Barrier billiards-a simple pseudo-integrable system

1990 J. Phys. A: Math. Gen. 23 887

(http://iopscience.iop.org/0305-4470/23/6/014)

J. Phys. A: Math. Gen. 23 (1990) 887-900. Printed in the UK

Barrier billiards—a simple pseudo-integrable system

J H Hannay† and R J McCraw‡§

† H H Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK
‡ Physics Department, Hamilton College, Clinton, NY 13323, USA

Received 1 December 1988, in final form 27 October 1989

Abstract. A hard line-segment barrier lies in a periodic cell in the plane, parallel to one edge, and half its length. This constitutes perhaps the simplest *pseudo-integrable* system, and the classical motion of a particle within it admits concise analysis. The hitting and missing of the barrier defines a quasi-periodic (doubly periodic) sequence with the period ratio depending on the invariant |gradient| of the trajectory. The consequent motion is expressed in terms of its continued fraction expansion. Rational gradients correspond to 'periodic' orbits which can be classified in full. Quadratic irrational gradients lead to trajectories of calculable fractal dimension (or power-law correlation decay).

1. Introduction

A plane containing a lattice of hard obstacles off which a particle bounces provides one of the simplest realisations of a two degree of freedom Hamiltonian dynamical system [1]. An example is the 'Sinai billiard' where the obstacle is a circle yielding *ergodic* motion [2, 3]. No obstacle at all, on the other hand, would trivially exhibit *integrable* motion. We shall consider perhaps the simplest obstacle: a hard line-segment barrier parallel to one edge of the periodic cell and half its length. In the full plane this produces an infinite set of parallel dashed lines with equal 'mark' and 'gap' lengths. It suffices to treat a standard representative form (figure 1) with barrier ends lying at integer lattice points. (Any other form is equivalent through a linear transformation.)

The motion in this system is certainly not ergodic: a particle, once started, explores only two different directions, not all directions. On the other hand, neither is the system integrable because the connectivity of the sheets explored in phase space is not topologically that of a torus, but rather a two-handled sphere. The multi-handledness arises from the constants of the motion failing to commute at isolated points (barrier ends), and means that the motion is described by an interval exchange mapping rather than a rotation mapping (appendix 3). Such systems have been called *pseudo-integrable* [4], (or earlier, 'almost integrable' [5], or 'A-integrable' [6]).

The essential distinction between integrable motion, say that in a square box, and our pseudo-integrable motion can be understood physically. In both the box problem and our barrier in a periodic cell problem, the true path of a particle (with its direction changes at bounces) can be conveniently represented by a straightened path: a straight line with the same initial position and gradient, and the same length (figure 2). For our problem every passage of the straightened trajectory through a barrier corresponds

§ Present address: H H Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK

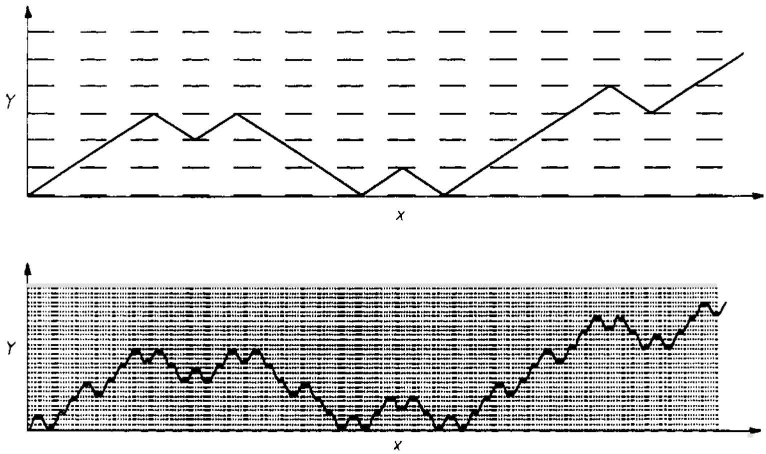

Figure 1. The trajectory from the origin with gradient $(\surd 5-1)/2$ looks similar on two different scales. The lower picture has been scaled anisotropically ('affinely') by a factor $9+4\surd 5$ horizontally and $3+2\surd 2$ vertically.

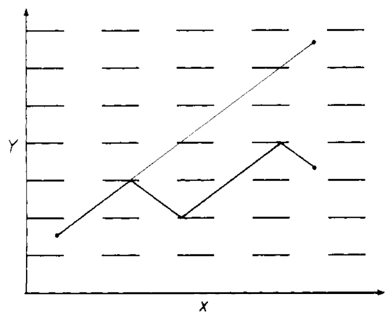

Figure 2. Construction of the straightened representative (faint line) of a trajectory (bold line)—each barrier passage of the former corresponds to a bounce of the latter.

to a bounce of the true trajectory. Similarly for the square problem, every passage of the straightened trajectory through lines of an infinite square net corresponds to a bounce on a wall of the box. In this latter case the position vector of any point on the straightened path *uniquely* determines the true position in the square box (without knowledge of the sequence of intermediate squares the straight line passes through). This characterises the system's integrability. For the pseudo-integrable system this is not so—there is a two-fold ambiguity in the true position (and gradient) in the periodic cell, depending on the number of barriers the straightened trajectory has crossed (i.e. the number of bounces of the true trajectory). Such ambiguities characterise pseudo-integrable systems. Our system is particularly simple because momentum parallel to the barrier is conserved, so this component of the motion is trivial.

Other examples of pseudo-integrable systems are motion inside a regular hexagonal enclosure, or, closer to our system, a 60° rhombus enclosure as studied by Eckhardt *et al* [6]. Indeed, these authors reduce the rhombus billiard to a barrier billiard (with, however, the gap length half the barrier length). An explicit formula is then written down for the trajectory as a sum of increments, positive or negative, according to whether or not a barrier is encountered. They show that the formula is algorithmically non-complex (unlike an equivalent prescription for a chaotic system like the Sinai billiard), and consider it therefore to constitute a solution. It is not, however, an effective algorithm in the sense that a computation is required for each step (to find out whether a barrier is hit or missed). We find for our problem an algorithm which is faster than merely following the trajectory itself: computations are of the order of ln(length) rather than of the order of length (see section 3 below). Actually Eckhardt *et al* also mention ([6], appendix A) the present equal barrier/gap problem and take preliminary steps towards the route developed here (we are grateful to a referee for pointing this out to us).

The character of the motion in our barrier system depends crucially on the rationality of the initial gradient. This can be understood by considering the straightened trajectory. For a rational initial gradient this sequence of barrier passages repeats itself periodically, while for an irrational one the sequence is quasi-periodic (doubly periodic). The true trajectory is therefore, respectively, a repeating zig-zag shape, and a non-repeating one, such as that in figure 1, which shows the trajectory with gradient $(\sqrt{5}-1)/2$, the golden ratio, starting from the origin. As our analysis shows, this is a self-affine fractal, $Y \sim x^{2-D}$, where the dimension D is $2-\ln(3+2\sqrt{2})/\ln(9+4\sqrt{5}) = 1.3984\ldots$. If it had not started from the origin it would generically be a statistically self-affine fractal with the same dimension. (A close analogue would be the Weierstrass function with or without random phases [7].)

The bound version of the problem (the single obstacle in a periodic cell) is easily inferred from that in the extended version. The rational trajectory becomes a periodic orbit, closing and retracing itself. The golden ratio trajectory does not close but passes arbitrarily close to every point in both available directions. The fractal dimension would show up through the power-law decay of the gradient $\langle \mathrm{d}Y/\mathrm{d}x\rangle \sim x^{1-D}$. Power-law correlation decay like this has long been conjectured in pseudo-integrable systems [8, 9].

In fact, we will not need to distinguish, in our calculations, between the cases of rational and irrational gradient. The latter will be treated, as is usual in quasi-periodic (doubly periodic) systems, as the limit of the sequence of rational approximations generated by the continued fraction expansion [10]:

$$|\text{gradient}| = a_0 + 1/(a_1 + 1/(a_2 + 1/(a_3 \ldots \tag{1}$$

For a rational gradient this terminates at some a_n. For quadratic irrationals (roots of quadratic equations with integer coefficients) like the golden ratio, the a_i become a repeating sequence, that for the golden ratio being $1, 1, 1, 1, \ldots$. For typical irrationals the a_i form a random sequence with known statistical properties.

Our goal, then, is to find the vertical displacement $Y(q, p)$ of every trajectory whose straightened representative runs from the origin to a primitive point (q, p) of the integer lattice (one whose q and p share no factors, so that the trajectory hits no lattice points on the way). More precisely, we seek to relate such displacements Y to each other so that long trajectories can be built from short ones or vice versa. This procedure deals with irrational trajectories like that of figure 1, which start from the origin. Ones which do not are also described by the displacements Y, albeit less explicitly: the displacements Y actually represent displacements of bands of trajectories, only the edge ones of which hit lattice points—these are the 'periodic orbit bands' in the bound version of the system (appendix 1). Any irrational trajectory is approximable for an arbitrarily large length by a member of a band of a close enough rational. An alternative explicit procedure for the analysis of a trajectory between arbitrary points is outlined in appendix 2.

The displacements $Y(q, p)$ for q, p coprime with $p \leqslant 8$ are shown in figure 3. We shall always assume $p > 0$ since $Y(q, -p) = -Y(q, p)$ by symmetry. It also suffices to show only those points with $|q| \leqslant p$ because $Y(q+2p, p) = Y(q, p)$ (under the implied even shear of the plane, barriers map into barriers). The asymmetry of the values of Y on the right and the left is striking (though we shall not use this explicitly). We may now anticipate, with some examples, the essence of the full analysis of the next section.

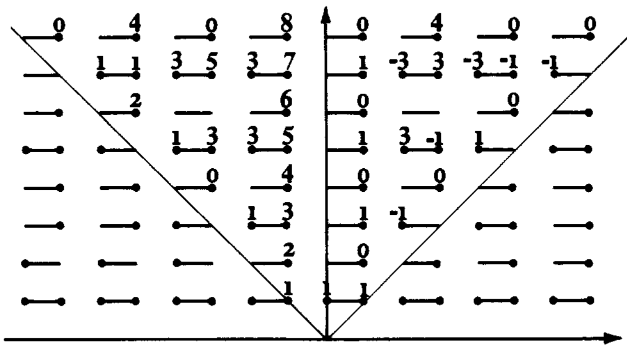

Figure 3. Vertical displacement values $Y(q, p)$ of trajectories whose straightened representatives run from the origin to primitive integer lattice points (q, p). The true trajectory thus zig-zags from the origin to $(q, Y(q, p))$.

Each displacement value can be expressed as the sum or difference of two other values. For example $Y(3, 8) = Y(2, 5) + Y(1, 3)$. The reason is that $(2, 5)$ and $(1, 3)$ span a parallelogram of unit area whose diagonal is $(3, 8)$. Since it has unit area it encloses no lattice points (barrier endpoints) so that the diagonal is 'deformable' into either of the 'two-leg' parallelogram routes. Now, the leg $(2, 5)$-$(3, 8)$ is equivalent to $(0, 0)$-$(1, 3)$ since both start at left barrier ends, hence the stated relation. In contrast, however, the leg $(1, 3)$-$(3, 8)$ is not equivalent to $(0, 0)$-$(2, 5)$, but rather it crosses the same sequence of barriers as $(0, 0)$-$(-2, 5)$, so from this route we have $Y(3, 8) = Y(1, 3) + Y(-2, 5)$. Combined with the first relation, evidently $Y(-2, 5) = Y(2, 5)$. Subtraction rather than addition occurs when the number of barriers crossed by the first leg is odd, so that the starting gradient of the second leg is negative, for example, $Y(2, 7) = Y(1, 3) - Y(-1, 4)$. The general rule (for $q'p'' - p'q'' = \pm 1$) is

$$Y(q'+q'', p'+p'') = Y(q', p') + (q''p' - p''q')(-1)^{(q'+1)(p'+1)/2} Y((-1)^{q'} q'', p'')$$

$$(= \text{the same with primes and double primes exchanged}). \qquad (2)$$

2. String construction

We recall the standard string construction of Klein [11] (figure 4) for continued fractions. A taut string runs from the origin to infinity with an irrational gradient whose continued fraction is sought. There is imagined to be a nail at each point of the unit lattice. When the origin end of the string is now moved one unit up to $(0, 1)$,

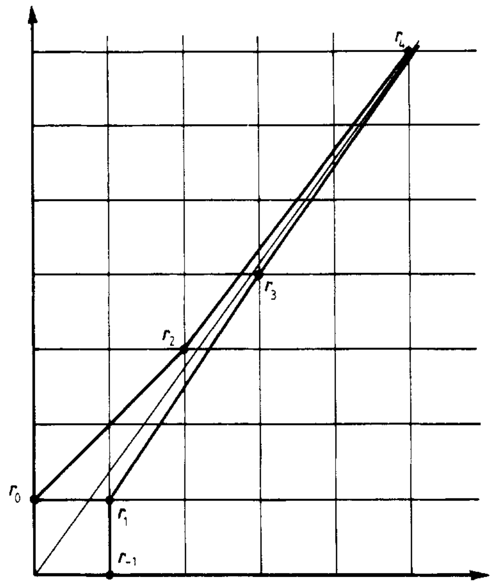

Figure 4. The string construction for continued fractions (from Arnold [12]). The positions $r_k = (q_k, p_k)$ where the displaced strings (bold line) bend correspond to successive rational approximants, p_k/q_k, in the continued fraction of the gradient of the undisplaced string (faint line).

the string catches on a certain sequence of nails above its previous path, bending at some and merely touching others without bending (e.g. $(1,2)$). Similarly, if the end is moved to $(1,0)$, another sequence below is picked out. Together the two sequences of *bending positions* (q_k, p_k) form the successive approximants p_k/q_k to the irrational gradient formed by truncating the continued fraction at a_{k-1} (the $k-1$, rather than k, here, follows Arnold's convention [12]). The upper sequence gives the even values of k, and the lower one the odd values of k. Between each bend the string is divided into a number of equal subsections by nails which it touches without bending. These are the a_i of the continued fraction $(1 < a_i < \infty)$. The successive corner positions (q_k, p_k) obèy the recursion relations (using Arnold's convention)

$$\begin{pmatrix} q_{k+1} \\ q_k \end{pmatrix} = \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_k \\ q_{k-1} \end{pmatrix} \qquad \begin{pmatrix} p_{k+1} \\ p_k \end{pmatrix} = \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_k \\ p_{k-1} \end{pmatrix} \tag{3}$$

and the unit area rule

$$p_k q_{k-1} - q_k p_{k-1} = (-1)^k. \tag{4}$$

We can now interpret this construction in terms of trajectories. The string from the origin represents a straightened irrational trajectory. Consider the barriers which it crosses (bounces of the true trajectory). When the string is pulled aside *it still crosses the same barriers* since the barrier ends have nails which the string does not pass over. This means that just as the irrational gradient g emerges from the sequence of bending points (q_k, p_k), so also the (vertical) displacement of the irrational trajectory emerges from the sequence of the displacements of the rational trajectories from the origin to (q_k, p_k). Each such rational trajectory crosses the same barriers as the corresponding stretch of the irrational line, namely $(0,0)$ to $(p_k/g, p_k)$, and therefore has equal displacement. The geometry which generates the recursion relation (3) for p_k generates a similar one for displacements, as follows.

Let $Y_k^+ = Y(q_k, p_k)$ denote the (vertical) displacement of the trajectory (whose straightened representative runs) from $(0, 0)$ to $\boldsymbol{r}_k \equiv (q_k, p_k)$. It will also be convenient to define $Y_k^- = Y(-q_k, p_k)$. Thus Y_k^- is the displacement in the 'dual' system in which barriers and gaps are exchanged (as achieved by reflection in the vertical axis). This allows us to restrict attention henceforth to $q > 0$, as is intended in the string construction.

Consider the triangle with vertices at the origin, $\boldsymbol{r}_{k-1}$, and $\boldsymbol{r}_{k+1}$. The edge between the last two vectors is divided, by lattice points, into a_k equal subsections. Equating the displacement along $\boldsymbol{r}_{k+1}$ to the sum of those along the other two sides, we have

$$Y_{k+1}^+ = Y_{k-1}^+ \pm Y_k^{\mp} \pm Y_k^{\mp} \ldots \pm Y_k^{\mp} \qquad (a_k \text{ such terms in } Y_k). \tag{5}$$

The sign and the superscript of each term here are decided by the lower endpoint of its subsection. For the mth subsection $(0 \leqslant m \leqslant a_k - 1)$, this point is $\boldsymbol{r}_{k,m} = (q_{k-1} + mq_k, p_{k-1} + mp_k)$. The superscript of the subsection is $ss(\boldsymbol{r}_{k,m})$ where

$$ss(q, p) = (-1)^q. \tag{6}$$

The sign is $(-1)^{k+1} s^+(\boldsymbol{r}_{k,m})$ where

$$s^+(q, p) = (-1)^{(q+1)(p+1)/2}. \tag{7}$$

The power $(q+1)(p+1)/2$, here, is the number of lattice points, excluding the origin, in or on the right-angled triangle with vertices the origin, $(0, p)$, and (q, p), (recalling that q, p are coprime). These lattice points are paired off by barriers except those (n of them, say) at the inner end of barriers which cross the hypotenuse. And it is $(-1)^n$ which gives the required sign of the subsection if k is even, and $-(-1)^n$ if k is odd (as then the point (q, p) does not count). There is a similar formula to (5) for Y_{k+1}^- with all the superscripts reversed, and a sign function $s^-(q, p) = -s^+(-q, p)$. Since the superscript and sign functions are all periodic with period 4 in q and p, the sum of the a_k terms reduces to a sum of four terms. With [ ] denoting integer part, we have

$$Y_{k+1}^{\mp} = Y_{k-1}^{\mp} + (-1)^{k+1} \sum_{m=0}^{3} \left[\frac{(a_k + 3 - m)}{4}\right] s^{\mp}(\boldsymbol{r}_{k,m}) Y_k^{\mp ss(\boldsymbol{r}_{k,m})} \tag{8}$$

or, collecting terms:

$$\begin{pmatrix} Y_{k+1}^+ \\ Y_{k+1}^- \\ Y_k^+ \\ Y_k^- \end{pmatrix} = \begin{pmatrix} M^{++} & M^{+-} & 1 & 0 \\ M^{-+} & M^{--} & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} Y_k^+ \\ Y_k^- \\ Y_{k-1}^+ \\ Y_{k-1}^- \end{pmatrix}$$

with

$$\begin{aligned} M^{\alpha\beta} = {} & (-1)^{k+1} \tfrac{1}{2}(1 + \beta ss(\boldsymbol{r}_{k,0}))\{[(a_k + 3)/4] s^\alpha(\boldsymbol{r}_{k,0}) + [(a_k + 1)/4] s^\alpha(\boldsymbol{r}_{k,2})\} \\ & + (-1)^{k+1} \tfrac{1}{2}(1 + \beta ss(\boldsymbol{r}_{k,1}))\{[(a_k + 2)/4] s^\alpha(\boldsymbol{r}_{k,1}) + [a_k/4] s^\alpha(\boldsymbol{r}_{k,3})\}. \end{aligned} \tag{9}$$

The sequence of 4×4 matrices above governs the growth of the displacement Y_k^+ of the trajectory through the continued fraction expansion of its gradient. Each matrix evidently has unit determinant. Its elements are determined by the current continued fraction element a_k, and the quantities $q_{k-1}, p_{k-1}, q_k, p_k \bmod 4$. These in turn are determined by the previous sequence of $a_i \bmod 4$ through the recursion relations (3).

For a trajectory with quadratic irrational gradient the sequence of elements a_k becomes periodic. This means the $q_{k-1}, p_{k-1}, q_k, p_k \bmod 4$, and hence the whole 4×4 matrix, will become periodic too. The asymptotic growth with k of the vertical displacement Y_k^+ (and Y_k^-) is governed by the largest eigenvalue λ of the product of all the 4×4 matrices in one period (l of them, say). Similarly, the asymptotic growth with k of the horizontal displacement q_k is governed by the largest eigenvalue μ of the product of l successive continued fraction matrices (3). Thus $Y_k \sim \lambda^{k/l}$ and $q_k \sim \mu^{k/l}$, so that $Y_k \sim q_k^{\ln\lambda/\ln\mu}$. As a non-integer power law, this represents a fractal with dimension $D = 2 - \ln\lambda/\ln\mu$. In figure 1, for example, the period l of the 4×4 matrix is 6, $\lambda = 3+2\sqrt{2}$, and $\mu = 9+4\sqrt{5}$.

The four quantities Y_{k-1}^+, Y_{k-1}^-, Y_k^+, Y_k^- are not independent. They are related by two equations based on the displacements along the edges of the parallelogram spanned by $\boldsymbol{r}_{k-1}$ and $\boldsymbol{r}_k$. The net displacements along either route from the origin to $\boldsymbol{r}_{k-1}+\boldsymbol{r}_k$ must be equal since the parallelogram encloses no lattice point (like any unit-area parallelogram). This gives one equation, (corresponding to equating the two alternative right-hand sides of (2)), and the other is the same requirement for the dual system. The two relationships allow the four displacements Y to be expressed either in terms of just Y_{k-1}^+ and Y_k^+, or Y_{k-1}^- and Y_k^-, depending on the set $q_{k-1}, p_{k-1}, q_k, p_k \bmod 4$. This means that the 4×4 matrix can be reduced to a 2×2 matrix with, again, unit determinant. The 2×2 matrix turns out (appendix 4) to have a remarkably simple form for each of the 24 possible sets mod 4 allowed by the unit-area rule (4).

3. Discussion

We have given the algorithm relating the vertical displacement $Y(x, y)$ of any trajectory, with horizontal displacement x from the origin, to the continued fraction of its gradient $(y/x) \equiv g$. It is exponentially fast in the sense that if the kth approximant of the continued fraction is (p_k/q_k) the algorithm requires k operations (4×4 matrix multiplications) to determine $Y(q_k, p_k)$ ($= Y(p_k/g, p_k)$), and typically p_k increases exponentially with k (see below).

Quadratic irrational gradients give orbits with calculable fractal dimensions. The question naturally arises: can one find a fractal dimension for a generic irrational gradient? We are not able to answer this question. Numerical investigation suggests that a generic trajectory has dimension 1.5 (as does the path in spacetime of a one-dimensional Brownian motion). Two methods were used.

Firstly the product of the matrices (9) was formed using continued fraction elements a_k generated by iterating the continued fraction mapping (reciprocation and truncation) with an arbitrary initial number g. The chaos of this mapping quickly simulates a generic number irrespective of g. The results were consistent with $Y_k \sim \exp(k\pi^2/24\ln 2)$. Since the mapping is known ([10] page 75, [13] page 320) to give $p_k \sim \exp(k\pi^2/12\ln 2)$, this would imply $Y_k \sim \sqrt{p_k}$, and hence $D = 1.5$ (where $Y_k \sim p_k^{2-D}$).

Secondly, the numerical average, $\langle Y\rangle$, over x, of the displacement of a (straightened) trajectory from $(0, 0)$ to (x, y), is consistent with $\langle Y\rangle \sim \sqrt{y}$. This relation is equivalent to a tantalising purely mathematical one: that the integral over one period of the product of square waves,

$$\int sq(z)sq(2z)sq(3z)\ldots sq(pz)\,\mathrm{d}z \qquad \text{with} \quad sq(z) = (-1)^{[z]} \tag{10}$$

diminishes as $1/\sqrt{p}$ as $p \to \infty$. (This integral is easily verified to be the increment of $\langle Y \rangle$ as y increases across p.) The integrand changes sign at values of z which are rational fractions with denominator $\leqslant p$ ('Farey fractions' [14]), and it is possible, as was pointed out to us by J P Keating, that the power $\frac{1}{2}$ we seek is connected with the $\frac{1}{2}$ of the Riemann hypothesis via the Farey fractions [15].

If the generic motion is indeed a fractal of dimension 1.5, it must differ from a random walk in an important respect [6]. A random walk is chaotic in the sense that its information content is proportional to its length (one bit of information per step), whereas the information content of our trajectory increases only logarithmically with its length, as discussed above. The random Weierstrass–Mandelbrot function [7] of dimension 1.5 is an example of a fractal having this property.

There remains the question of generalisation. The unit barrier to gap length ratio played an important role in our analysis. Can one progress without it, perhaps to rational ratios? Or, more interesting, does our special system have any implications for pseudo-integrable systems in general? Finally, one might seek to quantise the system, either as a two-dimensional Bloch problem or, using the feature noted in the introduction, as a one-dimensional quantum map. The straightforward enumerability of the periodic orbits makes it attractive for semiclassical orbit quantisation, though the orbits would not account for corrections due to scattering, inherent in any pseudo-integrable system, from the singularities ('Sommerfeld scattering' from a barrier end [16]).

Appendix 1. Periodic orbits

In an orchard with a square lattice of trees one sees 'avenues' in every rational direction, the width of each being inversely proportional to the spacing of trees along it, so that the low-order rationals are most conspicuous. On our unit lattice such an avenue can be considered as a band of (straightened) trajectories with rational gradient. All except the edge members of this band miss all lattice points and represent true trajectories which perform identical zig-zags. We shall call these trajectories 'periodic orbits' because in the torus picture of the single unit cell, each closes and retraces itself periodically.

The periodic orbits of this system, as of any Hamiltonian dynamical system, merit study because non-periodic orbits can be approximated by them. They fill phase space with uniform density [17, 18] and form the only known basis for the semiclassical quantisation of non-integrable systems. Their classification is simple for our system, and since each band has, on its edge, lattice points of which some are equivalent to the origin, their study reduces to that of the previous section of rational orbits starting from the origin.

The phase space surface associated with each rational direction (q, p) is covered by one, two or three bands of periodic orbits. These cases have, respectively, $(q, p) =$ (odd, odd), (odd, even), or (even, odd). To see this we first construct a one-dimensional spatial section consisting of one barrier and one gap wrapped into a circle (figure 5). To follow a trajectory, numbered marks would be made on this circle indicating the successive positions where the trajectory hits any barrier or gap. The numbers correspond to the y coordinate, and the mark mapping is simply rotation by $\pi q/p$. In fact the marks are superfluous because each arc of angle π/p (starting from a barrier end) maps rigidly, so the numbers can refer to arcs rather than marks. For a full description

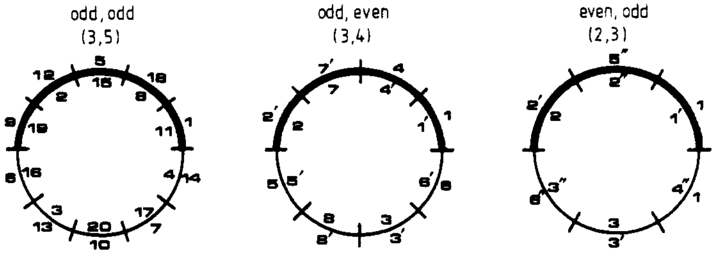

Figure 5. Periodic orbit bands displayed as numbered sequences of arcs mapping around a circle representing one barrier and one gap. Primed and double primed sequences represent the separate bands which are present if the gradient p/q of the orbits is not odd/odd. Numbers outside the circle indicate positive gradient, those inside indicate negative gradient.

one needs also to know the direction (positive or negative gradient) after the encounter, so we write the number outside the circle to indicate positive gradient, and inside for negative gradient. Thus any number on the gap semicircle is 'like' its predecessor (both outside or both inside), while any on the barrier semi-circle is unlike its predecessor.

The possibilities are, by symmetry, exhausted by the examples illustrated. For $(q, p) = (\text{odd}, \text{odd})$ the entire phase space sheet is covered by a single band of orbits with horizontal displacements $4q$. For (odd, even) there are two bands with $2q$. For (even, odd) there are three bands with horizontal displacements q, q and $2q$. All combinations have the same total displacement $4q$, as they must to cover the sheet fully. The total vertical displacement is, for the same reason, zero in each case: the $4q$ combination has zero vertical displacement, the two $2q$ ones have equal and opposite displacements $\pm(Y(q,p)+(-1)^{(q+1)(p+1)/2}Y(-q,p))$, and finally q and q have equal and opposites $\pm Y((-1)^{(q+1)(p+1)/2}q, p)$, while $2q$ has zero displacement.

Appendix 2. General reduction rule for gradients

Let A and B be general points in the plane (not lattice points). The straight line between them represents a straightened trajectory, and the product of the signs of the gradients at its ends is $(-1)^n$, where n is the number of barriers crossed by the straight line. This quantity can be easily and usefully related to an equivalent quantity $(-1)^{n'}$ for certain other lines $A'B'$ in the plane.

For example, if $A'B'$ is the reflection of AB in the diagonal mirror line $x = y$, then $n + n'$ can be considered as the number of barriers crossed by AB plus the number of *mirror image* (i.e. vertical) barriers it crosses. These two sets of barriers superimpose to form square enclosures in the plane, and since $n + n'$ is the number of entrances and exits from these, it is evidently even if A and B are both outside, or both inside these squares, and odd otherwise. The function $[x+1][y+1]$ is 1 inside and -1 outside such squares, so we have

$$\begin{aligned}(-1)^{n'} &= (-1)^n(-1)^{n+n'} \\ &= (-1)^n(-1)^{[x_A+1]}(-1)^{[x_B+1]}(-1)^{[y_A+1]}(-1)^{[y_B+1]}.\end{aligned}$$

Another example is horizontal shear: $x'_A = x_A + ay_A$, $y'_A = y_A$, and likewise for B, with a being an integer. If a is even, then the images of the barriers under this transformation are again barriers, so $n = n'$. If a is odd, then alternate rows of barriers are imaged to rows of barriers again, but the rows in between are imaged into gaps instead. The original barriers and their images then superimpose to form continuous horizontal lines at odd-integer heights and nothing at even-integer heights where the original barriers and their images cancel. Thus:

$$(-1)^{n'} = (-1)^n(-1)^{n+n'} = (-1)^n(-1)^{[y_A/2]}(-1)^{[y_B/2]}.$$

The transformations of these two examples taken in succession constitute the continued fraction transformation (3). The inverse transformation, with a chosen as the integer part of the gradient of AB (provided it is greater than unity) can be used for the systematic shortening of the separation AB until the gradients at either end are sure to be equal. Indeed, this is an alternative route to the equations (9) derived by the string construction. Its advantage is that it copes with trajectories which do not pass through the origin. Its disadvantage is that it supplies gradients, which need integrating to yield displacements.

Appendix 3. Phase space geometry: topology and interval exchange mappings

A typical trajectory in a (two degree of freedom) pseudo-integrable system explores a two-dimensional surface in its four-dimensional phase space which, however, is not a torus. It is easy to see that the topology of this surface for our system is a pair of tori connected by a 'neck'—i.e. a two-handled sphere (figure 6). The neck corresponds to the barrier which converts upward to downward motion and vice versa, as indicated

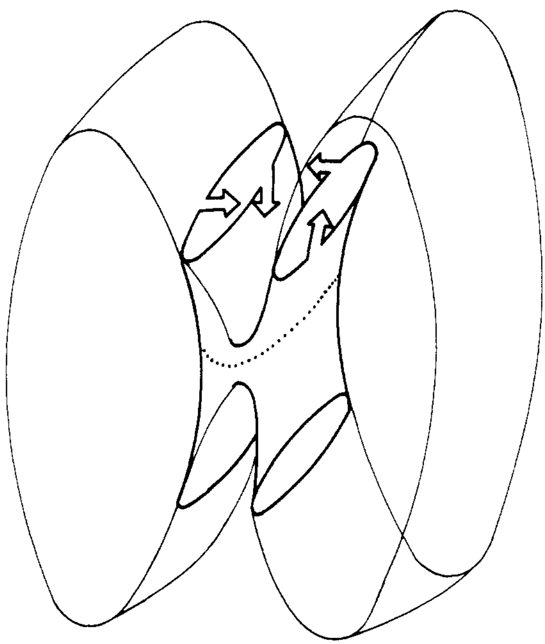

Figure 6. The double torus, or two-handled sphere surface explored by a trajectory in the barrier billiard. The neck joining the tori represents the barrier, and the dotted lines, the gap. Arrows show the directions of motion on the sheets, which are related indirectly to the directions on the plane.

by the arrows. It does not affect motion on the other two sheets of the tori, which correspond to passage through a gap (dotted), upward or downward. There is a simple rule for the number of handles of the phase space surface for motion in rational polygon enclosures [19].

The description of the motion as an interval exchange mapping is illustrated in figure 7. Here a vertical unit with an initially positive gradient (shaded strip) and one with an initially negative gradient (unshaded strip) are mapped by following their motion two units to the right. One section (interval) of the shaded strip has shifted into the unshaded area, indicating conversion to negative gradient, and vice versa. Sections which have escaped from the vertical unit have been shifted to an equivalent position within it. Such mappings where intervals are shifted so that their order is changed (interval exchange mappings) have been extensively studied [20]. They are known to be ergodic, indeed 'weak mixing'. This is expressed by:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} |\mu(I^j A \cap B) - \mu(A)\mu(B)| = 0$$

where A and B are arbitrary initial intervals with lengths $\mu(A)$ and $\mu(B)$, and $\mu(I^j A \cap B)$ is the length of the overlap of B with the image of A under j iterations (lots of little pieces as $j\to\infty$). If the mapping were mixing the summand itself would tend to zero as $j\to\infty$, meaning that the little pieces become uniformly spread out. The weak mixing condition allows occasional reclustering with diminishing frequency.

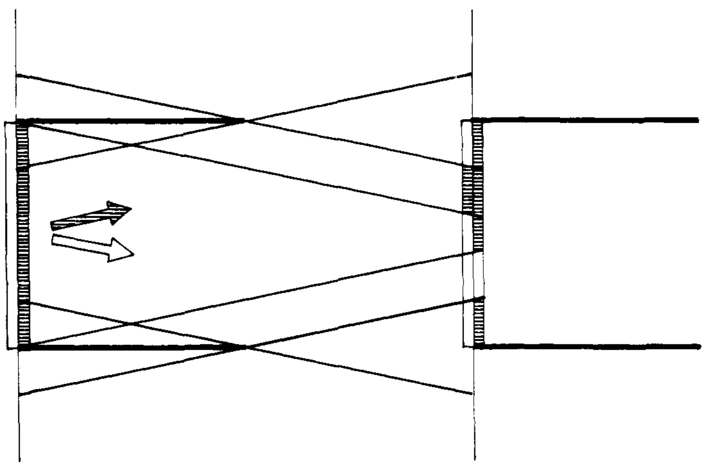

Figure 7. Interval exchange displayed as a mapping of unit vertical double strips (intervals) by horizontal motion (it is vice-versa in figure 5). Sections of positive gradient strip (shaded) are shifted into the negative gradient area by a bounce on a barrier (and vice-versa). Sections escaping from the unit vertical are considered restored to equivalent positions within it.

Appendix 4. Displacement recursion matrix reduced from 4×4 to 2×2

As indicated in the main text, only two of the four quantities Y_k^+, Y_k^-, Y_{k-1}^+, Y_{k-1}^- are independent. All four can be expressed in terms of Y_k^+, Y_{k-1}^+ if any one of the pairs

(q_k, p_k), (q_{k-1}, p_{k-1}), $(q_k+q_{k-1}, p_k+p_{k-1})$ is of the form $(3,0)$ mod 4, or $(3,2)$ mod 4. Otherwise (that is, if any one is of the form $(1,0)$ mod 4 or $(1,2)$ mod 4) all four are expressible in terms of Y_k^-, Y_{k-1}^-. If σ_k is defined as 1 for the former case, and -1 for the latter case, then

$$\begin{pmatrix} Y_{k+1}^{\sigma_{k+1}} \\ Y_k^{\sigma_{k+1}} \end{pmatrix} = \mathbf{M} \begin{pmatrix} Y_k^{\sigma_k} \\ Y_{k-1}^{\sigma_k} \end{pmatrix}$$

where $\mathbf{M}$ is given as follows:

for p_k even and q_{k-1} even $\quad \mathbf{M} = \begin{pmatrix} [(a_k+1)/2] & 1 \\ 1 & 0 \end{pmatrix}$

for p_k even and q_{k-1} odd $\quad \mathbf{M} = \begin{pmatrix} [(a_k/2)] & 1 \\ 1 & 0 \end{pmatrix}$

for p_{k-1} even and q_k even $\quad \mathbf{M} = \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix}$

for p_{k-1} even and q_k odd, with:

$a_k = 0 \bmod 4 \quad \mathbf{M} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

$a_k = 1 \bmod 4 \quad \mathbf{M} = \xi_k \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}$

$a_k = 2 \bmod 4 \quad \mathbf{M} = \xi_k \begin{pmatrix} 2 & -1 \\ 1 & -1 \end{pmatrix}$

$a_k = 3 \bmod 4 \quad \mathbf{M} = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}$

where $\xi_k = (-1)^{(p_k+p_{k-1}-1)/2}$

for (p_k+p_{k-1}) even and q_k even $\quad \mathbf{M} = \begin{pmatrix} a_k & 1 \\ 1 & 1 \end{pmatrix}$

for (p_k+p_{k-1}) even and q_k odd, with:

$a_k = 0 \bmod 4 \quad \mathbf{M} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

$a_k = 1 \bmod 4 \quad \mathbf{M} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$

$a_k = 2 \bmod 4 \quad \mathbf{M} = \eta_k \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$

$a_k = 3 \bmod 4 \quad \mathbf{M} = \eta_k \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix}$

where $\eta_k = (-1)^{(p_{k-1}-1)/2}$.

References

[1] Arnold V I 1978 *Mathematical Methods of Classical Mechanics* (Berlin: Springer)
[2] Sinai Ya G 1976 *Introduction to Ergodic Theory* (Princeton, NJ: Princeton University Press)
[3] Sinai Ya G 1970 *Russ. Math. Surv.* **25** 137
[4] Richens P J and Berry M V 1981 *Physica* **2D** 495
[5] Zemlyakov A N and Katok A B 1976 *Math. Notes* **18** 760
[6] Eckhardt B, Ford J and Vivaldi F 1984 *Physica* **13D** 339
[7] Berry M V and Lewis Z V 1980 *Proc. R. Soc.* A **370** 459
[8] Richens P J 1982 *PhD thesis* University of Bristol
[9] Heyney F S and Pomphrey N 1982 *Physica* **6D** 78
[10] Khinchin A Ya 1964 *Continued Fractions* (Chicago, IL: University of Chicago Press)
[11] Klein F 1939 *Geometry* (New York: Dover)
[12] Arnold V I 1983 *Geometrical Methods in the Theory of Ordinary Differential Equations* (Berlin: Springer)
[13] Levy P 1937 *Theorie de l'Addition des Variables Aleatoires* (Paris: Gauthier Villars)
[14] Schroeder M R 1984 *Number Theory in Science and Communication* (Berlin: Springer)
[15] Edwards H M 1974 *Riemann's Zeta Function* (New York: Academic)
[16] Sommerfeld A 1954 *Optics* (New York: Academic)
[17] Hannay J H and Ozorio de Almeida A M 1984 *J. Phys. A: Math. Gen.* **17** 3429
[18] Ozorio de Almeida A M 1988 *Hamiltonian Systems, Chaos and Quantization* (Cambridge: Cambridge University Press)
[19] Gutkin E 1986 *Physica* **19D** 311
[20] Veech W A 1982 *Ann. Math.* **115** 201