TOPICS IN MODERN PHYSICS

A Tribute to Edward U. Condon

Edited by Wesley E. Brittin and Halis Odabasi

COLORADO ASSOCIATED UNIVERSITY PRESS
BOULDER, COLORADO 1971

A PROBLEM IN THE STABILITY OF PERIODIC SYSTEMS

EDWIN M. MC MILLAN
University of California
Berkeley, California

In the Spring of 1967 I attended a theoretical seminar at which Professor René de Vogelaere spoke concerning the stability of non-linear periodic systems. The motivation was storage rings, with beams focused by azimuthally varying fields ("strong focusing"); the question, the effect of non-linear terms on an otherwise stable system; the presentation I found utterly fascinating. It recalled another seminar I attended at Princeton* over a third of a century earlier, at which G. D. Birkhoff discussed the stability of the solar system. I remember none of the detail of that earlier seminar, but I have a strong memory of how an apparently simple situation led rapidly and unavoidably into a maze of complexity, leaving the original question "Is the motion of the system stable for infinite time?" unanswered.

The common feature of the two situations is periodicity, the periodic perturbation of one planet by another, or the periodic variation of the focusing fields as seen by a circulating electron or proton. Because of the repetitive nature of the situation, the solution for many periods of the motion can be pieced together out of the general solution for a single period. If x, y represent the values of a coordinate and its conjugate momentum (coupling with other degrees of freedom is not included in this treatment) while x', y' are the values after one period, the general solution can be written:

$$\left.\begin{aligned} x' &= x'(x, y) \\ y' &= y'(x, y) \end{aligned}\right\} \tag{1}$$

where the Jacobian $(\partial x'/\partial x)(\partial y'/\partial y) - (\partial x'/\partial y)(\partial y'/\partial x)$ must be equal to unity if the equation of motion is derivable from a Hamiltonian. These equations describe a mapping in a plane, which moves point

* Where Professor Condon supervised the final year of my thesis work, on a problem in experimental physics.

x, y to point x', y', and whose repeated iteration depicts the motion of the system.

The unit Jacobian ensures that the mapping is area-preserving, that is, it carries a closed curve over into one of equal area. It may happen that a curve transforms into itself; it is then called an invariant curve. The points inside a closed invariant curve will remain inside, and the stability question is answered for such points; however they may move about as the transformation is repeatedly iterated, they cannot escape through the invariant boundary, and the motion that they represent is stable. Points not so constrained may go rapidly to infinity, representing an unstable motion, or they may simply wander about in an apparently aimless fashion, but with no clear tendency to get very far away. They may seem to be stable, but stability after an infinite number of iterations cannot be proved. These are the difficult cases that make the subject so fascinating.

Professor de Vogelaere was discussing the transformation:

$$\left.\begin{aligned} x' &= y + x^2 \\ y' &= -x + x'^2 = -x + (y + x^2)^2 \end{aligned}\right\}. \tag{2}$$

This was chosen because the linear terms taken alone give a stable motion, the non-linear terms are simple in form, and the transformation (using the first form given on the right side of the second equation) is particularly suited for iteration on a digital computer. It has two fixed points (points for which $x' = x$, $y' = y$), one at the origin, and one at $x = 1$, $y = 0$. Near the origin the motion approaches that given by the linear terms alone, which is stable; this is called a stable fixed point. The other is an unstable fixed point.

Before carrying the discussion farther, we should digress to consider the linear case, whose solution is well known. The most general area-preserving linear transformation about a fixed point is:

$$\left.\begin{aligned} x' &= ax + by \\ y' &= cx + dy \end{aligned}\right\} \tag{3}$$

$$\text{with } ad - bc = 1.$$

The stability is determined by the trace $(a + d)$ of the transformation matrix. If $(a + d)/2$ lies between -1 and $+1$, the result of n iterations of the transformation is:

$$\left.\begin{aligned} x_n &= x_0 \cos \mu n + \left[\frac{1}{2}(a - d)x_0 + by_0\right] \frac{\sin \mu n}{\sin \mu} \\ y_n &= y_0 \cos \mu n + \left[\frac{1}{2}(d - a)y_0 + cx_0\right] \frac{\sin \mu n}{\sin \mu} \end{aligned}\right\} \tag{4}$$

$$\text{with } \cos \mu = \frac{1}{2}(a + d).$$

If $(a + d)/2 > 1$, cos and sin are replaced by cosh and sinh in the above equations; if $(a + d)/2 < -1$, the same substitutions are made, and in addition $\cosh \mu$ and $\sinh \mu$ are multipled by -1, $\cosh \mu n$ and $\sinh \mu n$ by $(-1)^n$. The trigonometric functions show stability, the hyperbolic functions instability.

A relation of the form $f(x, y) = \text{const.}$ defines an invariant curve if $f(x', y') = f(x, y)$ for all values of x and y lying on the curve. It is easy to verify by direct substitution that:

$$-cx^2 + by^2 + (a - d)xy = \text{const.} \tag{5}$$

is an invariant curve for the transformation (3). These curves are ellipses for $-1 < (a + d)/2 < 1$, otherwise they are hyperbolas. In the stable case, successive iterations move a point around an elliptic path. If μ is a rational multiple of π, the point returns to its original position after some number of iterations; such points can be called fixed points of the corresponding order, e.g., if $a + d = 0$, $\mu = \pi/2$ and every point is a fixed point of the fourth order. If μ is not a rational multiple of π, the ellipse is eventually densely covered and can be visualized by simply plotting the results of many iterations. In the unstable cases, a point moves along one branch of a hyperbola (or alternates between the two branches if $(a + d)/2 < -1$), but it does this only once; it does not return from infinity. A given starting point, iterated forward (and backward using the inverse transformation) generates only the skeleton of a curve. We, with superior knowledge of the algebraic properties of the transformation, know that the hyperbola is an invariant; a computer blindly iterating the transformation does not know this and gives us only a disconnected set of points. When the non-linear part of the transformation is included, our superior knowledge vanishes, and we are left in a quandary; are there, or are there not, invariant curves?

The asymptotes of the hyperbolas constitute a privileged class. In these directions the hyperbolic functions combine to give single positive or negative exponentials. Successive iterations (forward along one pair of opposite directions, backward along the other) lead to an increasingly dense set of points as the fixed point is approached. Reversing the direction of iteration, a short densely occupied line segment in the neighborhood of the fixed point generates the whole asymptote. This remains true even when non-linear terms are added, and the asymptotic line leads into a curve. The initial direction is determined from the linearized form of the transformation, and the computer can then be turned loose to extend the curve, until one runs out of patience or precision. The initial directions can be found from Equation (5), or directly from the transformation, noting that along an asymptote $y'/x' = y/x$. They are:

$$\frac{dy}{dx} = \frac{1}{2b}(d - a \pm \sqrt{(a + d)^2 - 4}) \tag{6}$$

with points moving outward along the directions with the plus sign, inward along directions with the minus sign.

We can now return to the transformation (2) and the presentation of de Vogelaere. Differentiation about the point 1, 0 gives $a = 2$, $b = 1$, $c = 3$, $d = 2$; we then get $dy/dx = \pm\sqrt{3}$, and the initial directions are at 60° to the x-axis. When extended by computer, the lines to the right of the y-axis go harmlessly to infinity. Those to the left start to bend toward one another, as if trying to surround the stable fixed point at 0, 0, and at about -0.55 on the x-axis they intersect. This has great consequences. A given invariant curve cannot cross itself; this would imply that the transformation can go two different ways from a single point, the intersection, but the transformation is single-valued. However, two different invariant curves can cross, and when they do they are caught up in an endless dance. The intersection must transform to a single point, which lies on both curves, so there must be another intersection, and so on. The points on one curve are approaching the fixed point, and the intersections come closer and closer, with the other curve behaving like a wildly meandering river. We can call the region enclosed by the two curves up to their first crossing the "inside." The law of preservation of areas requires that all outside loops have equal area, and all inside loops have equal area. (The symmetry of the transformation, which we have not discussed, requires that these two areas are also equal to one another.) Outside loops become long, thin tentacles reaching toward infinity; inside loops become "worms" curled up in the interior. The development of this pattern, and the ingenuity displayed by the "worms" in avoiding contact with others of the same family, are fascinating to watch.

What does this have to do with stability? If the two invariant curves had joined smoothly on their first meeting*, they would have made a single closed invariant surrounding the stable fixed point, and all points inside would have been stable. (This does not violate the law against crossing; curves do not cross the unstable fixed point, they merely approach it.) Since this did not happen, every inside point is in possible jeopardy of escaping. But it seems that the "worms" approach the stable fixed point with exceeding slowness, implying that points

* "Joining smoothly" implies more than coincidence of slopes at one point. It requires superposition over a length sufficient to preclude the formation of loops, e.g. between points D and E in Fig. 2.

started near 0, 0 will work their way out, if they do, with corresponding slowness. Thus there is a region of ambiguity, in which both computer precision and patience fail, and in which there seems to be no means for decision. This is the fascination of the stability problem, which in one form or another has occupied mathematicians for many decades. I shall not go into further ramifications, such as the chains of "islands" formed by invariant curves of higher orders—the complexity that can be derived from the simple transformation (2) does not need more emphasis.

At this point the reader may wonder why I wrote this. Is it intended to be an elementary treatise on stability theory? No, but I felt that some background was necessary to introduce my attempt to contribute something to the subject. This arose from a remark at the seminar, that there was no known case of a non-linear area-preserving transformation with a finite region of guaranteed stability. This may have been an overstatement, but I took it as a challenge to find at least one case (not counting those that can be constructed by treating weak-focusing problems by strong-focusing methods, in which conservation of energy can be invoked). The first step was to choose a simple form for the transformation, which contains one non-linear function of one variable in one of the pair of equations, and which has unit Jacobian regardless of the form of the function. The form chosen was:

$$\left.\begin{aligned} x' &= y \\ y' &= -x + f(y) \end{aligned}\right\} \quad \text{with the inverse*:} \quad \left.\begin{aligned} x' &= -y + f(x) \\ y' &= x \end{aligned}\right\}. \tag{7}$$

This is more general than it may appear at first. For example, the transformation (2) can be converted to (7), with $f(y) = 2y^2$, by the change of coordinates $X = x$, $Y + X^2 = y$, the capital letters representing the variables in (2). Such a coordinate change, while it affects the appearance of a mapping, does not change its topology, which is what we really want to find. In representing physical systems, (7) is adequate insofar as the non-linear behavior can be treated as lumped at one place in each period. Let the basic linear system be represented by (3) (with capital letters), and interpolate just before the observation plane a thin lens, defined as an element which adds a slope $F(X)$ to a path which traverses it at a displacement X. At the lens, the displacement is X', and the change in slope is to be added to Y', giving:

$$\left.\begin{aligned} X' &= aX + bY \\ Y' &= cX + dY + F(X') \end{aligned}\right\}. \tag{8}$$

Now make the coordinate transformation $X = x$, $aX + bY = y$, with the result:

* The inverse transformation is here defined as the result of inverting the transformation, then interchanging the primed and unprimed quantities.

$$\left.\begin{aligned} x' &= y \\ y' &= -x + (a+d)y + bF(y) \end{aligned}\right\}. \tag{9}$$

Thus the function $f(y)$ in the form (7) contains a linear part $(a + d)y$ arising from the basic linear system, in addition to the thin-lens function $F(y)$, which may, of course, also have a linear term. For a check, consider the case of free motion, with $a = 1$, $d = 1$, and $F(y) = 0$. The function $f(y) = 2y$, and the invariants in the x, y plane are straight lines of unit slope, which convert to lines of zero slope in the X, Y plane, as expected. If $f(y) = 0$, the system is in the center of the region of stability, and the transformation is a clockwise rotation by 90°.

We must get back to our original task, to find functions $f(y)$ such that the transformation (7) gives a closed invariant boundary, ensuring the stability of points lying within the boundary. The fixed points, if any, lie at the solutions of the equations $x = y$, $f(y) = 2y$, where $x' = x$, $y' = y$. A fixed point is stable if $-2 < df(y)/dy < 2$, otherwise it is unstable. One could presumably carry out a computational program, following invariant curves from their beginnings at unstable fixed points, and adjusting parameters in $f(y)$ until they join smoothly, by which closed boundaries could be constructed. This is not what I mean; I am searching for cases which can be solved by analytical or geometrical arguments. The direct approach, starting with a given function $f(y)$, seems to be intractable. We therefore try an indirect approach: start with a given function $x = \phi(y)$, and require that it be an invariant under the transformation (7). This is the same as saying that if $x = \phi(y)$, then $x' = \phi(y')$, or in terms of the inverse function ϕ^{-1}, $y' = \phi^{-1}(x')$. With these substitutions, the second equation of (7) becomes:

$$\phi^{-1}(x') = -\phi(y) + f(y). \tag{10}$$

The first equation of (7) allows us to replace x' by y in (10), giving:

$$f(y) = \phi(y) + \phi^{-1}(y). \tag{11}$$

This is a remarkable result, of startling simplicity, which fell out almost without effort on my part. It leads to not just one, but to great families of functions $f(y)$ giving regions of stability. It also has a simple geometrical interpretation. The curve $x = \phi^{-1}(y)$ is the mirror image of $x = \phi(y)$ in the diagonal axis $y = x$. For, $x = \phi^{-1}(y)$ is equivalent to $y = \phi(x)$, which is derivable from $x = \phi(y)$ by simply interchanging the roles of x and y, which is the same as making the reflection about the diagonal. Any pair of curves related by this reflection symmetry can be invariants (with some restrictions that will be mentioned in a moment), with the appropriate function $f(y)$ being the sum of the two, treated as functions of y. It is clear from the symmetry that one could just as well think of the curves as functions of x, and derive the same function $f(x)$. If, for each value of x, the midpoint between the two curves is plotted, the resulting curve is $\frac{1}{2}f(x)$. Intersections of this curve with the diagonal axis are fixed points, and those at which the two symmetrically placed invariants also intersect are unstable. This curve is also the basis of a second reflection symmetry. The second symmetry is that of reflection parallel to the y-axis in the curve $y = \frac{1}{2}f(x)$, or parallel to the x-axis in the curve $x = \frac{1}{2}f(y)$; the first symmetry makes these operations equivalent.

One can also start with a single function ϕ that describes a closed curve with reflection symmetry about the diagonal. This function is its own inverse, that is, the same functional form is used in giving y as a function of x or x as a function of y. The sum in (11) is then the sum of the two values of the function, if it is a double-valued function. If curves are drawn that have more than two values, there may be trouble. The requirement is that the values can be taken in pairs whose sums are equal. This is a condition on the drawing of the curves; one can no longer use any curve with diagonal symmetry. Some cases of this kind are easily constructed, but the flexibility is lost.

The two-fold symmetry that we have found in the invariant curves should correspond to a two-fold symmetry in the transformation itself, and this is easily seen to be the case. Consider first the reflection symmetry about the diagonal $x = y$. This reflection is performed by interchanging the coordinates x and y. If this is done in the transformation (7), the inverse transformation is obtained. This means that every feature of the mapping is repeated as a mirror image reflected in the diagonal. If an invariant curve has this symmetry already, its image is itself, otherwise invariants come in symmetrical pairs.

The reflection corresponding to the second symmetry can be performed by leaving x unchanged, and replacing y by $f(x) - y$. This also converts the direct transformation (7) to its inverse, and the same statements made about the first symmetry can be repeated about the second. In fact, the transformation (7) is equivalent to the two reflections:

$$\left.\begin{aligned} x' &= y \\ y' &= x \end{aligned}\right\} \tag{12}$$

and

$$\left.\begin{aligned} x' &= x \\ y' &= f(x) - y \end{aligned}\right\} \tag{13}$$

performed in sequence, while the inverse is obtained by reversing the sequence. Each of the reflections has the determinant -1, but their product has the determinant unity as required.

We can now look at some examples of curves with this two-fold symmetry. The simplest case is that in which $f(y) = 2ky$, making the transformation linear. In the notation of (3), $a = 0$, $b = 1$, $c = -1$, $d = 2k$, and the invariant curves are:

$$x^2 + y^2 - 2kxy = \text{const.} \tag{14}$$

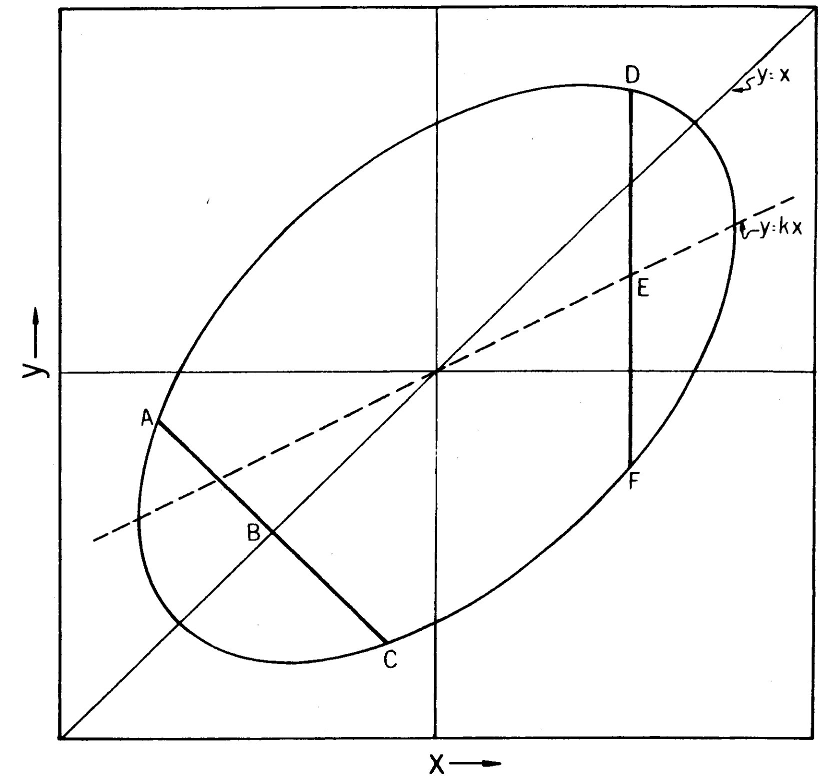

Fig. 1. The ellipse $x^2 + y^2 - 2kxy = \text{const.}$ is an invariant curve under the transformation $x' = y$, $y' = -x + 2ky$. It is used here to show the twofold reflection symmetry discussed in the text. The first symmetry operation is a mirror reflection in the diagonal $x = y$, illustrated by the equality $AB = BC$. The second is a reflection parallel to the y-axis about the line $y = kx$, illustrated by $DE = EF$.

The quantity μ of Eq. (4) is given by $\cos \mu = k$, and if μ is a rational multiple of π the iteration of the transformation will generate only a discrete set of points lying on the invariant curve. Thus, if $k = 0$, the transformation will give not a circle but four points equally spaced around a circle.

These are parabolas or hyperbolas with their principal axes at 45° to the x-axis, and the first symmetry is obvious. The second symmetry is seen by drawing a set of parallel chords, parallel to the y-axis, and connecting their centers. These lie on a straight line of slope k, and the second symmetry is thereby made visible (see Figure 1).

The simplest non-linear case, at least in functional form, is that in which $f(y)$ is quadratic. The transformation (2) used by de Vogelaere, as we have pointed out, can be changed into (7) with $f(y) = 2y^2$. What was not obvious in the beginning is that the change of coordinates introduces a new symmetry that did not appear in the original coordinate system, the symmetry about the diagonal; the "second symmetry" was already present in the form of a reflection about the x-axis. In this case the invariants do not join smoothly, so there are two complete sets of invariant curves which are mirror images of one another, and they are images in two different ways, because of the two-fold reflection symmetry. This is illustrated in Figure 2. In Figure 3 is shown a further extension of one of the curves, with "tentacles" and "worms." Around

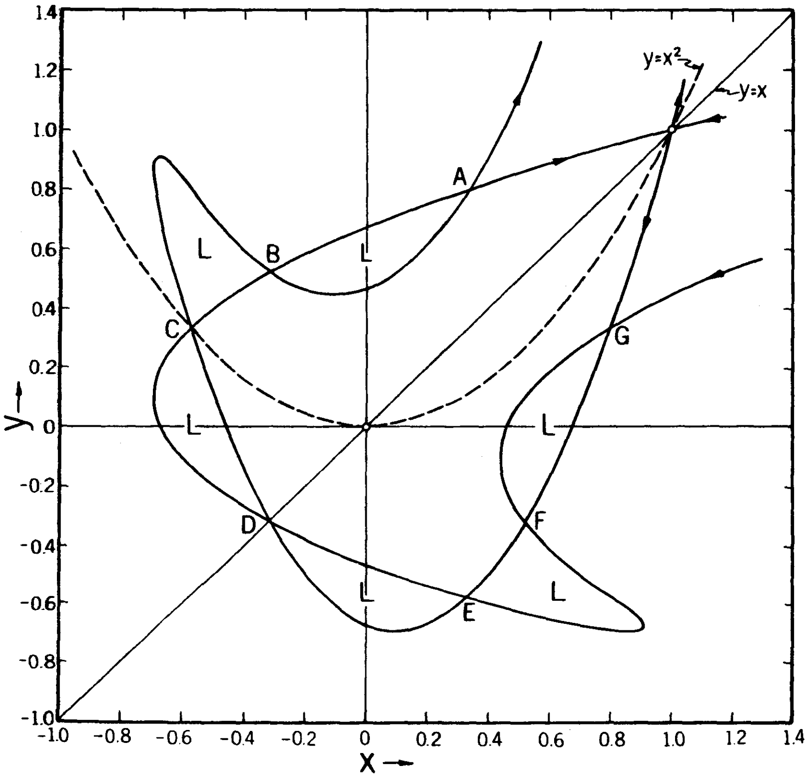

Fig. 2. Initial portions of invariants under the transformation $x' = y$, $y' = -x + 2y^2$, leading away from (or toward) the unstable fixed point at $x = 1$, $y = 1$. The arrows indicate the directions in which points are moved by the transformation. The point pairs AG, BF, CE illustrate the first symmetry, the point pairs AE, BD the second symmetry. The areas of the loops marked L are all equal.

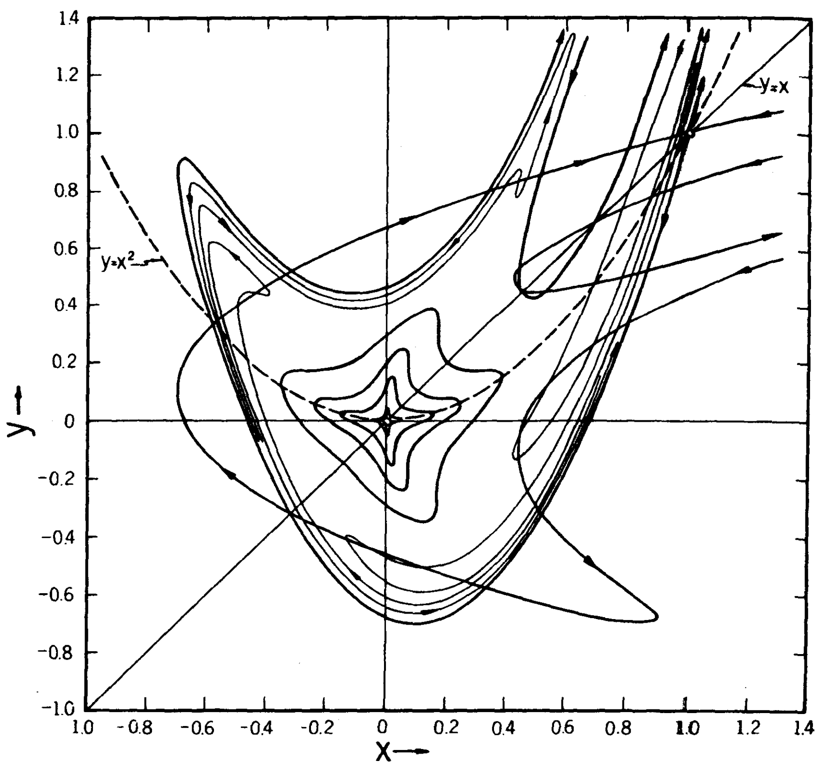

Fig. 3. A partial extension of the curves of Fig. 2, showing "tentacles" reaching off the figure and "worms" in the interior. Since the entrance channel for the "worms" becomes very narrow, the figure becomes difficult to draw completely as the iteration progresses.

Some apparently closed curves around the stable fixed point at $x = 0$, $y = 0$ are also shown. The peculiar behavior near the origin seems less mysterious if one recalls that the function $y = \frac{1}{2}f(x)$ approaches the limit of zero slope, where the curve degenerates to four points, and where the slightest perturbation can cause a slow migration about the center and a concomitant slow change in radius.

the origin are shown some apparently closed curves generated by starting at points on the diagonal and iterating the transformation many times, then joining the resulting points. Are these really closed curves? Do they represent motions that are stable for an infinite time? We don't know.

Some simple cases with families of closed invariants can be constructed in the following way. First write down an equation for an invariant curve, then solve it for y; if there are just two values for y, their sum must be $f(x)$. The first symmetry is satisfied if x and y occur in a symmetrical fashion in the equation, and there will be just two values for y if the equation is quadratic in y. The most general equation with these properties is:

$$Ax^2y^2 + B(x^2y + xy^2) + C(x^2 + y^2) + Dxy = \text{const.} \tag{15}$$

(A possible term in $x + y$ has been omitted since it can be removed by a translation of the coordinates along the diagonal.)

The function $f(x)$ is then:

$$f(x) = -\frac{Bx^2 + Dx}{Ax^2 + Bx + C}. \tag{16}$$

This function, divided by two, forms the curve about which the second symmetry exists for the curve represented by (15); the same function, with x replaced by y and put into (7), gives the transformation under which (15) is an invariant. Note that (16) does not depend on the constant on the right side of (15), so that a single transformation can generate the whole family (15) of invariants.

The number of parameters in (16) can be reduced to two (or less in some cases) since only the ratios of coefficients are significant, and one of the ratios can be adjusted by a change of scale of the coordinates. Some examples are given below:

(a) $$f(x) = \frac{2kx}{1 + x^2}, \text{ with the invariants:} \tag{17}$$

$$x^2y^2 + x^2 + y^2 - 2kxy = \text{const.} \tag{18}$$

If $-1 < k < 1$, there is a stable fixed point at the origin, and all the invariants are smooth closed curves (Figure 4). If $k > 1$, the fixed point at the origin is unstable, and stable fixed points occur at $x = y = \pm\sqrt{k - 1}$. A figure-eight shaped separatrix separates the plane into three regions, which are all stable, in spite of the unstable fixed point; points which move away from its vicinity always return along a closed path (Figure 5).

(b) $$f(x) = \frac{2kx}{1 - x^2}, \text{ with the invariants:} \tag{19}$$

$$x^2y^2 - x^2 - y^2 + 2kxy = \text{const.} \tag{20}$$

If $|k| > 1$, the system is unstable. If $0 < k < 1$, there is a stable fixed point at the origin and unstable fixed points at $x = y = \pm\sqrt{1 - k}$. The two rectangular hyperbolas $(x + 1)(y - 1) = -k$ and $(x - 1)(y + 1) =$

$-k$ are separatrices; stable invariants fill the region between these two curves, and unstable invariants lie outside (Figure 6).

(c) $$f(x) = \frac{x^2}{1-x}, \text{ with the invariants:} \tag{21}$$

$$x^2y + xy^2 - x^2 - y^2 = \text{const.} \tag{22}$$

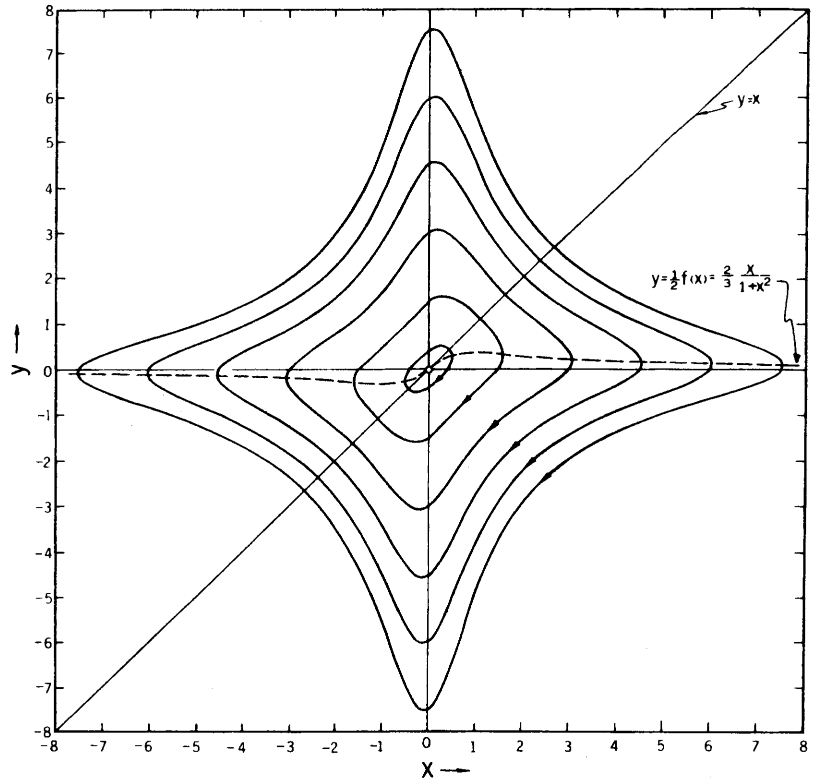

Fig. 4. The transformation $x' = y, y' = -x + f(y)$, with $f(y) = 2ky/(1 + y^2)$, has the family of invariant curves $x^2y^2 + x^2 + y^2 - 2kx = \text{const.}$ Some members of this family, for the case $k = \frac{2}{3}$, are shown. See (a) in text. The function $f(x)$ plotted in the figure is the same as the function $f(y)$ occurring in the transformation, with y replaced by x. Some may find it more convenient to use the inverse transformation, in which $f(x)$ appears directly. If this is done, the figures are unchanged except for the direction of the arrowheads.

There is a stable fixed point at the origin, and an unstable fixed point at $x = y = \frac{2}{3}$. There are four separatrices, the line $x + y + 2 = 0$, the two branches of the hyperbola $(x - 1)(y - 1) = -1$, and the curve given by setting the right side of (22) equal to $-8/27$. The only stable area is inside the loop of that curve (Figure 7).

In cases (a) and (b) it will be noticed that there are ranges of the parameter k in which the behavior was not specified. These were omitted in order to avoid a diversion into the matter of second order fixed points and invariants, but the reader may wish to make his own

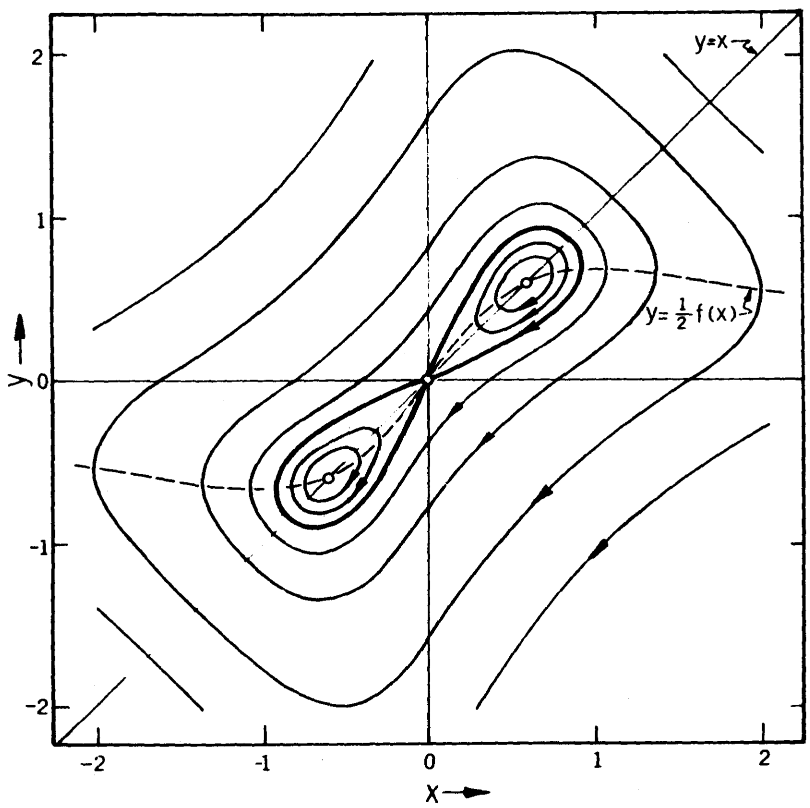

Fig. 5. Same as Fig. 4, for the case $k = 1.36$, plotted on a larger scale. The fixed point at the origin, stable in Fig. 4, has become unstable, and stable fixed points have appeared at $x = y = \pm 0.6$, surrounded by a figure-eight-shaped separatrix, given by Eq. (18) with the constant set equal to zero. Separatrices in this and following figures are drawn with heavy lines. See (a) in text.

investigation, and will find that it is not too difficult. In these omitted ranges the mappings look like those in Figures 5 and 6 rotated through 90°, and there are fixed points of the second order on the diagonal $y = -x$. The transformation carries one of these points into the other, and the next iteration returns it to the original point. The curves inside the two closed loops in the rotated Figure 5 are described in an inter-

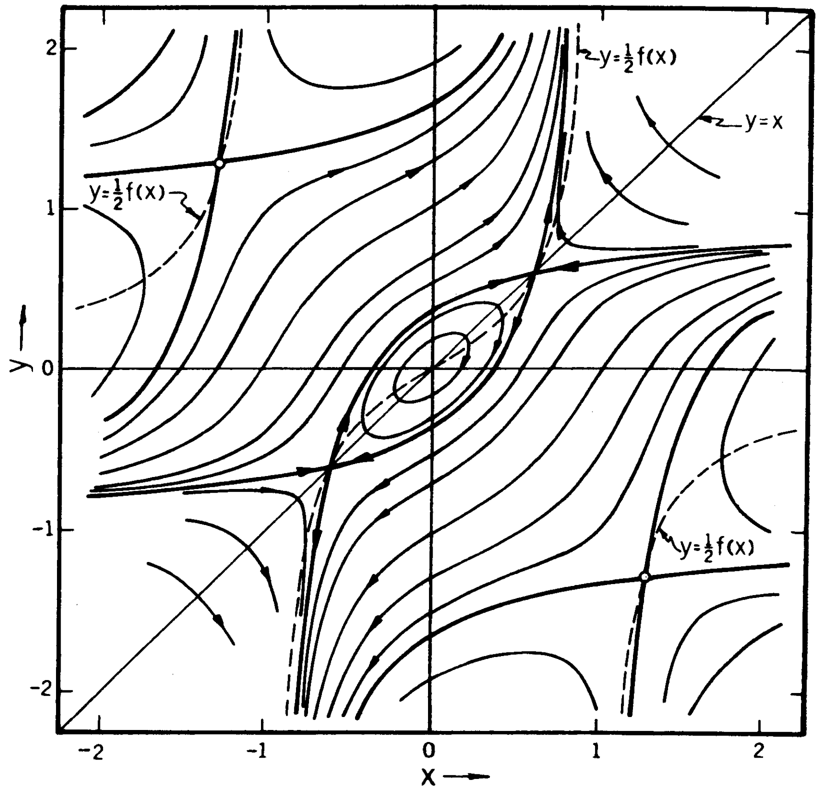

Fig. 6. The transformation $x' = y, y' = -x + f(y)$, with $f(y) = 2ky/(1 - y^2)$, has the family of invariant curves $x^2y^2 - x^2 - y^2 + 2kxy = \text{const}$. Some members of this family, for the case $k = 0.64$, are shown. The stable and unstable fixed points of Fig. 5 have here traded places, and the two separatrices mentioned in the text under (b) intersect at the unstable fixed points and bracket the central stable area. In addition, there are four separatrices consisting of two branches each of the rectangular hyperbolas $(x + 1)(y + 1) = -k$ and $(x - 1)(y - 1) = -k$, and two unstable fixed points of the second order at $x = -y = \pm\sqrt{1 + k}$. These appear near the lower right and upper left corners of the figure. Arrowheads are omitted from some of the curves, because the transformation moves one second order fixed point to the other, and causes points on associated curves to move to other branches. Arrowheads would have meaning only if correlated with the result of two iterations of the transformation, which returns a point to its original branch.

esting way, by a point which alternates between the interiors of the two loops.

The final examples I shall give are derived directly from Eq. (11), in that one draws any curve $y = \phi(x)$ and its inverse $\phi^{-1}(x)$, and then derives the function $y = f(x)$ which will make them invariants under the transformation (7). The only restrictions on ϕ and ϕ^{-1} are that they be continuous, and that considered together they form a double-valued function when solved for y (or x). Their derivatives need not

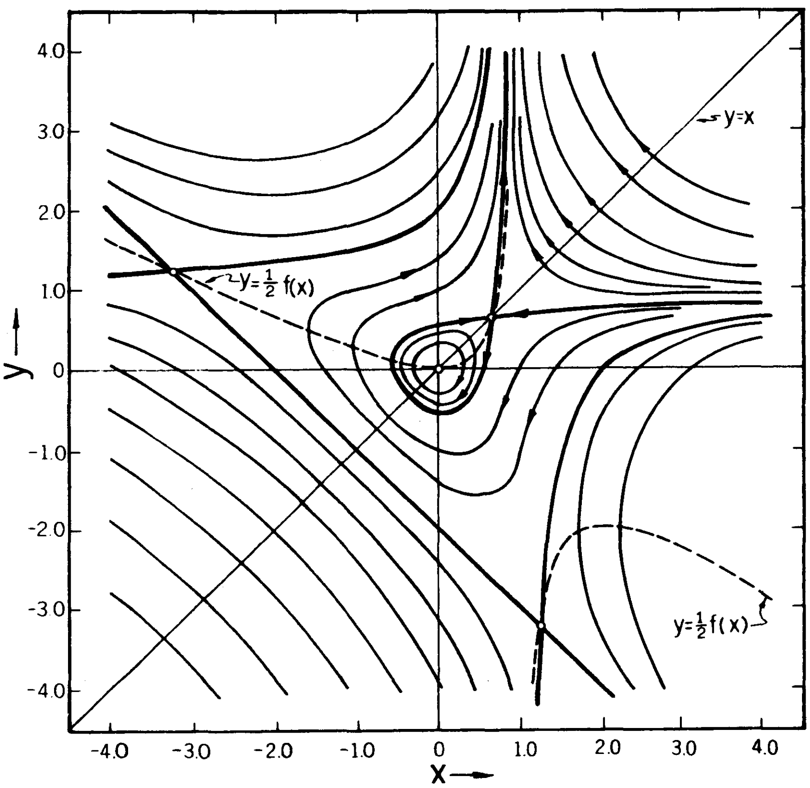

Fig. 7. Another family of invariant curves, with $f(x) = x^2/(1 - x)$. (Note that $f(y)$ in the transformation is the same function, with x replaced by y.) See (c) in text. This function approaches x^2 at the origin, and the separatrix forming a loop about the stable fixed point can be compared with the "wormy" curves of Figs. 2 and 3. Note also the absence of the four-pointed stars around the origin. The outlying separatrices intersect in fixed points of the second order.

be continuous; functions made of straight line segments are acceptable. The most obvious cases are those in which the two curves $\phi(x)$ and $\phi^{-1}(x)$ meet in two unstable fixed points on the diagonal $y = x$, as in Figures 6 and 8, but equally valid cases are given by closed curves containing one unstable fixed point (Figure 5 and the central loop of Figure 7) and by smooth closed curves (Figures 1 and 9). As an example of how one can construct a boundary, consider $\phi(x) = -x + 2x^2$, $\phi^{-1}(x) = \frac{1}{4}(1 \pm \sqrt{1 + 8x})$. For $x < 0$, $f(x)$ is the sum of the two values of $\phi^{-1}(x)$; for $x > 0$, $f(x)$ is the sum of $\phi(x)$ and the positive value of $\phi^{-1}(x)$. The reader can make his own figure for this case.

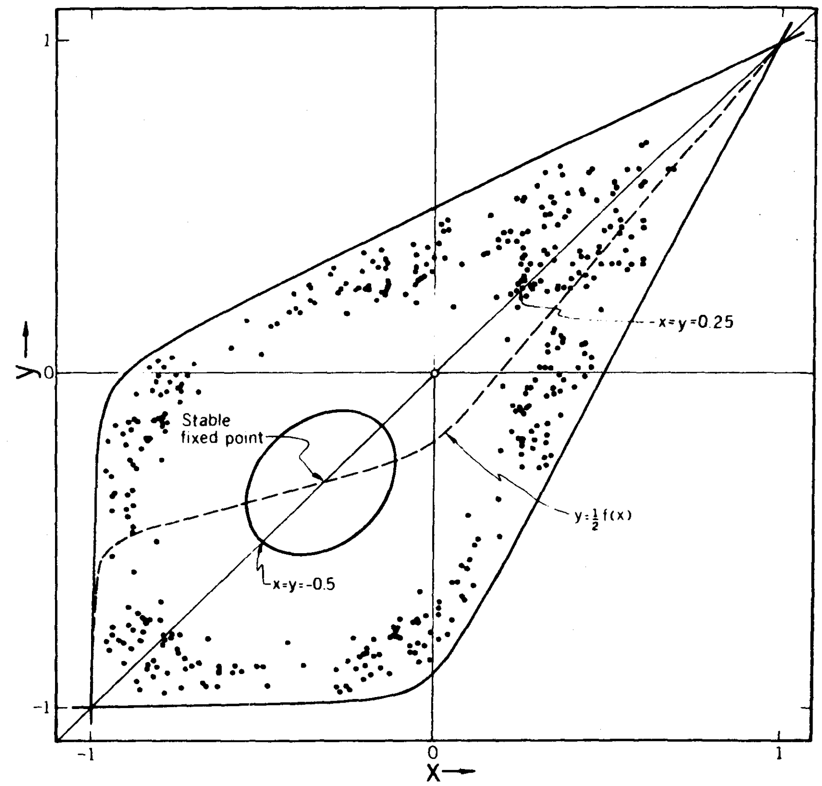

Fig. 8. The case $f(x) = (3x - 1)/2 - k^2/2(x + 1) + \sqrt{x^2 + k^2}$, with 0.1 as the value of k. The invariant boundary consists of two hyperbolas. The results of two computer runs are shown. A run starting at $x = y = -0.5$ generates the apparently smooth curve surrounding the stable fixed point at $x = y \approx -0.328$, and a run starting at $x = y = 0.25$ gives, for the first 400 iterations, the scattered points indicated as dots.

The enormous variety of such curves that can be drawn is obvious. A fascinating point here is that the inverse problem, given $f(x)$ to find $\phi(x)$, or even to establish the existence or non-existence of a closed invariant, seems completely intractable except by computer, or in a few cases with particular analytic forms.

Equally difficult seems to be the determination whether a curve (or pair of curves) so drawn are singular cases of closed invariants for the corresponding transformation, or whether they are members of a continuous family of closed invariants, or even whether there are any other closed invariants at all. As the reader may imagine, the first step

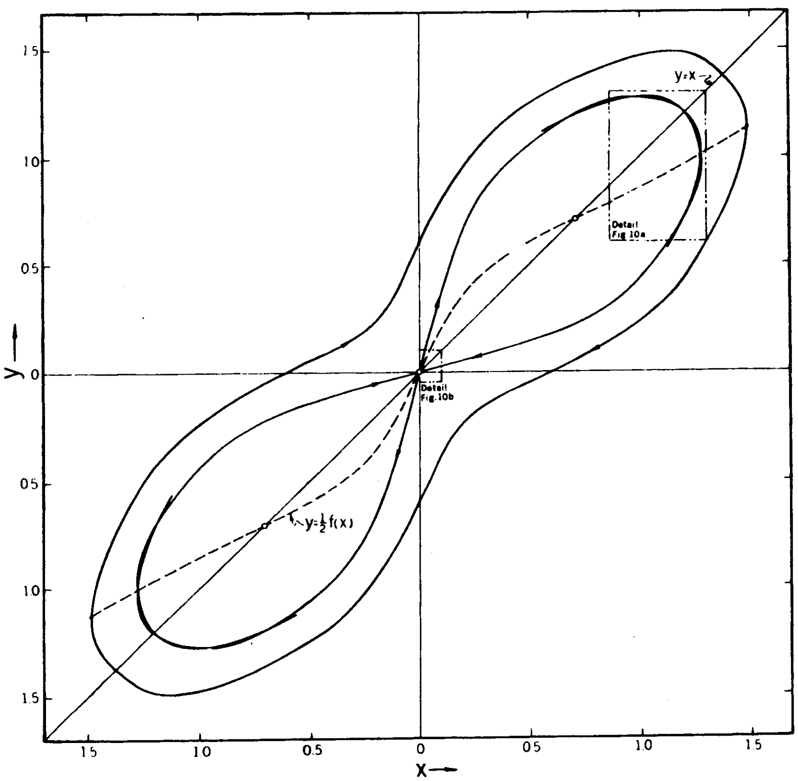

Fig. 9. The invariant boundary in this case is given by the curve $x^4 + y^2 - 4xy = \frac{1}{4}$. There is an unstable fixed point at the origin and stable fixed points at $x = y = \pm 1/\sqrt{2}$. The four invariant curves leaving the origin are followed through their first crossing of the diagonal $x = y$, where they meet in pairs at an angle of about 4°. Figure 10 shows the parts in the indicated rectangles, on an expanded scale.

after the discovery of Eq. (11) was to investigate a variety of cases. One of the earliest cases tried was that shown in Figure 6, with the rectangular hyperbolas taken as ϕ and ϕ^{-1}, and some computer runs were made which seemed to show that there were closed invariants in the area enclosed by the hyperbolas. The fact that all points are on closed invariants in this case, with the hyperbolas serving as separatrices of a family of curves, was pointed out by John M. Greene of

Princeton University. This observation gave the clue that led to the discovery of the wider set of cases described by Eq. (15).

Figure 8 demonstrates the fact that a closed invariant boundary does not necessarily imply that all interior points must lie on invariants. The boundary consists of the two hyperbolas $y = x - 1 + \sqrt{x^2 + k^2}$ and $y = (x + 1)/2 - k^2/2(x + 1)$. If the parameter k approaches zero, the hyperbolas approach their asymptotes, and the boundary approaches one made of straight line segments; $k = 0.1$ is the case shown in the figure. We know that all points starting inside the boundary will remain inside, but we do not know how they will move except by computation. Since there is no unstable fixed point inside, there is no way to compute an invariant curve; all that we can do is to pick a starting point, let the computer iterate the transformation repeatedly, and see where it goes. The results of two computer runs are shown in the figure. The first starts at $x = y = -0.5$, and generates an apparently closed curve surrounding the stable fixed point. After 10^6 iterations there is no evidence of deviation from a single smooth curve within the estimated precision of 10^{-18}. The second run shown starts at $x = y = 0.25$ and the behavior is entirely different, as illustrated by the first 400 iterations, plotted as points which scatter over a wide area. The distribution shows evidence of structure, indicating the presence of higher order fixed points and "islands." Similar runs started at $x = y = 0.5$ and 0.75 also give scattered points. We seem to lack completely any criterion, short of computation, for determining whether a given starting point will lead to a smooth curve or a set of scattered points, and in fact we have no way of being sure that the smoothness of the smooth curves is absolute.

Figure 9 shows another case, in which a set of invariant curves inside the stable boundary can be computed. The boundary is the curve:

$$x^4 + y^4 - 4xy = \frac{1}{4}. \tag{23}$$

The evaluation of ϕ and f requires the solution of a quartic equation. There are two stable fixed points inside the boundary, and an unstable fixed point at the origin, where the slope of $f/2$ is equal to 2. The four invariant curves leaving the origin with slopes $2 \pm \sqrt{3}$ can be followed by computer. It is found that they do not join smoothly to form two closed curves; they engage in the game of endless crossing and recrossing, all inside the outer invariant boundary. Figure 10 shows parts of Figure 9 on expanded scales, illustrating early stages of this process. We recall that the loops between crossings all have the same area, and wonder why the available area is not finally exhausted. We do not reckon with the ingenuity of the curves, which find a way to make loops overlap previous loops, without requiring that any curve cross itself.

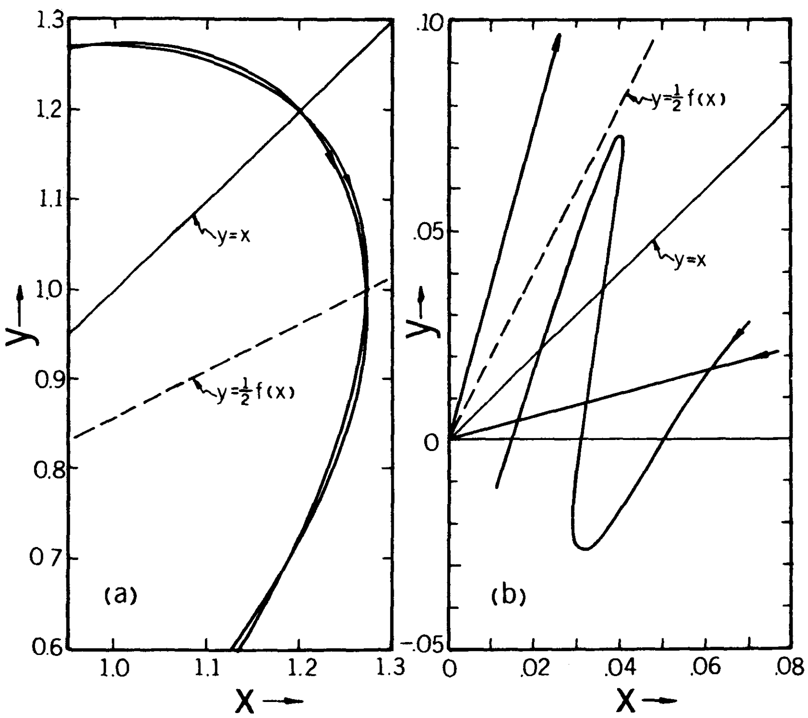

Fig. 10. Panel (a) shows the behavior of the invariants in Fig. 9 in the region of the first crossing at the diagonal; (b) shows a later stage, with incipient "worms."

Some examples of invariants made of straight line segments are shown in Figure 11. The functions $f(x)$ in these examples all have the same general form, consisting of a straight line through the origin extending from $x = -1$ to $x = 1$, and two lines with a different slope (but both the same) extending to the right and left of the central segment. If certain relations between the slopes are satisfied, the resulting transformations give closed invariants also consisting of straight line segments. The relations are easily found if the invariants are drawn first, then $\frac{1}{2}f(x)$ is constructed. In the cases illustrated, the central segment of $\frac{1}{2}f(x)$ has the slope $2/3$, while the outer segments have the slopes 5, $3/2$, $3/5$, and $-2/7$. There is a stable fixed point at the origin, and the first two cases have unstable fixed points at the vertices on the diagonal $x = y$. The interior is filled with elliptical invariants up to an

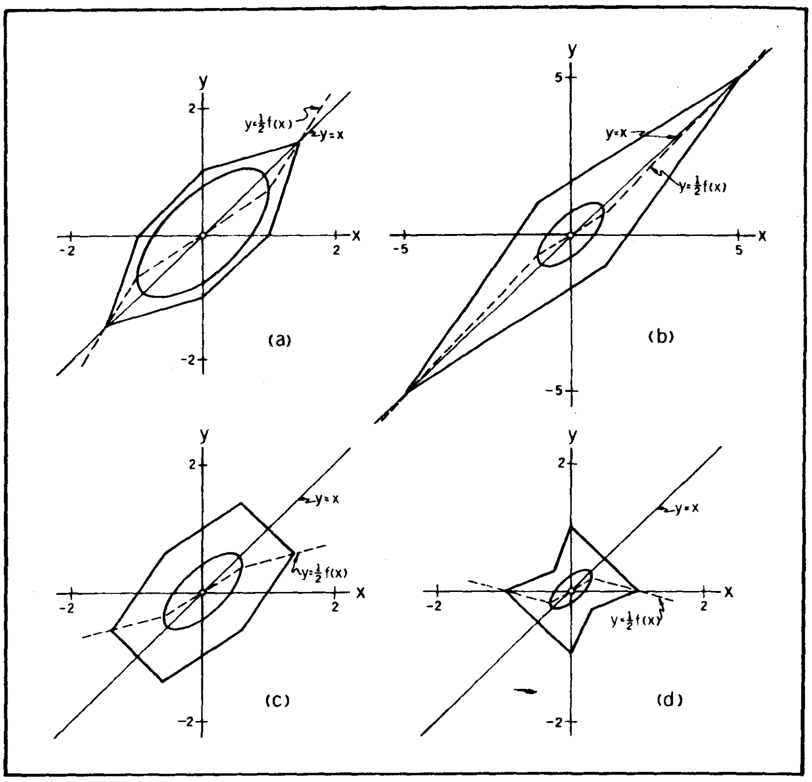

Fig. 11. Some examples of invariant boundaries made of straight line segments. These cases all have a center of symmetry, and were designed to give a function $\frac{1}{2}f(x)$ consisting of only three segments. The slope between $x = -1$ and $x = 1$ is equal to $\frac{2}{3}$ in all, and the slope of the outer segments is 5 in (a), $\frac{3}{2}$ in (b), $\frac{3}{5}$ in (c), and $-\frac{2}{7}$ in (d). In the central regions, the transformation gives a family of ellipses bounded by the ones drawn; between these ellipses and the straight-line boundaries the behavior is more complicated, with higher-order fixed points, islands, and areas of scattered points.

amplitude of ± 1 in x and y; between the last ellipse and the rectilinear boundary the motion of points under the transformation seems to be very irregular, but it is not clear whether it is ergodic. A still different choice of slopes can give a case that is an analogue of that shown in Figure 9, in the sense that it has one unstable and two stable fixed points inside a fixed boundary, and can be used (with less computational labor) to illustrate the formation of "worms" inside a boundary. Figure 12 shows the case with the slopes equal to 2 and $-1/4$, which has stable fixed points at $x = y = \pm 9/5$.

In an earlier paragraph a promise was made that cases could be constructed with boundaries that are more than double valued. I shall now fulfill that promise by describing a procedure by which this can be done. Take any known case with a center of symmetry at the origin, and erase the part of the boundary lying in the $+, +$ quadrant. Replace $f(x)$ by $f(x) = 0$ to the right of the origin. Fill in the erased part of the boundary by reflecting the part in the $+, -$ quadrant about the x-axis. The resulting completed boundary is an invariant under the transforma-

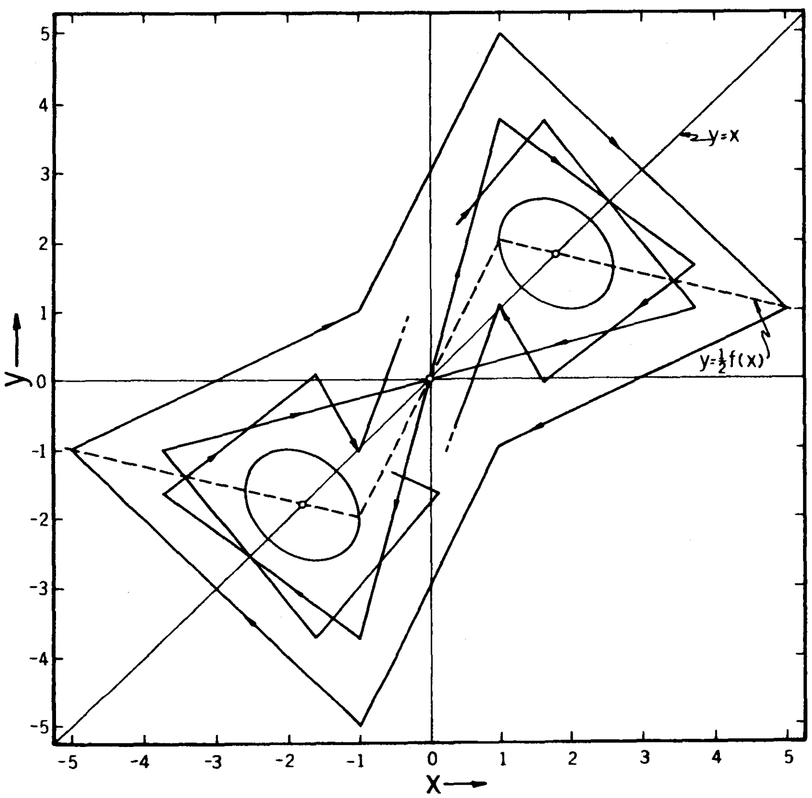

Fig. 12. Another example of an invariant boundary composed of straight lines, in which the function $y = \frac{1}{2}f(x)$ has the slope 2 between $x = -1$ and $x = 1$, and the slope $-\frac{1}{4}$ outside this range. There are stable fixed points at $x = y = \pm\frac{9}{5}$ and an unstable fixed point at the origin. Invariant lines leave the latter at slopes of $2 \pm \sqrt{3}$. The continuation of these lines is followed for a short distance. Their failure to join in closed figures will lead to the generation of "worms" made of joined line segments (angleworms?). The ellipses surrounding the stable fixed points are regions of regular behavior.

tion with the original $f(x)$ to the left of the boundary, and $f(x) = 0$ to the right. Starting with ellipses, we can get beheaded ellipses (Figure 13) or doubled-headed ellipses (Figure 14), and here we see a four-valued function acting as an invariant boundary.

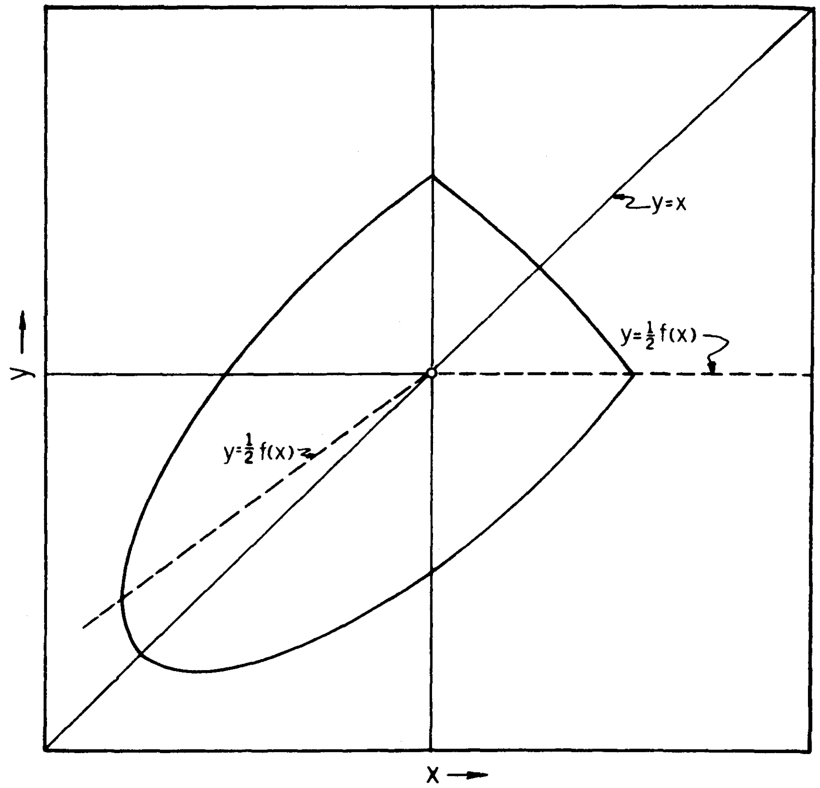

Fig. 13. The beheaded ellipse. See text.

I had intended to let these be the last examples, but Dr. Laslett has brought me one that is so striking that it would not be fair to the reader to leave it out, even though it does encroach into the domain of higher-order fixed points, which I had expected to say no more about. The function $\frac{1}{2}f(x)$ is very simple; it is zero for positive values of x, has a constant positive value $k(<1)$ below $x = -1$, and is a straight line of slope $-k$ between these regions. The resulting pattern of invariants is shown in Figure 15. In the center is the two-headed ellipse, as in Figure 14; similar shapes in its interior are also invariants, but outside is a region of irregular behavior. Embedded in this region there are seven "islands," each occupied by a family of invariants. The transformation carries points from island to island, in a sequence shown in the figure. The reader can easily verify this sequence, performing the transformation by reflecting first through the diagonal and then vertically through the function $y = \frac{1}{2}f(x)$. Each island contains a stable fixed point of the seventh order, and between neighboring islands there are unstable fixed points of the seventh order. If the seventh-order invariants are extended from these points, they do not join smoothly; the sea between the islands is inhabited by worms!

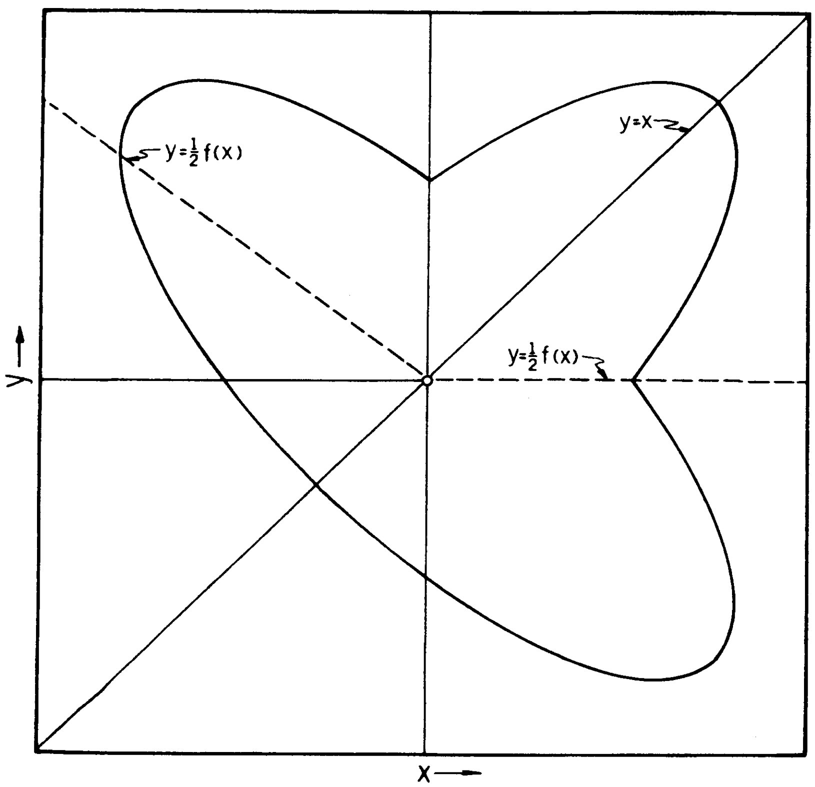

Fig. 14. The two-headed ellipse. See text.

How can one understand a figure of this complexity? The two-headed ellipse in the center suggests that the process described in connection with Figure 14 has been performed; this can be undone, leading to the restoration of the right half of the figure in a form centrally-symmetric with the left half. There are now six islands, but the transformation still moves a point two steps clockwise, and the fixed points divide into sets of the third order. The sequence is simplified if the whole figure is rotated by 90°. Since it has reflection symmetry about both diagonals, the requirement of diagonal symmetry is still satisfied, and the "second symmetry" is restored by redrawing $\frac{1}{2}f(x)$ as before, but with the opposite sign. In the rotated case (Figure 16) it is found that the transformation moves points from an island to the adjacent one, and we have fixed points of the sixth order. (Incidentally, by using the process of Figure 13 this can be reduced to the fifth order.)

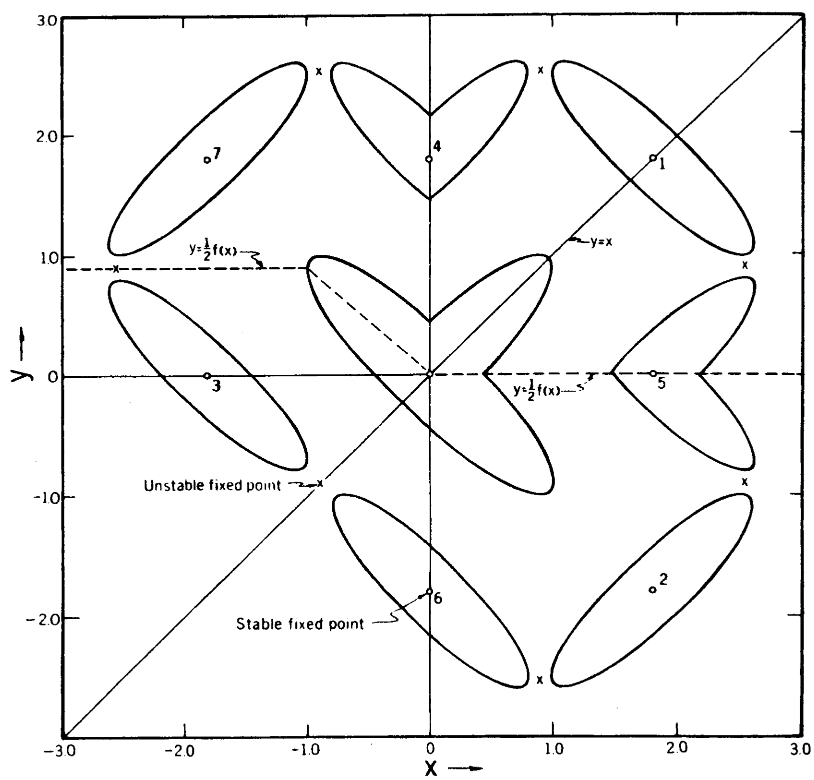

Fig. 15. An illustration of "islands." See text.

Let us now look at the stable fixed points in Figure 16. Points 1 and 2 (and 4 and 5) are equally spaced above and below the lines at $y = k$ (and $-k$); points 3 and 6 are equally spaced above and below the origin; therefore the "second symmetry" is satisfied. Point 1 is at $x = y = 2k$, point 2 at $x = 2k$, $y = 0$, etc. The limiting ellipses are similar to the central ellipse, but smaller; with a horizontal extent equal to $4k - 2$ rather than 2. The location of the unstable fixed points is slightly harder to find, because some of them lie above and below the sloping part of $y = \frac{1}{2}f(x)$. It turns out that the unstable point between islands 1 and 2 is at $x = k + 2k^2$, $y = k$, that between islands 2 and 3 is at $x = k$, $y = -k$, etc., ending with $x = k$, $y = k + 2k^2$ for the point between islands 6 and 1. The mid-point between the latter and the one below it is at $y = \frac{1}{2}(k + 2k^2 - k) = k^2$, in agreement with the value of $\frac{1}{2}f(x)$ at $x = k$.

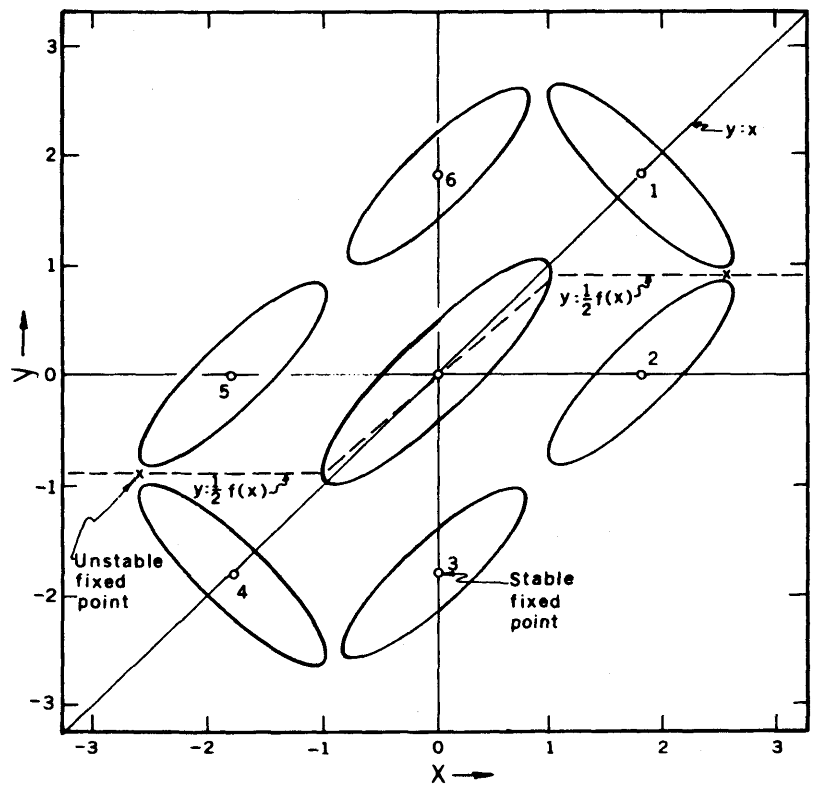

Fig. 16. Another set of "islands," related to Fig. 15. See text.

Earlier I said that this is not a treatise on stability theory; it is an account of the discovery, interpretation, and use of the relation expressed in equation (11), with illustrative examples. This relation does not by any means solve the difficult problems of stability theory, but it may be helpful, and it leads to some interesting territory, only a small part of which has been explored. A brief report of the work discussed in this paper was issued as UCRL-1795 (University of California Lawrence Radiation Laboratory, September 5, 1967) with an

Addendum dated March 29, 1968. The literature of stability theory is very extensive; I give here one reference which can help to lead the interested reader into that literature. This is a paper by M. Hénon, "Numerical Study of Quadratic Area-Preserving Mappings," in the *Quarterly of Applied Mathematics* XXVII, 291–312 (1969). (The particular transformation that Hénon considers can, by a coordinate transformation, be put into the form (7) with $f(y) = 2y \cos\alpha + y^2 \sin\alpha$, where α is a parameter.)

I want to express my appreciation to L. Jackson Laslett and René de Vogelaere for arousing my interest in this subject and for many useful discussions, and to Claude Froeschlé, John M. Greene, Michel Hénon and Jürgen K. Moser for help through correspondence. The computational work in connection with the preparation of this paper was done by Dr. Laslett, using computer facilities supported by the U.S. Atomic Energy Commission, and the drawings for the figures were made by Mrs. Jacqueline Dols.