

# THE PATHOLOGY OF DIFFERENTIAL EQUATIONS

T. M. CHERRY\*

1. When a system of ordinary differential equations, say

$$(1) \quad \dot{x}_r = \frac{dx_r}{dt} = X_r(x_1, \dots, x_n), \quad (r = 1, \dots, n),$$

or in vector notation  $\dot{\mathbf{x}} = \mathbf{X}(\mathbf{x})$ , arises in a physical or even a purely mathematical context, it is usually desired to solve it; to 'find', in some adequate sense, a specified particular solution  $\mathbf{x} = \mathbf{x}(t)$  or the set of all its solutions. If the range of  $t$  over which the solution is desired is unbounded this problem is usually intractable; for example, there have been two centuries of inconclusive effort to establish the stability of the solar system — as idealized by the equations of motion of a set of mutually gravitating particles. There are however, related questions on which progress has been made over the last 70 years. For example (1) we can inquire, for functions  $X_r$  of a given species, say polynomials, what sorts of solution are possible, and what sorts of structure are possible for the set of all solutions; (2) the question of solution near an equilibrium point, though in one sense 'local', has something of a global character, and notable progress with it has recently been made.

I propose in this address to speak about these questions at the level of the 'general reader', trying to give interesting indications rather than a complete review. Some of the material has not hitherto been published (so far as I know), and for the rest I give some historical notes and main references.

The chief object of the address is to illustrate the fact that *the solutions of systems (1) can have pathological features, even when the  $X_r$  are functions of the simplest sort, e.g. polynomials*. Here there is of course a contrast with the easily soluble systems which one meets in the elementary text books. This fact was adumbrated by Poincaré in his great memoir of 1890 on the equations of dynamics, but I think that it was first conclusively demonstrated by Hadamard in 1898 (for the set of all solutions), and Morse in 1921 (for individual solutions) [1]. I forbear to suppose the  $X_r$  anything but analytic, for this would invite further complication; for definiteness they are hereafter restricted to be polynomials, or at worst integral functions.

\* Presidential address, August 1957.

## Solution Curves

2. There is a well known geometrical representation in which, restricting all the variables to be real, we take  $x_1, \dots, x_n$  as coordinates in some 'phase-space', e.g. Euclidean  $n$ -space, and  $t$  a time variable. Then (1) specify a velocity-field in the space, and a solution  $\mathbf{x}(t)$  is represented by a parametrized curve called a *trajectory* or (if the  $t$  dependence is disregarded) a *path*. Since (1) is autonomous ( $t$  absent from  $\mathbf{X}$ ) the solutions fall into sets  $\mathbf{x}(t - t_0)$ , where  $t_0$  is arbitrary, and for any such set there is only one path. The range of  $t$  is  $(-\infty, \infty)$ , so each trajectory is an open curve. (It is possible for  $\mathbf{x}(t)$  to become infinite for a finite limiting value of  $t$ , but in the cases to be here considered this will not occur, and for simplicity the possibility will be neglected in general statements.)

There are standard theorems of existence, uniqueness and continuity which state (i) that the trajectory determined by any initial point  $\mathbf{x}(t_0)$  is a regular curve for  $-\infty < t < \infty$ , (ii) that two non-identical paths cannot intersect, (iii) that for any *finite*  $t$ ,  $\mathbf{x}(t)$  varies continuously with  $\mathbf{x}(t_0)$  — but the continuity is not uniform over  $-\infty < t < \infty$ . Accordingly, it is in the behaviour of trajectories when indefinitely continued ( $t \rightarrow \pm \infty$ ) that the pathological possibilities lie.

To focus this matter, suppose for the initial  $\mathbf{x}(t_0)$  in question that  $\mathbf{x}(t)$  remains bounded, and let  $(t_r)$  be any sequence increasing to  $+\infty$ . Then the set of points  $\mathbf{x}(t_r)$  has limit points, and the set of all such limit points for all such sequences  $(t_r)$  is called *the  $\omega$ -limit set of the trajectory  $\mathbf{x}(t)$* ; from the continuity theorem it is easily proved to consist of a closed set of trajectories. *The  $\alpha$ -limit set of the trajectory  $\mathbf{x}(t)$*  is similarly defined via sequences  $(t_r)$  decreasing to  $-\infty$ . For example the system

$$\dot{x} = x(1 - x^2 - y^2) - y, \quad \dot{y} = y(1 - x^2 - y^2) + x$$

has in polar coordinates a solution  $r^2 = \frac{1}{2}(1 + \tanh t)$ ,  $\theta = t$ , whose  $\omega$ -limit set is the circle  $r = 1$  (itself a solution) and  $\alpha$ -limit set the point  $r = 0$  (an equilibrium solution).

These definitions and the associated theory and due to G. D. Birkhoff [2], and the  $\alpha$   $\omega$ -terminology is one of his happiest appropriations.

The preceding adumbrations about 'pathological features' can now be clarified: *a system (1) can have a solution whose  $\omega$ - or  $\alpha$ -limit set is pathological*. An example will be given in § 5.

3. *Second order systems*. When  $n = 2$  in (1) the phase-space is naturally taken as a plane in which  $x_1, x_2$  are cartesian coordinates. The topological possibilities regarding the trajectory-field were investigated by Poincaré in the 1880's, and the theory is accessible in recently published books [3]. The phase-plane provides a topological straight-jacket, as it were, to restrict the possible waywardness of the trajectories, so that in general the  $\omega$ - or

$\alpha$ -limit set of a trajectory  $T$  is a single closed curve  $C$ . It is possible — as in the example below — for  $C$  to have angular points on it, but in general  $C$  is everywhere smooth; and then it is a closed *path* of (1), which is either (i)  $T$  itself; or (ii) a ‘limit cycle’ or asymptote which  $T$  approaches spirally from one side. In case (ii) neighbourhoods of  $C$  on *each* side are filled by paths which, like  $T$ , spiral towards  $C$ , and this happens in case (i) also if  $T$  is a limit cycle. But in case (i) it is possible — more exceptionally — that a full neighbourhood of  $C$  is filled by a family of closed paths; this commonly occurs when the differential equations specify the motion of a conservative dynamical system. There are no really pathological possibilities, but the following example shows that the trajectory-field may have unexpected features.

Let  $f(x, y) = (x^2 + y^2)^2 + 2(y^2 - x^2)$ , so that  $f = \text{const.}$  gives a family of Cassini ovals; for  $f > 0$  these enclose the figure-of-eight curve or lemniscate  $f = 0$ , and for  $f < 0$  we get ovals within each eye of the  $\infty$ . This family of ovals has the differential equation  $dy/dx = -f_x/f_y$ . Then the equations

$$(2) \quad \dot{x} = f_y + f_x \tan \beta, \quad \dot{y} = -f_x + f_y \tan \beta$$

specify paths, say  $\beta$ -paths, which cut the ovals at a sensed angle  $\beta$ , which need not be a constant. These  $\beta$ -paths are hence easily sketched. If we take  $\tan \beta = xf(x, y)$  the lemniscate  $f = 0$  is a  $\beta$ -path; inside each of its eyes the  $\beta$ -paths spiral from the equilibrium point  $y = 0, x = \pm 1$  out to the lemniscate as asymptote; but outside the lemniscate every  $\beta$ -path is closed, on account of the anti-symmetry of  $dy/dx$  as regards  $x$ . Thus we obtain an analytic family of curves of which some are spirals and ‘equally many’ are closed ovals.

By taking in (2)  $f = y^2 - (x^2 - 1)^2(x^2 - 2)^2$  with  $\tan \beta = xf$  we get  $\beta$ -paths showing a similar phenomenon, where now a central compartment containing ovals is flanked on each side by compartments containing spirals.

4. *Paths on a torus.* Next in simplicity to the case  $n = 2$  in (1) comes  $n = 3$ , with the  $X_r$  so chosen that there is an integral

$$(3) \quad F(x_1, x_2, x_3) = \text{const.},$$

where  $F$  is a polynomial or other single-valued function. Then each path is confined to one of the surfaces (3), and if  $x, y$  are curvilinear coordinates on this surface the paths on it are specified by differential equations

$$(4) \quad \dot{x} = X(x, y), \quad \dot{y} = Y(x, y).$$

The decisive distinction from the case  $n = 2$  of (1) is that the surfaces (3) need not be topologically plane.

The simplest new case arises when the surface (3) in question is a torus, on which we take  $x, y$  to be angular coordinates of the usual sort, each of period 1, say. An auxiliary plane on which  $x, y$  are cartesian coordinates

is the universal covering surface of the torus, and any path on the torus has  $\infty^2$  congruent representations on the plane. In (4),  $X$  and  $Y$  are periodic in each of  $x$  and  $y$ . In the trivial case where  $X, Y$  are constant the paths in the plane are parallel straight lines, which give on the torus a family of closed curves if  $X/Y$  is rational; but if  $X/Y$  is irrational each path spirals round the torus without closing, so that it is everywhere dense on the torus and has this whole surface as its  $\alpha$ -limit set and as its  $\omega$ -limit set. Each such path cuts any section  $x = \text{const.}$  in a countable infinity of points, so there is an uncountable infinity of distinct paths.

Imagine now that these paths are drawn on a deformable membrane stretched over the torus, that the membrane is slit for a short distance along the path through a chosen point  $P$ , and that the slit is then opened out into a region of finite breadth. Hereby the part of the membrane near  $P$  is pushed aside, so that the paths on it become deformed. Now suppose the slit is continually extended along the path  $T$  through  $P$ , in both senses, and that it is continually opened out so that a strip of finite breadth is generated, with an associated deformation of the neighbouring paths (which will include 'later' stretches of the path  $T$ , not yet reached by the slit); any part of the strip already generated is supposed to have its edges held fixed, so the strip keeps encroaching on more and more of the surface, and its breadth must diminish towards zero as it is indefinitely followed in either sense. Finally we can suppose that the strip is filled with new paths strung along it, so that the torus is completely covered by a set of paths, which can be specified by differential equations (4) in which, clearly,  $X, Y$  can be supposed continuous functions which never simultaneously vanish.

This set of paths has pathological properties, which are clearest in the case (which we may suppose achieved) where the strip is everywhere dense on the torus. If we take the strip to be open (i.e. edges excluded) the residual set  $R$  on the torus is perfect and nowhere dense and consists of an uncountable set of paths; and every path on the torus has  $R$  as its  $\alpha$ - and  $\omega$ -limit sets. The paths which constitute  $R$  are called recurrent (because they belong to their limit sets) and of discontinuous type; the contrast is with the paths in the case where  $X/Y$  is an irrational constant, which are recurrent and of continuous type.

For half a century it was a question of some notoriety whether this pathological case could occur for differential equations (4) in which  $X, Y$  are analytic functions, but in 1932 Denjoy [4] disproved the conjecture; the argument is too long and subtle to be given here, but the key fact can be informally stated: the 'positive half' of any trajectory (corresponding to  $t > 0$ ) partitions the surface into 'a set of strips nearly all of which must be infinitely narrow', and the negative half of the trajectory is forced to thread its way along these strips.

5. *A pathological example* [5]. We can achieve a pathological set of paths closely resembling the set just discussed, *but whose differential equations (4) have  $X, Y$  analytic*, provided we take  $X, Y$  to vanish simultaneously at two points  $A, C$  of the torus. Specifically, the paths are defined by (2), in which  $\beta$  is a constant to be suitably chosen, and  $f$  has level curves which, in the  $xy$ -plane, consist of  $\infty^2$  repetitions of the unit shown by thin lines in figure 1; on the torus, the family consists of a set of closed curves, for

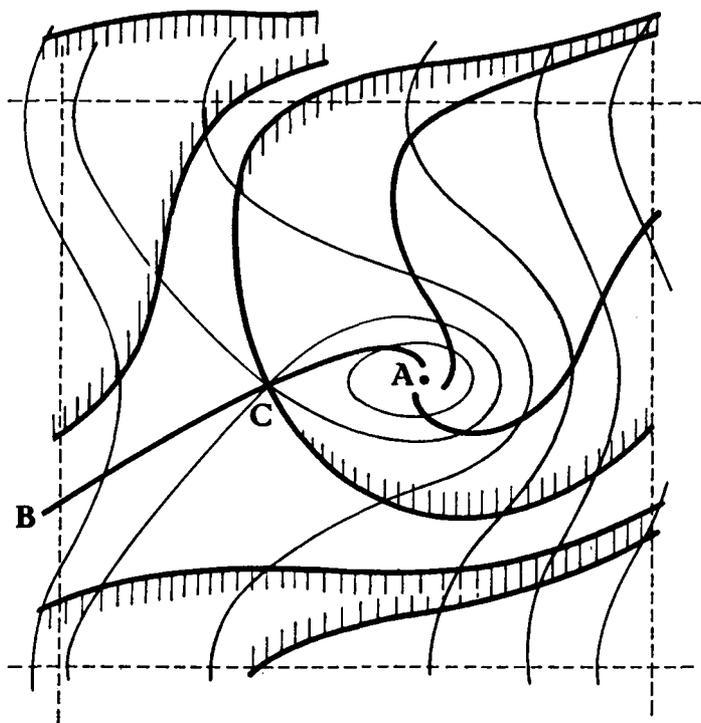


Figure 1. Unit of guide curves (thin) and  $\beta$ -paths (thick)

which (for example) we can take its intersections with a family of cylinders, whose cross sections are congruent equiangular spirals and whose polar axis is parallel to the axis of symmetry of the torus but not coincident with it; the singular points  $A, C$  arise from tangencies of two of the cylinders with the torus — external tangency for  $A$ , internal for  $C$ .

We call the family  $f = \text{const.}$  the guide-curves, and (2) specify trajectories cutting the guide curves at a constant angle  $\beta$ , which are easily sketched as on the figure. A set of them spiral outwards from  $A$ , for  $t$  increasing (say) from  $-\infty$ ; one of these goes for  $t \rightarrow +\infty$  into the col  $C$ , and the others form a strip bordered by the two trajectories running away from  $C$ . This strip plays the part of the strip described in § 4, with the unessential difference that its continuation leftwards consists of a single path  $BC$  running into  $C$ . By considering the  $\infty^2$  congruent representations in the  $xy$ -plane it is intuitive (i) that there is an average gradient  $g$  for each indefinitely extended trajectory, the same for all of them; (ii) that if  $g$  is rational the strip (considered on the torus), when followed in the sense of  $t$  increasing, abuts at

some stage on to  $C$  and then divides (in general) into two ribbons which border the earlier part of the strip; (iii) that if  $g$  is irrational the continuation of the strip never abuts in this way on  $C$ ; (iv) that  $g$  varies monotonically and continuously with  $\beta$ , but the variation is pathological in that  $g$  remains fixed at any *rational* value as  $\beta$  increases over a finite range — during which the continuation of the strip is traversing over  $C$ . Hence an irrational  $g$  can be achieved by suitable choice of  $\beta$ .

When  $g$  is irrational, (iii) gives us the pathological situation that the strip is cut by any section  $x = \text{const.}$  of the torus in an infinite set of disjunct intervals, and there are trajectories residual to the strip which are recurrent and of discontinuous type; in certain cases all the residual trajectories are of this type, but the discrimination regarding this is not a trivial matter.

Finally we can exhibit a third order system (1), in which the  $X_r$  are polynomials, having this sort of pathological solution. We start from two integrals such as (3), one representing a family of tori and the other a family of equiangular spiral cylinders, and by eliminating the arbitrary constants arrive at a system (1) whose trajectories are the guide curves. This system is then modified so as to give on each torus  $F(x_1, x_2, x_3) = \text{const.}$  the  $\beta$ -paths, and finally we take  $\tan \beta = F$  so as to secure continuous variation of  $\beta$ .

### Hamiltonian Systems

#### 6. Differential equations of Hamiltonian form

$$(5) \quad \dot{x}_r = \frac{\partial H}{\partial y_r}, \quad \dot{y}_r = -\frac{\partial H}{\partial x_r}, \quad H = H(x, y), \quad (r = 1, \dots, m)$$

are of outstanding interest because of their dynamical applications and their beautiful formal properties. To provide a background for later sections of this address I shall say a little about one of these properties.

There is a theorem of Liouville, that if  $m$  independent involutory\* integrals of (5) are known, then the remaining  $m$  integrals that are required to complete the solution of the system can be found by quadrature. The classical elementary dynamical problems, such as that of motion under two fixed centres of gravitation, are in effect 'soluble' because such a primary set of  $m$  integrals  $F = \text{const.}$  can be found in closed form; these  $F$  moreover are regular single-valued functions, so the hypersurfaces  $F = \text{const.}$  partition the  $2m$ -dimensional phase-space into regular  $m$ -dimensional regions, and for any trajectory the  $\omega$ - and  $\alpha$ -limit sets can at most fill one such region. Now an arbitrary system (5) can be said to do its best to behave in this same manner, as follows:

\* The set is involutory if for each pair the Poisson bracket  $(F_1, F_2)$ , defined as in (6), is zero.

For a general theory we naturally try for developments in series, and the simplest case in which we can expect to obtain something approaching global validity from series is to use them for the neighbourhood of an equilibrium point, say the origin. Then the development of  $H$  in powers of the  $x_r, y_r$  starts with quadratic terms, and by a preliminary contact transformation these terms may in general be supposed of the form

$$\lambda_1 x_1 y_1 + \cdots + \lambda_m x_m y_m.$$

The condition for  $F$  to be an integral is

$$(6) \quad (F, H) \equiv \sum_r \left( \frac{\partial F}{\partial x_r} \frac{\partial H}{\partial y_r} - \frac{\partial F}{\partial y_r} \frac{\partial H}{\partial x_r} \right) = 0.$$

A power series  $F = x_p y_p + \cdots$  will satisfy this provided its constituent homogeneous polynomials,  $T$  say, of successive degrees satisfy equations of the form

$$\sum_r \lambda_r \left( x_r \frac{\partial T}{\partial x_r} - y_r \frac{\partial T}{\partial y_r} \right) = \Sigma A \Pi (x_r^{a_r} y_r^{b_r})$$

where the coefficients  $A$  have a complicated dependence on the terms of lower degree; and the solution is a polynomial

$$(7) \quad T = \Sigma \frac{A \Pi (x_r^{a_r} y_r^{b_r})}{\Sigma \lambda_r (a_r - b_r)}$$

provided only that 'critical' terms, for which the denominator vanishes, 'happen' to have their coefficients  $A$  zero. If  $\lambda_1, \cdots, \lambda_m$  are mutually incommensurable the only critical terms are those for which  $a_r = b_r$  for each  $r$ , and their coefficients  $A$  are in fact zero, but the proof of this is not trivial [6]. Hence if the series converge we have the desired set of integrals. However, in the most interesting case where two at least of the  $\lambda_r$  have a real ratio the divisor on the right of (7) can for suitable  $a, b$  be arbitrarily close to zero, and the convergence of the series remains doubtful; this is an instance of the notorious 'small divisor' problem'.

In fact, there seems little doubt that the series are in general essentially divergent, and that the solutions of general Hamiltonian systems have pathological features which distinguish them sharply from the elementary 'soluble' systems [7]. An example will now be given, and the vital fact will be that two functions which might, on analytical continuation, be identical are not so; in other words, an infinite set of constants which might all be zero are not so. The preceding fact about formal series-integrals shows that the non-zero property must not blithely be assumed 'on general grounds', but must be proved to the hilt.

7. *Pendulum with forced oscillations.* Consider a rigid rod  $AB$ , of length  $3l/2$  capable of swinging in a vertical plane about a frictionless pivot  $A$ .

The pivot is forced by an external agency to oscillate in a vertical line, so that its displacement  $sl$  is a given periodic function of  $t$ , of period  $2\pi\mu$ , say. The equations of motion in Hamiltonian form are

$$(8) \quad \begin{cases} \dot{x} = y + \dot{s} \sin x = \partial H / \partial y, \\ \dot{y} = n^2 \sin x - y\dot{s} \cos x - \frac{1}{2}\dot{s}^2 \sin 2x = -\partial H / \partial x, \\ H = \frac{1}{2}y^2 + n^2 \cos x + y\dot{s} \sin x - \frac{1}{2}\dot{s}^2 \cos^2 x, \end{cases}$$

where  $n^2 = g/l$ ,  $x$  is the inclination of  $AB$  to the upward vertical, and  $y$  is proportional to the angular momentum of the pendulum about the moving point  $A$ .

This system is converted into an autonomous one of the third order by putting  $t = \mu\theta$ , so that  $\dot{s}$  is periodic in  $\theta$  with period  $2\pi$  and mean value zero, say  $f(\theta)$ . Also if  $p^2$  is the mean value of  $\dot{s}^2$  we put  $\dot{s}^2 = p^2 + 2g(\theta)$ , where  $g(\theta)$  is periodic with mean value zero. Then the equivalent system is

$$(9) \quad \begin{cases} \dot{\theta} = 1/\mu \\ \dot{x} = y + f(\theta) \sin x \\ \dot{y} = n^2 \sin x - \frac{1}{2}p^2 \sin 2x - yf(\theta) \cos x - g(\theta) \sin 2x \end{cases}$$

and is equivalent to a 4th order Hamiltonian system since we can write

$$\dot{\theta} = \frac{\partial}{\partial \phi} \left( H + \frac{\phi}{\mu} \right).$$

The solutions of (9) are appropriately represented by paths in a space for which  $x$  and  $\theta$  are angular coordinates. For example we can take

$$\exp(e^y \cos x) \cos \theta, \quad \exp(e^y \cos x) \sin \theta, \quad e^y \sin x$$

as rectangular coordinates; for  $y \sim -\infty$  there is a limiting circle, which is to be regarded as an improper path since by (9)  $|y|$  can have at most exponential increase. The path through any point  $P$  of the half-plane  $\theta = 0$  intersects this half-plane again (for  $\theta = 2\pi$ ) at a point  $T(P)$ , where  $T$  indicates an analytic transformation of the half-plane into itself which has a single-valued inverse  $T^{-1}(P)$ . Writing  $T(T(P)) = T^2(P)$ , etc., an arbitrary arc  $\alpha$  joining  $P$  to  $T(P)$  determines a set of arcs

$$\dots T^{-1}(\alpha), \alpha, T(\alpha), T^2(\alpha), \dots$$

which abut so as to form an 'invariant curve'; in general there will be angular points where the arcs abut, but for suitable choice of  $\alpha$  the arcs may join regularly. The paths that cut such an invariant curve form a surface, and a family of such surfaces corresponds to an integral of (9).

The essential fact that is to be demonstrated is that there are invariant curves which are locally regular but globally pathological. The argument will be most vivid if stated largely in physical terms, but its mathematical translation presents no real difficulty.

We take the case where the vertically upward position of the pendulum is unstable, which will be so if  $0 < p^2 < n^2$  and  $f(\theta)$ ,  $g(\theta)$  are sufficiently small; this position is given by the periodic solution, or path,  $x = y = 0$ ,  $\theta = t/\mu$ , which cuts the half-plane  $\theta = 0$  at a point  $U$ , say. For the solutions near this periodic solution there is a theory initiated by Poincaré [8] and recently completed by Moser [9], who has shown that (9) has an integral  $F = \text{const.}$  which is periodic in  $\theta$  and regular for  $|x|$ ,  $|y|$  sufficiently small. To this corresponds a set of surfaces, and these are cut by the half-plane  $\theta = 0$  in a set of invariant curves which are topologically a family of hyperbolas, for the neighbourhood of  $U$ , as in figure 2. The waviness in these curves answers to the occurrence of  $f(\theta)$ ,  $g(\theta)$  in (9); but for the following argument this is irrelevant.

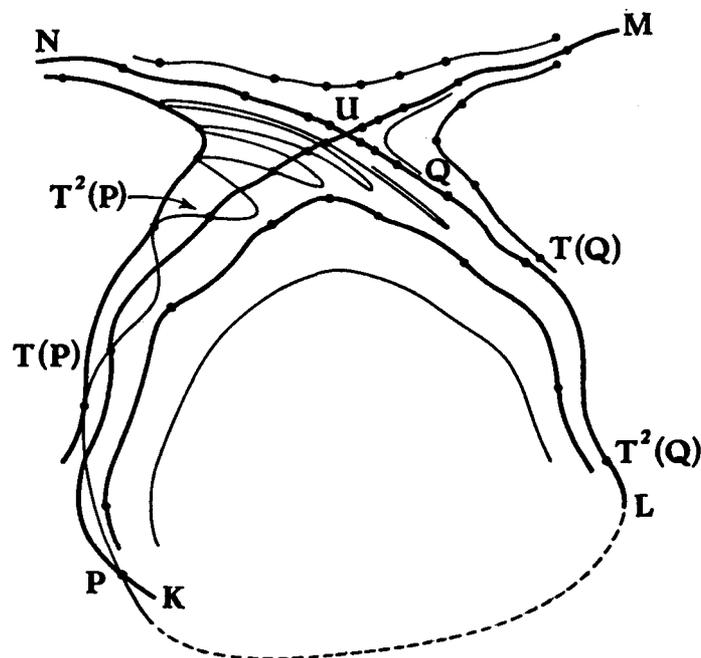


Figure 2. Invariant curves in the half-plane  $\theta = 0$ , for neighbourhood of  $U$

A chain of points  $T^n(P)$  on the curve  $UK$  represents a motion in which as  $t \rightarrow +\infty$  the pendulum makes an asymptotic approach, from the left side, to the unstable vertical position, and chains on  $UM$  represent similar motions in which the approach is from the right side. Points on  $UL$ ,  $UN$  similarly represent motions falling away from the vertical as  $t$  increases from  $-\infty$ . These four are called asymptotic curves. Chains of points in the sectors above and below  $U$  represent motions in which the vertical is overshoot, from right to left or left to right, and the sectors to either side of  $U$  similarly belong to motions in which the pendulum approaches the vertical but falls back without attaining it. It is an important feature that the points of a chain are the more tightly bunched in so far as they are the nearer to  $U$ .

It has already been remarked that any invariant curve can be indefinitely

continued, and it will now be shown that *the continuations of the asymptotic curves UK, UL are intersecting but not coincident*. For this, take first the case where the displacement  $s$  of  $A$  is an even function of  $t$ , with  $\dot{s}$  practically constant over each half-period, so that there are rapid transitions between its positive (ascending) and negative (descending) values. Then the angular motion of the pendulum proceeds almost as though  $A$  were at rest, except when  $A$  is jerked from its ascending to its descending phase or vice versa. At a downward (upward) jerk there will be an upward (downward) increment in the angular velocity\*, proportional roughly to  $|\sin x|$ .

Suppose first that the pendulum is swinging with angular velocity  $\omega$  through its bottom position  $x = \pi$  when a jerk, either upwards or downwards, occurs, at  $t = 0$ , say. Then the jerk is without effect on the angular velocity,  $x - \pi$  is an odd function of  $t$ , and by an argument involving continuity we can choose  $\omega = \omega_0$  so that the top position is approached as  $t \rightarrow \pm\infty$ . Hence the asymptotic curves intersect.

Now suppose that  $s = 0$ , with  $\dot{s}$  positive, when the pendulum is swinging with angular velocity  $\omega_0$  through  $x = \pi$  to the left; this is at an instant midway between two jerks, of which the first is upwards, and the second downwards. At the first jerk the pendulum was to the right and swinging down, and the jerk increased its clockwise angular velocity, while at the second jerk the pendulum will be to the left and swinging up, and its clockwise angular velocity will be again increased. From the asymmetry it is almost clear that the downward motion must have originated from a position, short of the upward vertical, where  $\dot{x} = 0$ , and that the upward motion must overshoot the vertical. This is indeed quite clear when the period  $2\pi\mu$  is such that  $x$  is about  $\frac{1}{2}\pi$  and  $\frac{3}{2}\pi$  at the respective jerks (so that they give the maximum effect), and the jerks are fairly violent. Moreover, a motion which originated from  $x = 0$  at  $t = -\infty$ , and reaches  $x = \pi$  midway between the upward and downward jerks, will there have  $\omega > \omega_0$ , and a fortiori the subsequent ascent will overshoot the vertical. Hence there are points on one asymptotic curve which are not on the other, and this proves our assertion under the assumed restrictions on the function  $s$  and the period  $2\pi\mu$ . An argument resting on analyticity shows finally that for general values of  $\mu$  and general even functions  $s(t)$  the asymptotic curves  $UK, UL$  intersect but are not coincident. (By further argument the restriction that  $s(t)$  be even can be removed.)

Let  $P$  be one such intersection. Since to a single motion corresponds the complete chain of points  $T^n(P)$ , the asymptotic curves must intersect at each point of this chain, so *the intersections accumulate at U on each of the curves*.

\* There is however, no sudden increment in the angular momentum variable  $y$ .

Now let  $\alpha$  be an arc of the continuation of  $UL$  which cuts  $UK$  near  $U$ . The character of the transformation  $T$  for the neighbourhood of  $U$  is known, and the arcs  $T(\alpha)$ ,  $T^2(\alpha)$ ,  $\dots$  are then determinate, as in the figure\*, in so far as they lie in this neighbourhood. It is seen that the set of arcs  $T^n(\alpha)$  has every point of  $UL$  and  $UN$  within the neighbourhood as a limit point; and by a continuity-argument the restriction to the neighbourhood can be removed. *Hence the asymptotic curves, globally considered, are pathological.*

This conclusion can be presented as an analytical theorem. The family of invariant curves was initially defined via an integral  $F = \text{const.}$  of (9) regular near the periodic solution  $x = y = 0$ ,  $\theta = t/\mu$  (where we can take  $F = 0$  on this solution), and the theorem is that *the function  $F(x, y, \theta)$  has continuations along its level surface  $F = 0$  which are essentially singular at all points of this surface.* This conclusion extends to *any* integral  $F_1 = \text{const.}$  which is locally regular near the periodic solution, for by considering as in the preceding paragraph the march of chains of arcs  $T^n(\alpha)$  it is seen that  $F_1$  can be locally regular only if  $F_1$  is constant on the surface  $F = 0$ .

The pathology of this example can be further pursued, but I forego this in order to call attention to a physical paradox. When  $\mu$  is small the oscillations of the pivot  $A$  are rapid, and we can plausibly approximate to the solution of (8) by replacing  $\dot{s}$  and  $\dot{s}^2$  by their mean values, i.e. by putting  $f = g = 0$  in (9); a strict solution would be obtained from an approximate solution by superposing a small tremor. On this approximation the upward vertical position  $\alpha = 0$  of the pendulum is stable if  $p^2 > n^2$ , i.e. if the mean speed of the oscillation is sufficiently rapid; and this can be verified by physical experiment. Moreover a given mean speed can be attained for an oscillation of arbitrarily small amplitude, provided the period is small enough, so the stabilization can be achieved by means of invisible oscillations! [10].

On investigating higher approximations for the case  $\mu \sim 0$  it is found that (9) has a formal integral

$$\frac{1}{2}y^2 + n^2 \cos x - \frac{1}{4}p^2 \cos 2x + \mu F_1 + \mu^2 F_2 + \dots = \text{const.},$$

where  $F_1, F_2, \dots$  are polynomials in  $y$  whose coefficients are periodic in  $x$  and  $\theta$ , e.g.

$$F_1 = (y^2 \cos x + n^2 \sin^2 x - \frac{1}{2}p^2 \sin x \sin 2x)f_1(\theta) + \frac{1}{4}y \sin 2x g_1(\theta) + \lambda \sin x \sin 2x,$$

where  $f_1, g_1$  are those periodic integrals of  $f, g$  whose mean values are zero and  $\lambda$  is the mean value of  $\frac{1}{4}(f_1 g - f g_1)$ . The preceding pathological result shows that the series cannot converge uniformly, for any range of  $\mu$ , all

\* They are the parts of the thin-line curve near  $P, T(P), T^2(P), \dots$ .

real  $x$ ,  $\theta$  and  $|y| < 2n$ , say, and it is likely that the series is essentially divergent.

8. *Transitivity.* It was remarked in § 6 that, for the classical ‘soluble’ Hamiltonian systems of order  $2m$ , the  $\omega$ - or  $\alpha$ -limit set of any trajectory is at most of dimension  $m$ . In Statistical Mechanics, systems for which  $m$  is enormously large are considered, which naturally cannot be explicitly solved and for which indeed the only known global integral is that of energy. It was seen by Maxwell and Boltzmann that a fundamental hypothesis of this subject would be effectively proved if a ‘general’ trajectory was everywhere dense on the appropriate energy-hypersurface, so that its  $\omega$ - and  $\alpha$ -limit sets were of dimension  $2m - 1$ . Birkhoff has called such a trajectory, and Hamiltonian systems possessing such trajectories, ‘transitive’; other authors, e.g. Artin [11], following Boltzmann, have used the name ‘quasi-ergodic’\*.

I believe that the only Hamiltonian systems that have been proved to be transitive are those that specify the geodesics on certain closed surfaces of negative curvature, and related systems that are only piecewise analytic; such surfaces are not of physical interest since they cannot be imbedded in Euclidean 3-space. The first such example was given by Artin [11] and starts from the well known modular figure in which the half-plane  $y > 0$  is divided into an infinite set of triangular cells by semi-circles. This figure is invariant under the transformations of the modular group, and the ‘closed surface’ arises by identifying all the ‘congruent’ points, viz. points of any set that is invariant under this group. The geodesics are then defined via the Poincaré metric  $ds = d\sigma/y$ , where  $d\sigma$  is Euclidean arc-length; they are semi-circles orthogonal to  $y = 0$ .

A transitive geodesic is one that, in its representation on the closed surface, is not merely everywhere dense, but traverses an arbitrary neighbourhood of each point of the surface, on different occasions, in directions that are dense in the set of all directions; it is the proviso about directions which makes the corresponding path in the phase-space everywhere dense on a 3-dimensional region, in contrast to the 2-dimensional density of unclosed geodesics on a torus. The translation of this into a property of the modular figure is easy, but there seems to be no way of making the property ‘evident’. A hint is however provided by the remark that the geodesic semi-circle will be required to traverse an infinity of the triangular cells near  $y = 0$ , and that in this neighbourhood the semi-circle is (in the Euclidean sense) almost straight but the cells show gross departure from similarity.

\* It is therefore misleading that Birkhoff and von Neumann used the name ‘ergodic theorem’ for what is *not* a proof that Hamiltonian systems in general have the quasi-ergodic property. The root trouble is perhaps that Boltzmann used ‘ergodic’ in two different, but related, senses; see P. and T. Ehrenfest, *Encycl. Math. Wiss.* IV 32, §§ 10a, b.

There is a modification in which, in the plane, only one cell is considered and a geodesic is continued through it as a succession of arcs by reflection at the boundaries of the cell. This cell can be represented, with Euclidean metric, on a pseudosphere, and the conclusion is that 'the hyperbolic billiard table permits transitive shots'.

In these examples it is far from the case that *every* geodesic is transitive. We have, rather, a complicated and intimate mixture of geodesics of many different sorts, both recurrent and asymptotic, amongst which for example the set of periodic geodesics is everywhere dense.

### Conclusion

9. In recent years there has been a considerable resurgence or work on differential equations, mainly directed towards non-linear oscillations. For example there is van der Pol's equation with forcing term,

$$(10) \quad \ddot{x} - k(1 - x^2)\dot{x} + x = bk\lambda \cos \lambda t,$$

whose feature is that it is dissipative for  $|x| > 1$  but 'explosive' for  $|x| < 1$ . If a dissipative system is subject to an external periodically varying force we may expect that, independently of the initial conditions, its motion will settle down to some stable oscillation, represented in the phase-space by a limit cycle (cf. § 3); and a good deal of the modern work has been devoted to investigating such conjectures and to methods for approximating to limit cycles. For equations like (10) the matter is naturally more complicated; we may expect that in an ultimate motion there will be a rough balance between the dissipative and explosive tendencies, and typical investigations rest upon the piecing together of approximations — buttressed by error-estimates — for motions under the two regimes. There is an extreme and characteristically written example of this by Littlewood [12], and general accounts of the subject are readily accessible [3].

On two of the classical problems there has recently been progress.

(i) For the differential equations

$$(11) \quad \dot{x}_r = \lambda_r x_r + \dots = X_r \quad (r = 1, \dots, n)$$

where  $+\dots$  indicates a power series in the  $x$  commencing with quadratic terms, we can seek for the general solution near the origin in the form

$$(12) \quad x_r = \alpha_r e^{\lambda_r t} + P_r,$$

where on the right we have power series in  $\alpha_1 e^{\lambda_1 t}, \dots, \alpha_n e^{\lambda_n t}$ , the  $\alpha$  being arbitrary constants. It is easily found that to obtain the term in  $\prod (\alpha_r e^{\lambda_r t})^{p_r}$  in the series for  $x_s$ , a division by

$$(13) \quad p_1 \lambda_1 + \dots + p_n \lambda_n - \lambda_s$$

is involved; the series is hence encumbered by 'small divisors' if the ratio of two of the  $\lambda$  is real negative, but is formally valid provided the  $\lambda$  are incommensurable. Now it is well known that the possible smallness of  $p - q\mu$  (where  $\mu$  is a positive irrational and  $p, q$  positive integers) for a given large  $q$  depends on the arithmetical character of  $\mu$ ; there are transcendentals  $\mu$  and associated sequences  $(p, q)$  such that  $p - q\mu$  tends to zero as fast as we please; but for almost all irrational  $\mu$ ,  $|p - q\mu|$  exceeds  $Aq^{-\nu}$ , for some constant  $A$  and  $\nu$ . It is therefore plausible that the convergence properties of the series (12) may be decisively affected by the arithmetical character of the ratios of the  $\lambda$  when these are real negative. In 1904 Dulac [13] showed that the series could be essentially divergent in the transcendental case, and recently Siegel by a tour de force [14] has shown that they are 'in general' convergent when the  $|\alpha_r e^{\lambda_r t}|$  are sufficiently small.

(ii) When the system (11) is Hamiltonian, with  $n = 2m$ , the condition that the  $\lambda$  be incommensurable is never satisfied since they fall into the pairs  $\pm \lambda_1, \dots, \pm \lambda_m$ ; but if  $\lambda_1, \dots, \lambda_m$  are incommensurable there is a formal solution resembling (12), in which the coefficients  $\lambda$  in the exponents are replaced by series involving the integration-constants  $\alpha$ . The small-divisor problem here regarding convergence appears to be less tractable than that treated by Siegel, and even the case where all the ratios of the  $\lambda$  are unreal, so that the divisors are not 'small', was outstanding until a recent publication by Moser [15].

10. If in (11) the  $\lambda$  are not incommensurable, the attempt to solve in the form (12) will generally fail through the vanishing of one of the divisors (13), and instead of a pure-exponential term in the series we get a 'secular term' carrying  $t$  as a factor. The greater formal complication of this case has led to its receiving less attention, but I believe that here we may in fact have something of a heel of Achilles. The following are two indications.

(i) There is both an analogy and a connection between differential equations and iteration equations such as

$$z_{n+1} = f(z_n) = \mu z_n + a z_n^2 + \dots,$$

whose 'solution' involves the effective determination of the sequences  $(z_n)$  of complex numbers defined by given  $z_0$ . The formal solution of this near the origin is made by series in the argument  $z_0 \mu^n$  which involve divisors of the form  $\mu^p - 1$ , for  $p = 1, 2, 3, \dots$ ; these divisors are 'small' if  $|\mu| = 1$ , with  $(2\pi i)^{-1} \log \mu = \lambda$  irrational, while if  $\lambda$  is rational the formal solution fails. However, *when  $\lambda$  is rational the solution is easily found by other means* [16].

(ii) We know that formal series solutions can be essentially divergent, and it is a challenging problem then to find an interpretation for them. Are they asymptotic expansions, in some sense, of analytic or perhaps pathological functions? I believe that this question will not be properly answered

unless we allow all the variables concerned to range over complex values. Here is an elementary example: In § 6 I have mentioned formal integrals of Hamiltonian systems (5) in the case where the  $\lambda$  are incommensurable. Regarding the commensurable case there is complete formal knowledge in the case  $m = 2$ : there does exist a series-integral  $G$  independent of  $H$  [17]. For example, if

$$H = x_1 y_1 - x_2 y_2 (1 - x_1 x_2) - b x_2^3,$$

then

$$G = x_1 x_2^2 y_2 + b x_1 x_2^4 \left( 1 + \frac{4}{3} x_1 x_2 + \frac{4 \cdot 5}{3^2} x_1^2 x_2^2 + \dots \right).$$

The series  $(\dots)$  is essentially divergent, but it is the asymptotic expansion of

$$\frac{3}{x_1^4 x_2^4} \exp \left( - \frac{3}{x_1 x_2} \right) \int_{x_1 x_2}^{-0} t^2 e^{3/t} dt,$$

valid for  $|\arg(-x_1 x_2)| < 3\pi/2$ ; and when  $G$  is modified by so replacing the series it is actually an integral. It would be rash to assert that this example is typical, but I would lay a large stake against anyone finding an equally non-typical example of interpretation of a divergent small-divisor series.

## References

- [1] Poincaré, H., *Acta Math.* 13 (1890), 1.  
Hadamard, J., *J. de Math.* [5] 4 (1898), 29. Considering a surface of negative curvature of genus exceeding 1, he shows that the set of those geodesics which do not run off ultimately to infinity is a pathological component of the set of all geodesics.  
Morse, M., *Trans. American Math. Soc.* 22 (1921), 84. Shows that a closed surface of negative curvature contains geodesics that are recurrent and of discontinuous type.  
Birkhoff, G. D., *Dynamical Systems*, American Math. Soc. Colloq. Pub. 9 (1927). For general reference.
- [2] Birkhoff, G. D., *Bull. Soc. Math. France* 40 (1912), 305, and *Dynamical Systems*, 197.
- [3] Lefschetz, S., *Differential equations; geometric theory* Interscience, New York, 1957.  
Stoker, J. J., *Non-linear Vibrations*, Interscience, New York, 1950.  
Andronow, A. A. and Chaikin, C. E., *Theory of Oscillations*, Princeton 1949.
- [4] Denjoy, A., *C.R. Acad. Sci. Paris* 194 (1932), 830 and *J. de Math.* [9] 11 (1932), 333.
- [5] Cherry, T. M., *Proc. London Math. Soc.* [2] 44 (1938), 175.
- [6] Cherry, T. M., *Proc. Cambridge Phil. Soc.* 22 (1924), 325. The result was known in 1920 by Birkhoff and guessed, but not proved, by Whittaker, E. T., *Proc. Roy. Soc. Edinburgh* 37 (1916), 95. Apparently, and surprisingly, it seems to have remained unnoticed by Poincaré.
- [7] The evidence as to essential divergence is usually indirect: that convergence would be inconsistent with properties otherwise established; for example, inconsistent with the pattern of periodic solutions near a given periodic solution.
- [8] Poincaré, H., *loc. cit.* and *Les Méthodes Nouvelles de la Mécanique Céleste*, Chap. 7.
- [9] Moser, J., *Comm. Pure Appl. Math.* 9 (1956), 673.
- [10] I believe that this phenomenon was discovered by Rayleigh or Maxwell, but the earliest reference I know is Stephenson, A. *Phil. Mag.* [6] 15 (1908), 233. The device of taking  $\delta$  to be piecewise constant was used by van der Pol, B. *Physica* (1925), 157

in discussing the linear approximation for the case  $x \sim 0$ . If the forced oscillation of  $A$  is horizontal instead of vertical, the downward position  $x = \pi$  of the pendulum becomes unstable when  $p > n$ , and there are then two inclined positions of stable equilibrium!

- [11] Artin, E., *Abh. Math. Sem. Hamburg* 2 (1923), 170.
- [12] Littlewood, J. E., *Acta Math.* 97 (1957), 267 and 98 (1957), 1.
- [13] Dulac, H., *J. École Polytech.* 9 (1904), 1.
- [14] Siegel, C. L., *Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl.* 2a (1952), 21.
- [15] Moser, J., *Comm. Pure Appl. Math.* 11 (1958), 257.
- [16] See for example Fatou, P., *Bull. Soc. Math. France* 47 (1919), 161 at Sec. 8.
- [17] Cherry, T. M., *Proc. Cambridge Phil. Soc.* 22 (1925), 510.

University of Melbourne