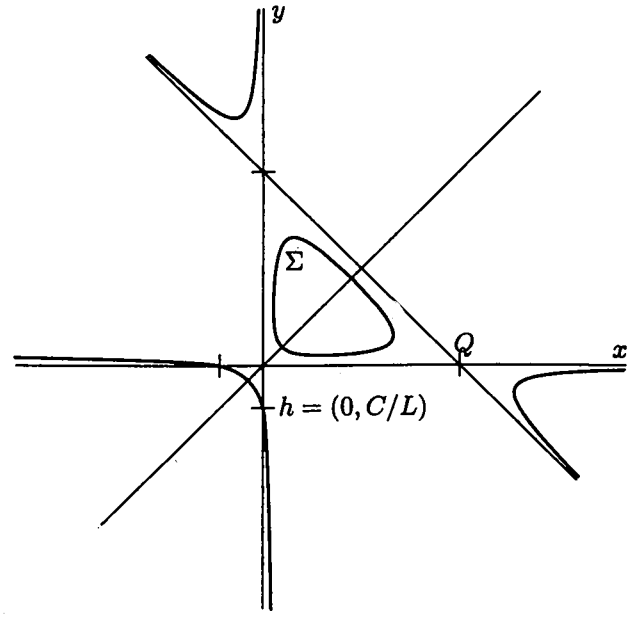


FIGURE 5. The cubic $X_{\mu\nu}$, for $\mu = 0.3$ and $\nu = 0.625$.

Remark 3.1. We have $X_{\mu\nu} = X_{\nu\mu} = X_{1-\mu,1-\nu}$, because the coefficients Q, L and C are the same in each case. Thus the same curve is associated with up to four standard initial trapezoids (congruent in pairs). Corresponding diagonal lengths start and remain the same for these quadrilaterals. See also Figure 9.

The usefulness of the shoe is explained by Figure 6. First, it is easy to see that $\Sigma_{\mu\nu}$ contains exactly those pairs (x, y) that can be achieved as squared diagonal lengths of quadrilaterals with the given sides; other points of $X_{\mu\nu}$ are not physically realizable. Now suppose we start with a quadrilateral corresponding to a point $p \in \Sigma_{\mu\nu}$. Any fold corresponds to a movement along either a vertical or a horizontal line, since it fixes one of the diagonals. Successive folds alternate between vertical and horizontal movements. We can think of this evolution as a game of billiards, somewhat unconventional in that it violates the law of conservation of momentum, but quite interesting in its own right. (At exactly four points on the shoe, these rebounds are degenerate, and the billiard "spins" before moving on; compare Proposition 4.3.)

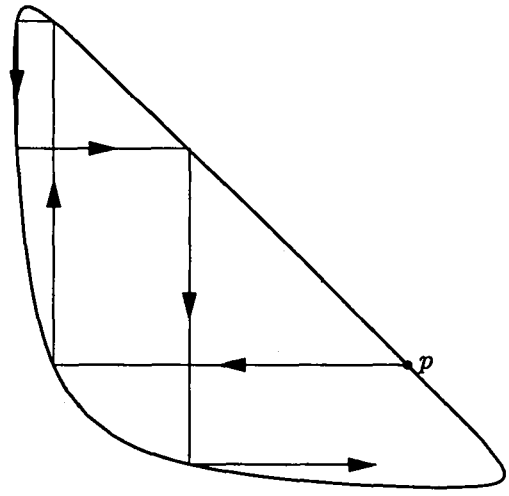


FIGURE 6. Folding along a diagonal amounts to moving directly across or vertically in the shoe.

An Algebraic Dynamical System

As already mentioned, bounded polypaths tend to appear distributed densely (uniformly, in fact) on a curve. The intricate curves of Figure 1 are closures of polypath orbits, just as the circle is the closure of the orbit of an irrational translation. This behavior is partially explained by the existence of dense billiard orbits on the shoe—partially, because the position of a vertex does not depend only on the values of (x, y); one can imagine dense or periodic behavior in terms of the diagonal lengths, together with an unbounded polypath. For the periodic case, see Proposition 4.1 and the ensuing discussion.

We start by showing that alternate hits of the billiard are governed by a simple rule, amounting to translation on an elliptic curve. Recall (from [Koblitz 1984], for example) that, given an elliptic curve X over $\mathbf{R}$ (say) and a point $\mathbf{0}$ on X, possibly at infinity, there is a natural abelian group law on X with $\mathbf{0}$ as its identity element. To add two points $a, b \in X$, draw the line through a and b (or the tangent line if $a = b$), and call c the third intersection of this line with X. Then draw the line through $\mathbf{0}$ and c; the third intersection of this line with X is $a + b$. A particularly nice case of this geometric construction is when $\mathbf{0}$ is a point of inflection: then three points on the curve are collinear if and only if their sum is $\mathbf{0}$.

Let $\mathcal{U}$ be the open square $(0, 1)^2$ minus the diagonals $\mu = \nu$ and $\mu + \nu = 1$. We have $C - LQ \neq 0$ in $\mathcal{U}$ (the locus of $C = LQ$ is the union of the lines $\mu = \nu$, $\mu + \nu = 0$, $\mu + \nu = 2$).

Theorem 3.2. *Suppose $(\mu, \nu) \in \mathcal{U}$. Then the cubic $X_{\mu\nu}$ is an elliptic curve. If p is a point on the shoe $\Sigma = \Sigma_{\mu\nu}$, its image after two iterations of the billiard map (one horizontal and one vertical) is $p+h$, where h is the point on the curve with homogeneous coordinates $[0, C, L]$, and the group law is defined by setting $\mathbf{0} = [1, -1, 0]$. In particular, alternate hits in the full billiard orbit of p form a coset of $\{h\}$, the cyclic subgroup generated by h.*

Proof. We reduce (3.1) to Weierstrass normal form, by performing the following changes of variable in succession (each row indicates that we replace x and y simultaneously by the given expressions):

$$x \to \frac{1 + Qx + y}{2x}, \qquad y \to \frac{1 + Qx - y}{2x},$$

$$x \to x + B, \qquad y \to y,$$

$$x \to (C - LQ)^{-1/3} x, \qquad y \to y,$$

where

$$B = \frac{4L - Q^2}{12(C - LQ)}.$$

The result is $y^2 = 4x^3 - g_2x - g_3$, where

$$-g_2 = \frac{2(6B^2(C - LQ) + B(Q^2 - 4L) + Q)}{(C - LQ)^{1/3}},$$

$$-g_3 = 4B^3(C - LQ) + B^2(Q^2 - 4L) + 2BQ + 1.$$

Therefore the discriminant $\Delta = g_2^3 - 27g_3^2$ equals

$$\frac{4096\,\mu^2\nu^2(\mu - 1)^2(\nu - 1)^2(\mu + \nu - 1)^2}{(\mu - \nu)^2(\mu + \nu)^2(\mu + \nu - 2)^2},$$

which is positive for $(\mu, \nu) \in \mathcal{U}$. This proves that $X_{\mu\nu}$ is irreducible and has no singularities; and so is an elliptic curve.

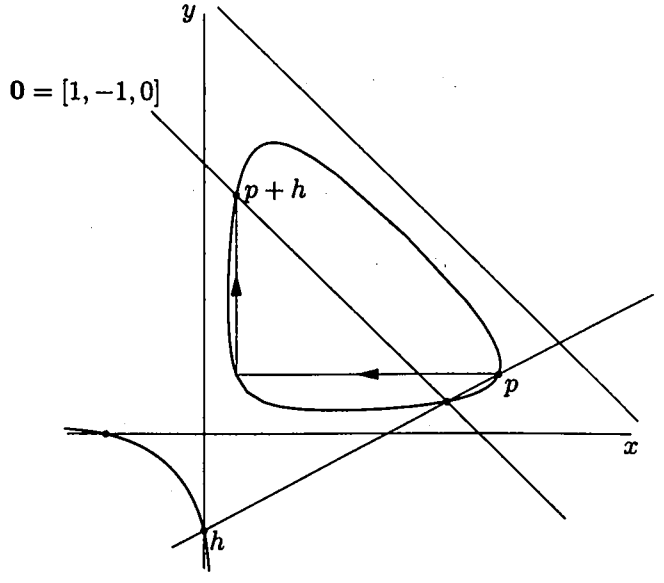


FIGURE 7. Two applications of the billiard map, one across and one up or down, amount to translation by h in the shoe.

We use Figure 7 as a guide in proving the rest of the theorem. Let p' be the unique point on the shoe horizontally across from p and let p'' be the unique point on the shoe vertically up or down from p'. By the group law, we can write

$$\mathbf{0} = p + p' + [1, 0, 0],$$
$$\mathbf{0} = p' + p'' + [0, 1, 0],$$
$$\mathbf{0} = \mathbf{0} + [1, 0, 0] + [0, 1, 0],$$

since $[1, -1, 0]$ is an inflection point and each triple of points on the right is collinear. Eliminating p' and $[1, 0, 0]$, we get

$$p'' = p + [0, 1, 0] + [0, 1, 0].$$

But $[0, 1, 0] + [0, 1, 0]$ is the intersection of the vertical tangent (the y-axis) with $X_{\mu\nu}$, and this, as already mentioned, is the point $h := [0, C, L]$. Therefore we can write $p'' = p + h$, as we wished to show. □

Remarks. (1) When $\mu = \nu$ the initial trapezoid is a rectangle. Then $C = LQ$ and $X_{\mu\nu}$ is reducible. It becomes $(xy - L)(x + y - Q) = 0$, and the shoe contains a source-sink pair at the points of intersection of the linear and quadratic components (this was noted by a referee).

(2) All along the line $\mu + \nu - 1 = 0$ the discriminant is zero, and one may easily verify that $X_{\mu\nu}$ has a double point on the line $y = x$, i.e., is a nodal cubic.

(3) The point h is at infinity if and only if $L = 0$, and this happens only when $3\mu + \nu - 2 = 0$ or $3\nu + \mu - 2 = 0$ (along the antennae of the butterfly).

The group structure on a real elliptic curve makes it isomorphic to the direct sum $\mathbf{R}/\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$; in fact, the isomorphism is analytic. For $(\mu, \nu) \in \mathcal{U}$, the shoe corresponds to $\mathbf{R}/\mathbf{Z} \times \{1\}$ (the connected component not containing the identity), and h corresponds to a point in $\mathbf{R}/\mathbf{Z} \times \{0\}$. Thus Theorem 3.2 implies that the billiard dynamics on the shoe is analytically equivalent to translation on the circle. In particular, either all points on the shoe have the same period, or there are no periodic points, depending on whether the order of h is finite or infinite. For instance, when $L = 0$ (see

Remark (3) above), we have $h+h+h = \mathbf{0}$, so every orbit has period three.

If we are satisfied with billiard dynamics as a kind of coarse-grained representation of a polypath, the study of polypaths involving isosceles quadrilaterals (see Remark 2.1) gives us nothing new beyond what we get by looking at polypaths that start from a standard isosceles trapezoid. Indeed, an isosceles quadrilateral with sides 2μ, λ, 2ν and λ can first be normalized so $\mu+\nu+\lambda = 2$, then placed on the shoe of $X_{\mu\nu}$ according to the length of its diagonals. Its dynamics will, in essence, be the same as that of the standard isosceles trapezoid corresponding to one of the intersections of the same shoe with the line $x = y$.

4. A CLOSER LOOK AT PERIODICITY

Proposition 4.1. *For $(\mu, \nu) \in \mathcal{U}$, the following statements are equivalent:*

(i) $h = [0, C, L] \in \Sigma = \Sigma_{\mu\nu}$ *has finite order.*

(ii) *Some iterate of π acts on all isosceles quadrilaterals of sides $2\mu, \lambda, 2\nu, \lambda$, where $\lambda = 2-\mu-\nu$, as a rigid motion (that is, an orientation-preserving isometry of the plane).*

(iii) *Some iterate of π acts on some such quadrilateral as a rigid motion.*

Proof. Let $PQRS$ be a quadrilateral as in (ii). It follows from the discussion in the previous section that the map D that assigns to a quadrilateral its squared diagonal lengths makes the diagram

$$\begin{array}{ccc} \mathcal{O} & \xrightarrow{\pi^2} & \mathcal{O} \\ {\scriptstyle D}\downarrow & & \downarrow{\scriptstyle D} \\ \Sigma & \xrightarrow{\tau} & \Sigma \end{array} \qquad (4.1)$$

commute, where $\mathcal{O}$ is the π-orbit of the quadrilateral and $\tau : \Sigma \to \Sigma$ is translation by h.

(There is a slight subtlety here: π is not just a folding but a folding plus relabeling of vertices, so the action of π really amounts to a horizontal move in the shoe followed by interchange of x and y. Two of these, however, give the same as the horizontal-plus-vertical move used in establishing Theorem 3.2.)

If h has period m, the directed quadrilateral $\pi^{2m}(PQRS)$ is congruent to $PQRS$, since it has same diagonals and same sides. By doubling again we ensure that the isometry between the two is orientation-preserving. Thus (i) implies (ii).

Clearly (ii) implies (iii). To show that (iii) implies (i), we take a power of π that acts as a rigid motion, and double it if odd. Using (4.1), we see that the image of the initial quadrilateral under D is a point in Σ that is taken to itself by a power of τ; this implies that h has finite order. □

Proposition 4.1 allows us to characterize polypaths that are unbounded in the case that h has finite order. If the rigid motion of statement (ii) has a nontrivial rotational component, or is the identity, the corresponding polypath is bounded (either finite or dense in a curve); otherwise the rigid motion is a translation, and the polypath "walks away" as the translation is iterated.

Example 4.2 (butterfly's antennae). A standard isosceles trapezoid with $2\mu = \lambda$ is on the line $3\mu+\nu = 2$, and, as mentioned near the end of Section 3, h has period three in this case. Thus the twelfth iterate of π is an isometry (this is also easy to see directly). It is a translation by a distance $4\nu(1+\sin(\alpha-\pi/2))$, where α is the angle PQR (Figure 8).

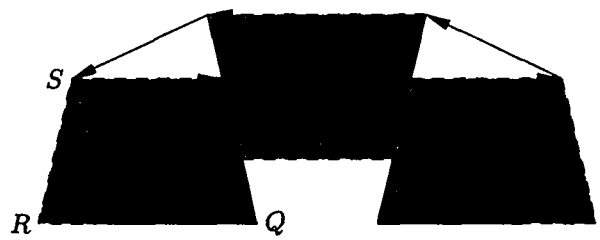


FIGURE 8. The image of a trapezoid $PQRS$ with three equal sides under three iterations of π (black outline), six iterations (upside-down gray), nine and twelve iterations. The last of these is a translate of the original trapezoid.

Proposition 4.3. *Let $(\mu, \nu) \in \mathcal{U}$, and let $\{p, p'\}$ be the intersection of $\Sigma = \Sigma_{\mu\nu}$ with the diagonal $x = y$.*

Let B be the set of contacts with Σ of the forwards and backwards billiards trajectory starting at p (Figure 5). Then (i)–(iii) *in Proposition* 4.1 *hold if and only if either* (a) $p' \in B$, *or* (b) B *contains one of the four points where the slope of Σ is horizontal or vertical.*

Proof. It is enough to show that B is finite if and only if (a) or (b) hold. If B is finite, the billiards trajectory is a closed polygonal path, symmetric with respect to the diagonal (because of the symmetry of Σ). If the path never hits Σ at a point where Σ is parallel to either axis, it never goes back on itself; in other words, each point of B is hit only once per period. Since horizontal and vertical legs alternate, B has an even number of points; by symmetry, p' is one such point.

The converse is obvious, again by symmetry. □

Note that in case (a) of the proposition only every other point of B is in the τ-orbit of p. In particular, p' need not be in that orbit.

Example 4.4. The translation constant h has order two if and only if $C = 0$, which happens on the curves

$$\mu = -2 + 2^{3/2}\sqrt{1-\nu} + \nu \tag{4.2}$$

and

$$\mu = 2 - 2^{3/2}\sqrt{\nu} + \nu. \tag{4.3}$$

Each of these curves is symmetric with respect to the diagonal $\mu = \nu$ of $\mathcal{U}$, and one is mapped to the other by reflection in the diagonal $\mu + \nu = 1$. For points on these curves, the set B has exactly four points (see Figure 9, left). Thus the standard (μ, ν)-isosceles trapezoid is congruent to its image under π^4. On (4.2) this congruence is a 180° degree rotation, so that π^8 acts as the identity and the polypath has only a finite number of distinct points. By contrast, on (4.3), π^4 is a glide-reflection, so π^8 is a translation and the corresponding polypath walks away (Figure 9, right). Compare Example 4.2, where the polypath always walks away.

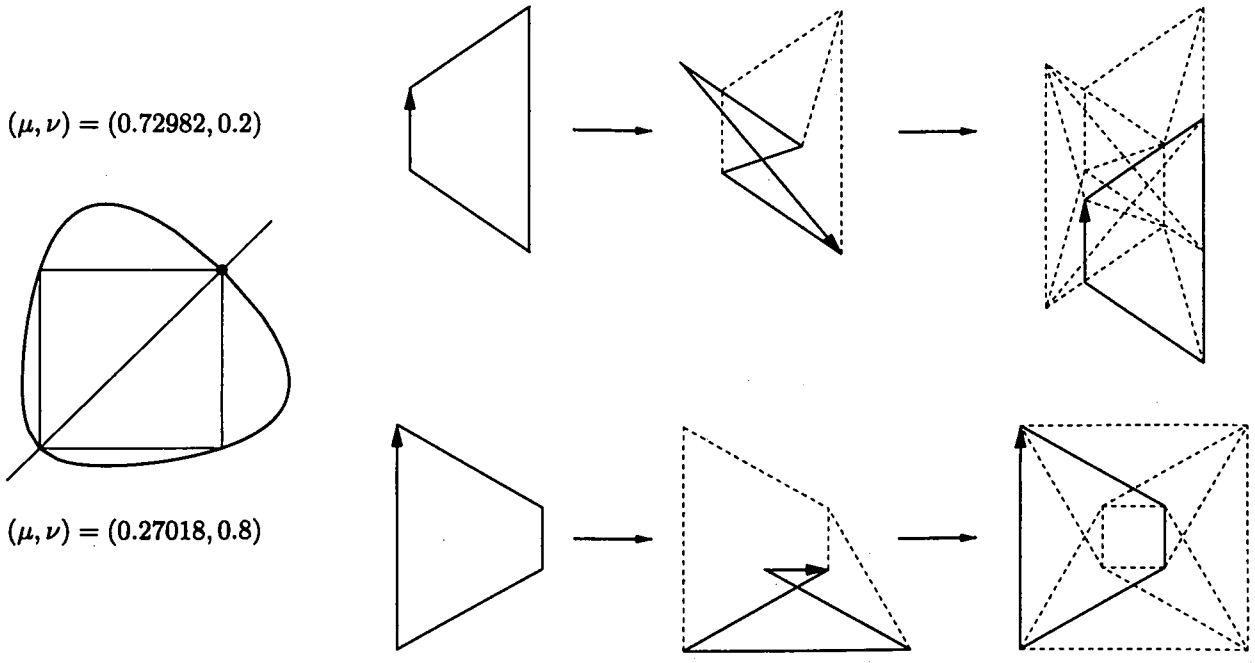


FIGURE 9. Dynamics when h has order two ($C = 0$). Both pairs (μ, ν) correspond to the same shoe (compare Remark 3.1). The polypath is finite or unbounded, depending on whether (μ, ν) satisfies (4.2) or (4.3). Each row shows the original standard trapezoid and its image under π^2 and π^8.

Example 4.5. An even simpler case is when $\nu = 0$ or $\mu = 0$. Then the standard quadrilateral is an isosceles triangle, and π^4 acts as a rotation by twice the apex angle α. If α is an irrational multiple of π, the polypath is dense in a circle; if it is a rational multiple, the polypath is finite.

By perturbing an isosceles triangle T_n with apex angle $\alpha = 2\pi/n$, where n is an integer, we can obtain trapezoids with arbitrary τ-period n. Specifically, we fix $\nu = 0.05$ (say), and consider a standard (μ, ν)-isosceles trapezoid, with μ close to its value for the triangle T_n. We apply π^n to this trapezoid, and record the lengths of its diagonals. Then we solve numerically for the value of μ that makes the two diagonals have the same length. This amounts to solving for the value of μ that makes the τ-orbit of p in Proposition 4.3 have period $m = \operatorname{lcm}(n, 2)/2$: see Figure 10. Then π^{4m} acts on the trapezoid as a rigid motion, which turns out always to be a translation, possibly trivial.

In Example 4.5, one can equally well start from an isosceles triangle with apex angle

$$\alpha = 2\pi\frac{n}{p} < \pi.$$

For small ν, experimentation shows that one can find a value of μ such that the rotation of S^1 to which τ is conjugate has rotation number $2p/n$ (expressed as a fraction of a full turn—that is, the rotation numbers of Figure 10 are $\frac{2}{3}, \frac{2}{4}, \ldots, \frac{2}{10}$).

This leads to the following conjecture. For every real number r with $0 < r < \frac{1}{2}$ there should be a curve K_r in the lower quadrant of $\mathcal{U}$ (defined by

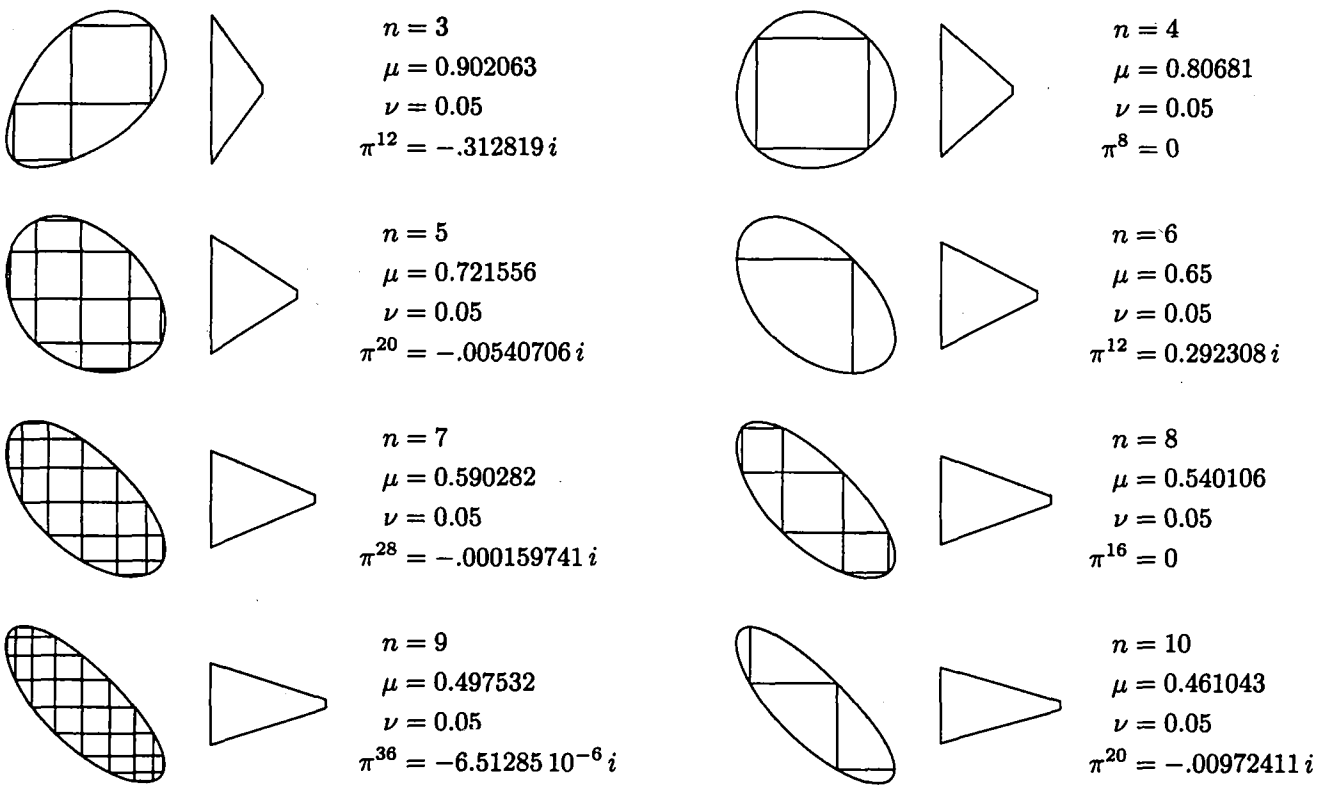


FIGURE 10. Somewhere near the isosceles triangle T_n with apex angle $2\pi/n$ is a standard trapezoid that, like T_n, is mapped to a translate of itself under $\pi^{2\operatorname{lcm}(n,2)}$. Whereas for T_n this translate coincides with the original, this is not always so for the perturbed trapezoid. (The translation vector is indicated as a complex number.) Note that $n = 4$ is the situation of Figure 9 (bottom), and $n = 6$ that of Figure 8.

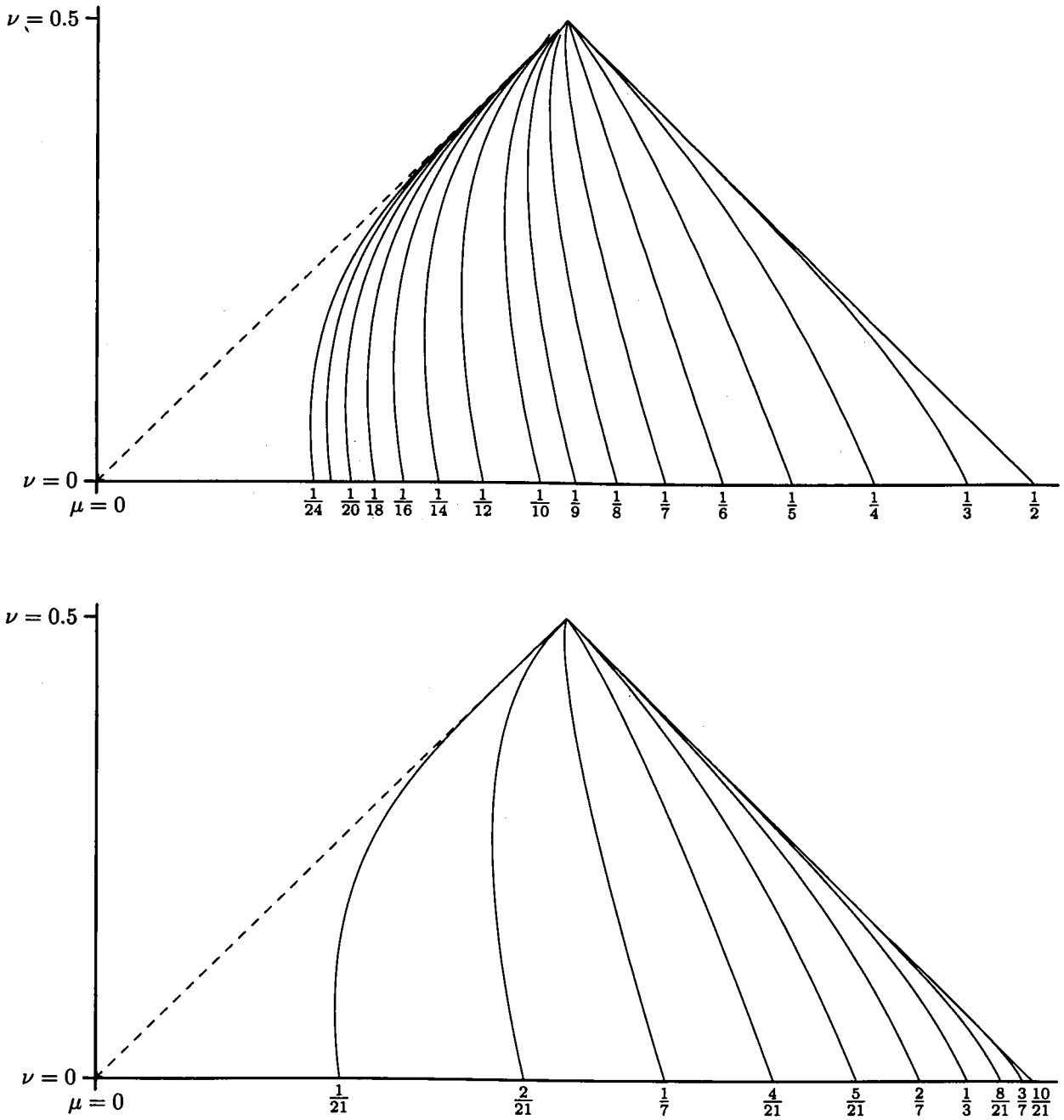


FIGURE 11. Curves $K_{p/n}$ for various values of p and n.

$\nu < \mu$ and $\nu + \mu < 1$), starting at the isosceles triangle with apex angle $2\pi/r$ and tending toward $(\frac{1}{2}, \frac{1}{2})$, such that the translation map τ has rotation number $2r$ along this curve. Some of these curves are illustrated in Figure 11. (The case $p = 1, n = 3$ gives one of the butterfly's antennas.)

Obviously, the picture is the same in the other quadrants as far as τ is concerned, by reflection in the lines $\mu = \nu$ and $\mu + \nu = 1$. However, regarding the asymptotic behavior of polypaths, we have seen (see Example 4.4) that $\nu + \mu = 1$ is not a symmetry axis, although $\mu = \nu$ is. We conjecture that, above the diagonal $\nu + \mu = 1$, all polypaths are unbounded, while below the diagonal, the polypath for a pair (μ, ν) not on a rational curve $K_{p/n}$ is dense on a bounded curve, while a polypath for $(\mu, \nu) \in K_{p/n}$ either is finite or walks away, depending on p/n.

REFERENCES

[Berger 1987] Marcel Berger, *Geometry I*, Springer, Berlin, 1987.

[Galperin 1987] G. A. Galperin, "Non-periodic and not everywhere dense billiard trajectories in convex polygons and polyhedrons", *Comm. Math. Phys.* **91** (1987), 187–211.

[Koblitz 1984] N. Koblitz, *Introduction to Elliptic Curves and Modular Forms*, Graduate Texts in Math. **77**, Springer, New York, 1984.

[Lax 1990] P. D. Lax, "The ergodic character of sequences of pedal triangles", *Math. Mon.* **97** (1990), 377–381.

[Mendès-France 1983] M. Mendès-France, "Entropie, dimension et thermodynamique des courbes planes", in pp. 153–177 in *Séminaire de théorie des nombres, Paris* 1981–82 (edited by Marie-José Bertin), Birkhäuser, Boston, 1983.

Kevin Charter, Department of Mathematics, University of Alberta, Edmonton, Canada T6G 2G1 (kcharter@deneb.math.ualberta.ca)

Thomas Rogers, Department of Mathematics, University of Alberta, Edmonton, Canada T6G 2G1 (author for correspondence) (mathdept@sirius.math.ualberta.ca)

Received January 18, 1993; accepted in revised form December 6