CHAPTER VI

APPLICATION OF POINCARE'S GEOMETRIC THEOREM

1. Periodic motions near generalized equilibrium ($m = 1$). Poincaré's last geometric theorem and modifications thereof* yield an additional instrument for establishing the existence of periodic motions. Up to the present time no proper generalization of this theorem to higher dimensions has been found, so that its application remains limited to dynamical systems with two degrees of freedom. It is our aim in this chapter to give some of the fundamental ideas involved in the theorem and its application.

It will be remembered that motion near to a periodic motion of a Hamiltonian or Pfaffian system, with m degrees of freedom and not involving the time explicitly, can be reduced to that of a similar system with only $m-1$ degrees of freedom but with an independent variable involved of period 2π. Here the periodic motion itself appears as generalized equilibrium. This reduction is accomplished by means of an analytic device (chapter IV, section 1).

In the present section we shall take up the question of the existence of motions with period $2k\pi$ near the position of generalized equilibrium for a single degree of freedom. We shall prove the existence of infinitely many such nearby periodic motions in the general stable case by a process of reasoning which, while not employing Poincaré's geometric theorem explicitly, is precisely that which establishes the theorem in certain simple cases. Later (section 3) these results are interpreted with reference to the original dynamical problem with two degrees of freedom.

* See my paper, *An Extension of Poincaré's Last Geometric Theorem*, Acta Mathematica, vol. 47 (1926).

Let the single pair of variables be $p, q,$ so that the Hamiltonian function H involves $p, q, t,$ being periodic in t of period 2π, and vanishes at the origin $p = q = 0$, together with its first partial derivatives, for all values of t.

If then (p_0, q_0) is any point near to the origin, there is a unique solution

$$p = f(p_0, q_0, t), \qquad q = g(p_0, q_0, t)$$

which for $t = 0$ takes on the values p_0, q_0, and which is analytic in p_0, q_0, t for t arbitrarily large and p_0, q_0 sufficiently small. Let p_1, q_1 denote the values of p, q after a complete period 2π. Evidently we have

$$p_1 = f(p_0, q_0, 2\pi), \qquad q_1 = g(p_0, q_0, 2\pi)$$

where f and g are analytic in p_0, q_0, and vanish with these variables.

In this way a transformation T is defined, of the same nature as the transformation of the surface of section obtained in the preceding chapter (section 10). For if we write $r = t$, the pair of Hamiltonian equations may be replaced by the equivalent set

$$\frac{dp}{dt} = -\frac{\partial H}{\partial q}, \qquad \frac{dq}{dt} = \frac{\partial H}{\partial p}, \qquad \frac{dr}{dt} = 1,$$

in which H is a function of p, q, r of period 2π in r. Here the manifold of states of motion is the three-dimensional p, q, r space in which the r axis represents a periodic motion, namely that corresponding to generalized equilibrium; it must not be forgotten that r is an angular variable. Now $\varphi = r = 0$ will serve as a surface of section according to our earlier work, although here we are limited to a certain neighborhood. A point $(p_0, q_0, 0)$ in this surface of section is taken along its stream line to $(p_1, q_1, 2\pi)$, i. e., to $(p_1, q_1, 0)$. Thus the transformation written above is indeed a transformation T of a surface of section S which is, however, only locally defined. Such 'local surfaces of section' can of course be constructed near a periodic motion in any dynamical problem by merely taking an element of surface which intersects but is not tangent to the corresponding stream line in the manifold of states of motion.

Now the fluid motion defined by the above three equations is that of an incompressible fluid, since the divergence of the right-hand members is 0. Consequently if we follow any tube of fluid made of sections of stream lines between the parallel planes $r = 0$ and $r = 2\pi$, which will move constantly with unit velocity in the r direction, we infer that the loss of volume at one base in time $\varDelta t$ is nearly $\sigma_0 \varDelta t$ (σ_0, area of first base), while the equal gain at the other is nearly $\sigma_1 \varDelta t$ (σ_1, area of second base). By allowing $\varDelta t$ to approach 0 we infer that $\sigma_0 = \sigma_1$.

Since σ_0 is an arbitrary area in $r = 0$, it is clear that T must be an area-preserving transformation of the variables p_0, q_0. This important property of T corresponds to a general property of the surface transformations associated with dynamical problems.

It is necessary now to state the conditions to be imposed upon the generalized equilibrium, with the aid of which the conclusion stated may be established.

We assume in the first place that the generalized equilibrium is of general stable type, and therefore completely stable. The significance of the normal form (chapter III, section 9) is that the solution may be written

$$p = p_0\, e^{M(p_0 q_0)t} + \Phi, \qquad q = q_0\, e^{-M(p_0 q_0)t} + \Psi,$$

in properly chosen conjugate variables p, q. Here Φ, Ψ are given as convergent power series in p_0, q_0 with initial terms of arbitrarily high degree $2\mu + 1$, and with all coefficients analytic functions of t; these series converge absolutely and uniformly for any fixed range of values for t, such as $|t| \leqq 2\pi$, when p_0, q_0 are small. The function M can be taken as a polynomial of degree not more than μ in the product $p_0 q_0$, with pure imaginary constant coefficients, of the form

$$\lambda + lp_0 q_0 + \cdots + sp_0^\mu q_0^\mu$$

with λ the multiplier. By the hypothesis of stability $\lambda/\sqrt{-1}$ is not rational, and in particular is not 0.

Our second assumption is that l is not 0. In case l vanishes but some other coefficient in M is not 0, essentially the same argument would apply. Thus the only case of failure is that in which the formal series M in the complete normal form reduces to a mere constant λ. This is a highly degenerate case, and actual examples can be constructed to show that an analogous conclusion cannot then be drawn.

The normal form gives a means of studying the transformation T. The property of the transformation T necessary for our present purposes is embodied in the following lemma whose proof is deferred to the next section:

LEMMA. For $l \neq 0$, upon suitable choice of the variables p, q, the positive quantity ε may be taken arbitrarily small, and then the integer n so large that any transformation T^ν ($\nu \leqq n$) takes the circle $r \leqq \varepsilon$ about the invariant point $r = 0$ into a region within the circle of radius 2ε, while the angular rotation effected by T^n increases from $n\lambda/\sqrt{-1}$ with r along any radial line for $r \leqq \varepsilon$, being at least 2π greater for $r = \varepsilon$ than for $r = 0$.

Let r, θ be polar coördinates and let (r_n, θ_n) denote the iterate of (r, θ) under T^n, where the rectangular coördinates p, q, the radius ε, and the integer n are selected as in the lemma. For any fixed θ_0 the difference $\theta_n - \theta_0$ will then increase from $n\lambda/\sqrt{-1}$ at $r = 0$ to a quantity at least 2π greater at $r = \varepsilon$. Hence there will be a unique solution of the equation

$$\theta_n - \theta_0 = 2k\pi$$

along a fixed radius vector, where $2k\pi$ is the least integral multiple of 2π exceeding $n\lambda/\sqrt{-1}$. Hence the analytic curve C given by this equation is met once and only once by each radius sector.

Now let us consider the image C_n of this curve under the transformation T^n; the curve C_n intersects C at some point Q, since if C_n is wholly within C or outside C, T^n would not be area-preserving in the original variables. The point Q is obtained from some point P, also on C, by the transformation T^n. Moreover P and Q have the same θ, by definition of C. Thus P and Q must coincide, and P is invariant under T^n. Since ε is arbitrarily small, we obtain the result stated:

In the case of generalized equilibrium of general stable type for a Hamiltonian problem with one degree of freedom ($l \neq 0$), *there exist infinitely many periodic motions in the vicinity.*

2. Proof of the lemma of section 1. Let us define $F(u)$ by the equation

$$l\,u\,F^2(u) = M(u) - \lambda.$$

It is clear that $F(u)$ is the square root of a real polynomial of degree $\mu - 1$ and constant term 1. If then we write further

$$\bar{p} = F(pq)p, \qquad \bar{q} = F(pq)q,$$

it is found that the above normal form for T is further simplified and may be written

$$\bar{p} = \bar{p}_0 e^{(\lambda + l\bar{p}_0\bar{q}_0)t} + \bar{\Phi}, \qquad \bar{q} = \bar{q}_0 e^{-(\lambda + l\bar{p}_0\bar{q}_0)t} + \bar{\Psi},$$

so that all of the terms of M except the first two disappear. When the associated real variables are introduced, and we let σ and s denote the real constants $2\pi\lambda/\sqrt{-1}$ and $2\pi l/\sqrt{-1}$ respectively, we obtain formulas defining T,

$$\begin{aligned} p_1 &= p_0\cos(\sigma + s r_0^2) - q_0 \sin(\sigma + s r_0^2) + P \quad (r_0^2 = p_0^2 + q_0^2), \\ q_1 &= p_0 \sin(\sigma + s r_0^2) + q_0 \cos(\sigma + s r_0^2) + Q, \end{aligned} \tag{1}$$

where P, Q are real power series in p_0, q_0 with initial terms of degree $2\mu + 1$ at least. It is apparent then that with these variables T is an ordinary rotation through a variable angle $\sigma + s r_0^2$, except for terms of order $2\mu + 1$ in the distance from the origin. It is such a choice of variables that will be adopted. If $l \neq 0$, we may take s as positive.

Hence we have, in some fixed neighborhood of the origin and for a fixed K.

$$(2) \qquad |P|,\ |Q| \leqq K r_0^{2\mu+1}.$$

From the above formulas we find at once

$$(3) \qquad r_1^2 = r_0^2 + R,$$

where the series R begins with terms of at least the degree $2\mu+2$. This gives, for the increment $\Delta r_0^2 = r_1^2 - r_0^2$,

$$(4) \qquad |\Delta r_0^2| \leqq L r_0^{2\mu+2}$$

where L is a fixed constant.

From (4) it is seen that r_n^2 increases less rapidly upon successive iteration than if

$$d r_n^2 / d n = L r_n^{2\mu+2}$$

But this yields

$$r_n = r_0 / (1 - L\mu r_0^{2\mu} n)^{1/2\mu}.$$

Hence r_n can only increase to twice the initial value r_0 after $n \geqq \nu$ iterations; where

$$L\mu\nu = (1 - 2^{-2\mu}) / r_0^{2\mu},$$

that is for n of at least the order of $r_0^{-2\mu}$. Likewise r_n can only decrease to half the initial value for n of the same order. These results may be combined in the form

$$(5) \qquad \frac{1}{2} \leqq r_n / r_0 \leqq 2 \qquad (n \leqq N r_0^{-2\mu}).$$

This is our first important conclusion.

Likewise since we have from (1)

$$\frac{q_1 p_0 - p_1 q_0}{p_1 p_0 + q_1 q_0} = \frac{(p_0^2 + q_0^2) \sin(\sigma + s r_0^2) + p_0 Q - q_0 P}{(p_0^2 + q_0^2) \cos(\sigma + s r_0^2) + q_0 Q + p_0 P},$$

it follows that we have

$$(6) \qquad \theta_1 = \theta_0 + \sigma + s r_0^2 + \Theta$$

where Θ is a power series in r_0 beginning with terms of degree 2μ at least, with coefficients of simple trigonometric type in θ_0. Moreover Θ is uniformly and absolutely convergent for r_0 sufficiently small, and its partial derivatives are given by derived series of a similar sort. This formula shows that for $r_0 = 0$, we have $\theta_n = \theta_0 + n\sigma$, while for $r_0 > 0$, the difference $\theta_n - \theta_0 - n\sigma$ can be made arbitrarily large by taking n sufficiently large but in the range (5).

By the aid of (3) and (6) we obtain

$$(7) \qquad \begin{cases} \left|\dfrac{\partial \varrho_1}{\partial \varrho_0} - 1\right| \leqq A\varrho_0^{\mu}, & \left|\dfrac{\partial \varrho_1}{\partial \theta_0}\right| \leqq A\varrho_0^{\mu+1}, \\[2ex] \left|\dfrac{\partial \theta_1}{\partial \varrho_0} - s\right| \leqq A\varrho_0^{\mu-1}, & \left|\dfrac{\partial \theta_1}{\partial \theta_0} - 1\right| \leqq A\varrho_0^{\mu}, \end{cases}$$

where we have written $\varrho = r^2$, and where A is a suitably chosen positive constant.

But the identities

$$\frac{\partial \varrho_n}{\partial \varrho_0} = \frac{\partial \varrho_n}{\partial \varrho_{n-1}} \frac{\partial \varrho_{n-1}}{\partial \varrho_0} + \frac{\partial \varrho_n}{\partial \theta_{n-1}} \frac{\partial \theta_{n-1}}{\partial \varrho_0},$$

$$\frac{\partial \theta_n}{\partial \varrho_0} = \frac{\partial \theta_n}{\partial \varrho_{n-1}} \frac{\partial \varrho_{n-1}}{\partial \varrho_0} + \frac{\partial \theta_n}{\partial \theta_{n-1}} \frac{\partial \theta_{n-1}}{\partial \varrho_0},$$

may be written

$$(8) \qquad \begin{aligned} u_n &= (1 + \epsilon_1) u_{n-1} + \epsilon_2 v_{n-1}, \\ v_n &= (s + \epsilon_3) u_{n-1} + (1 + \epsilon_4) v_{n-1}, \end{aligned}$$

in which we put

$$u_n = \partial \varrho_n / \partial \varrho_0, \qquad v_n = \partial \theta_n / \partial \varrho_0,$$

while, by (5) and (7), for $n \leqq N\varrho_0^{-\mu}$,

$$|\epsilon_1| \leqq 2^{\mu} A\varrho_0^{\mu}, \qquad |\epsilon_2| \leqq 2^{\mu+1} A\varrho_0^{\mu+1}, \qquad \epsilon_3| \leqq 2^{\mu-1} A\varrho_0^{\mu-}$$

$$|\epsilon_4| \leqq 2^{\mu} A\varrho_0^{\mu}.$$

These equations (8) enable us to determine u_n, v_n in succession for $n = 1, 2, \cdots$, with the initial conditions $u_0 = 1$, $v_0 = 0$.

Suppose now for a moment that the small terms be neglected in these equations (8). They then take the form

$$u_n = u_{n-1}, \qquad v_n = s u_{n-1} + v_{n-1},$$

whence, by elimination of u, we obtain

$$\Delta^2 v_n = v_{n+2} - 2 v_{n+1} + v_n = 0.$$

It is easily verified that the complete result of elimination yields similarly

$$(9) \qquad \Delta^2 v_n = \epsilon_5 \Delta v_n + \epsilon_6 v_n,$$

in which we have

$$(10) \qquad |\epsilon_5|, \; |\epsilon_6| \leqq B \varrho_0^{\mu-1} \qquad (n \leqq N \varrho_0^{-\mu})$$

within a small region about the origin. Furthermore, the initial conditions may be written

$$(11) \qquad v_0 = 0, \qquad \Delta v_0 = s + \epsilon_8^0,$$

where ϵ_8^0 denotes the value of ϵ_8 when $\varrho_{n-1}, \theta_{n-1}$ are replaced by ϱ_0, θ_0 respectively.

It is obvious that $v_1 = s + \epsilon_8^0$ is positive and thence, by use of (9), $v_2, v_3, \cdots$ are also positive for $n = 1, 2, \cdots$ until n becomes large if ϱ_0 is small enough, the approximate value of v_n being ns. We desire to obtain a more definite idea of the range of values for ϱ_0 and n for which v_n remains positive. During this range the angular variable θ_n increases with r_0, for a fixed angle θ_0.

Now the equation (9) is a homogeneous linear difference equation of the second order in v_n, and we are considering the particular solution satisfying (11). Evidently v_n will remain positive so long as Δv_n continues positive. A first question is then to determine the range of values of n throughout which both v_n and Δv_n necessarily remain positive. But the linear difference equation yields

$$\Delta^2 v_n \geqq -B\varrho_0^{\mu-1}[|\Delta v_n| + |v_n|],$$

so that clearly v_n and Δv_n diminish less rapidly than if

$$\Delta^2 v_n = -B\varrho_0^{\mu-1}[\Delta v_n + v_n],$$

while v_n and Δv_n remain positive. Thus v_n and Δv_n will remain positive for $n = 1, 2, \cdots$, at least as long as for the solution of the linear difference equation of the second order with constant coefficients

$$v_{n+2} - (2 - B\varrho_0^{\mu-1})\, v_{n+1} + v_n = 0$$

satisfying the initial conditions (11). But this solution is

$$v_n = (s + \varepsilon_3^0)\frac{e^{\alpha_1 n} - e^{\alpha_2 n}}{e^{\alpha_1} - e^{\alpha_2}} \tag{12}$$

where α_1, α_2 are defined by the equation

$$\alpha_i = \log\left(1 - \frac{1}{2}B\varrho_0^{\mu-1} \pm \sqrt{-\left(B\varrho_0^{\mu-1} - \frac{1}{4}B^2\varrho_0^{2\mu-2}\right)}\right) \qquad (i = 1, 2).$$

But v_n and Δv_n as thus determined will certainly remain positive until dv_n/dn vanishes, i. e.

$$e^{(\alpha_1 - \alpha_2)n} = \alpha_2/\alpha_1.$$

Since the leading term in $\alpha_1 - \alpha_2$ is clearly

$$2\sqrt{-B}\,\varrho_0^{(\mu-1)/2}$$

while the leading term in α_2/α_1 is -1, this relationship shows that n must be of the reciprocal order $\varrho_0^{-(\mu-1)/2}$.

Hence we infer that so long as $n \leqq N^* r_0^{-(\mu-1)}$ (compare with (5)), the angle θ_n will increase with r_0 for fixed θ_0 in the prescribed neighborhood.

The nature of the inequalities derived above makes it clear that we can select a value of r_0 so small, and then of n so large, that the conditions laid down in the lemma are satisfied.

3. Periodic motions near a periodic motion ($m = 2$). We have already seen (chapter IV, section 1) how the general Pfaffian system in which the time t does not appear explicitly admits of a reduction to a similar system with one less degree of freedom, provided that we are considering motions near a given periodic motion. In the reduced equations, however, an angular variable of period 2π appears in the differential equations, and the given periodic motion takes the form of generalized equilibrium.

In this section we propose to consider the periodic motions near a given periodic motion for the special Hamiltonian case ($m = 2$)

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \quad (i = 1, 2), \tag{13}$$

in which H is an analytic function of p_1, q_1, p_2, q_2, not involving t. However such a periodic motion admits of analytic continuation with variation of the energy constant $H = h$ (chapter V, section 9), and so is not isolated. Our aim then will be to consider only those nearly periodic motions which belong to the same value of h as the given periodic motion; this value may be taken to be $h = 0$.

The possibility of reduction to a Pfaffian case $m = 1$, combined with the results of the preceding section renders it highly probable at the outset that there will in general be infinitely many nearby periodic motions of long period, provided that the given periodic motion is of stable type.

In considering this question, we shall make the further assumption that the given Hamiltonian problem is associated with an ordinary Lagrangian problem whose principal function L

is quadratic in the velocities. If q_1, q_2 are the coördinates in this Lagrangian problem, the equations

$$p_i = \frac{\partial L}{\partial q_i'} \qquad (i = 1, 2)$$

serve of course to define the variables p_1, p_2.

Let $q_1 = q_1(t)$, $q_2 = q_2(t)$ be the equations yielding this periodic motion of period t, and consider the corresponding analytic curve in the q_1, q_2 plane.

Evidently we can introduce a modified system of coördinates $\bar{q}_1$, $\bar{q}_2$ such that $\bar{q}_2$ vanishes along the motion, while $\bar{q}_1$ increases by 2π as a point makes a circuit of the motion. For instance, if the curve of motion is without double points, it may be deformed into a circle concentric with the origin, in which case $\bar{q}_1$ and $\bar{q}_2$ may be taken as angle and radial displacement respectively. It is clear indeed that we may take $\bar{q}_1 = 2\pi t/\tau$ along the periodic motion. Of course such a change of variables from q_1, q_2 to $\bar{q}_1$, $\bar{q}_2$ does not affect the Lagrangian character of the dynamical problem, although the new principal function L is periodic of period 2π in the variable $\bar{q}_1$.

The corresponding Hamiltonian problem will have the form (13) in which H is periodic of period 2π in q_1, while for the periodic motion under consideration we have $q_1 = 2\pi t/\tau$, $q_2 = 0$. From the Hamiltonian equations in these variables, we have also along the periodic motion

$$2\pi/\tau = \partial H/\partial p_1, \qquad 0 = \partial H/\partial p_2.$$

It is obvious then that we may solve the equation $H = h$ for p_1 in the form

$$p_1 + K(q_1, p_2, q_2, h) = 0 \tag{14}$$

where K is a real, single-valued, analytic function of its four arguments, periodic in q_1 of period 2π. Furthermore we may regard h, q_1, p_2, q_2 as the dependent variables instead of p_1, q_1, p_2, q_2; we observe that (14) may be solved explicitly for h, since from the relation $H = h$, we derive

$$\frac{\partial H}{\partial p_1}\frac{\partial p_1}{\partial h} = 1$$

so that $\partial p_1/\partial h \neq 0$ along the motion. When these variables are used instead of p_1, q_1, p_2, q_2, the variational principle (chapter II, section 10) takes the form

$$\delta \int_{t_0}^{t_1} (-Kq_1' + p_2 q_2' - h)\, dt = 0, \tag{15}$$

which leads to the four equations

$$\frac{\partial K}{\partial h}\frac{dq_1}{dt} + 1 = 0, \quad \frac{\partial K}{\partial h}\frac{dh}{dt} + \frac{\partial K}{\partial p_2}\frac{dp_2}{dt} + \frac{\partial K}{\partial q_2}\frac{dq_2}{dt} = 0,$$

$$\frac{dq_2}{dt} = \frac{\partial K}{\partial p_2}\frac{dq_1}{dt}, \quad \frac{dp_2}{dt} + \frac{\partial K}{\partial q_2}\frac{dq_1}{dt} = 0.$$

From these equations we infer directly $h = \text{const.}$, which we know to be true of course.

Now it is evident that near the given periodic motion q_1 can serve as independent variable as well as t. If we eliminate t in the above equations, we find

$$\frac{dp_2}{dq_1} = -\frac{\partial K}{\partial q_2}, \quad \frac{dq_2}{dq_1} = \frac{\partial K}{\partial p_2}. \tag{16}$$

Here we are to set $h = 0$ in the function K, and K is periodic of period 2π in q_1. The given periodic motion corresponds to

$$p_2 = \varphi(q_1), \quad q_2 = 0$$

where φ is periodic of period 2π in q_1. These equations are clearly in Hamiltonian form ($m = 1$), with a generalized equilibrium point at the origin, at least after the simple modification

$$\bar{p}_2 = p_2 - \varphi(q_1), \; \bar{q}_2 = q_2,$$

in which we may take

$$\overline{K} = K - \varphi'(q_1) q_2 .$$

Conversely, if we have a solution of (16), we can determine t from the equation

$$\frac{dt}{dq_1} = -\frac{\partial K}{\partial h},$$

and obtain a solution of the original system when t is taken as independent variable. Thus (13) and (16) are equivalent.* Periodic motions near the given periodic motion for (13) correspond to motions near the origin of period $2k\pi$ for (16).

For a Hamiltonian problem (13) *which reduces to a generalized equilibrium problem* (16) *of stable type* ($l \neq 0$), *there will exist infinitely many periodic motions in the vicinity of the given periodic motion, making in general many circuits of that motion before re-entering.*

This result is of course obtained as the direct application of section 1 for the reduced problem.

4. Some remarks. The general conclusion which appears in consequence of the preceding sections is that, for a given value of the energy constant, there exist *in general* periodic motions in the vicinity of a periodic motion of stable type, at least when the dynamical system has two degrees of freedom and is of ordinary type. The fact that there may exist isolated periodic motions of stable type, even for dynamical systems with two degrees of freedom, may be brought out by means of the following elementary example.

Let us write

$$H = \frac{1}{2} k^2 (p_1^2 + q_1^2) + \frac{1}{2} l^2 (p_2^2 + q_2^2),$$

where the quantities k, l are incommensurable with one another, of which the general solution is

* For the reduction employed, cf. Whittaker, *Analytical Dynamics*, chap. 12.

$$p_1 = A\cos kt + B\sin kt, \quad q_1 = -A\sin kt + B\cos kt,$$
$$p_2 = C\cos lt + D\sin lt, \quad q_2 = -C\sin lt + D\cos lt.$$

The energy constant h is defined by the relation

$$H = \frac{1}{2}k^2(A^2+B^2) + \frac{1}{2}l^2(C^2+D^2) = h.$$

The only periodic solutions are the two analytic families

$$p_1 = q_1 = 0 \text{ and } p_2 = q_2 = 0,$$

all of which are of stable type. For an assigned value of the energy constant, there are only two such periodic motions; thus all periodic motions of the second family with assigned A^2+B^2 represent the same closed curve in the three-dimensional manifold $H=h$ in four-dimensional p_1, q_1, p_2, q_2 space. If the transformation T be set up for this case as in the preceding section 2, it is found to be essentially a rotation through an angle incommensurable with 2π, and so to correspond precisely to the highly degenerate case there excluded from consideration.

A first question as to a possible generalization of the above results in the case $m=2$ is the following: Suppose the origin is a point of generalized equilibrium of general stable type for a given differential system which is, in addition, completely stable; if the constant l is not 0, does it follow that there will always exist infinitely many periodic motions in the vicinity of the origin?

It seems to me very doubtful that the answer is in the affirmative. In the preceding argument the area-preserving property played a vital part. For this more general completely stable type, there is no reason to believe that this property continues to hold, even in the Pfaffian case.

The example can be generalized so as to indicate a preliminary necessary requirement if the conclusion that there are infinitely many periodic motions near a given stable periodic motion is to hold for Hamiltonian systems with more than two degrees of freedom.

In fact, consider the case of a dynamical system

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \qquad (i = 1, \cdots, m)$$

in which H is a function of the m products $p_1 q_1, \cdots, p_m q_m$ with p_i, q_i conjugate imaginary variables, namely

$$H = \sum_{j=1}^{m} c_j p_j q_j + \frac{1}{2} \sum_{j,k=1}^{m} d_{jk} p_j q_j p_k q_k.$$

Here the coefficients c_i, d_{ij} are periodic of period τ in t. The origin is a point of generalized equilibrium with multipliers

$$\lambda_i = \int_0^{\tau} c_i(t)\, dt \qquad (i = 1, \cdots, m),$$

so that it will be of general stable type if these m quantities and $2\pi\sqrt{-1}/\tau$ have no linear commensurability relations. If for the sake of brevity we write

$$x_i = \int_0^t \left[c_i + \sum_{j=1}^{m} d_{ij} p_j^0 q_j^0 \right] dt,$$

the general solution is at once found to be

$$p_i = p_i^0 e^{-x_i}, \quad q_i = q_i^0 e^{x_i} \quad (i = 1, \cdots, m).$$

Moreover if we write

$$C_i = \int_0^{\tau} c_i\, dt, \quad D_{ij} = \int_0^{\tau} d_{ij}\, dt,$$

the condition that the solution is periodic of period $k\tau$ is

$$C_i + \sum_{j=1}^{m} D_{ij} p_j^0 q_j^0 = 2\pi k_i \sqrt{-1}/k \qquad (i = 1, \cdots, m),$$

where $k_1, \cdots, k_m$ are integers. But these form m linear, non-homogeneous, algebraic equations in $p_i^0 q_i^0$, $(i = 1, \cdots, m)$, which can be solved if the determinant $|D_{ij}|$ is not 0. Moreover by making the ratios k_i/k small, the periodic motion can be taken near the origin. On the other hand if $|D_{ij}| = 0$, such a determination will be impossible in general.

In the particular case when the c_i's and d_{ij}'s are constants, the system appears in complete normal form, and the c_i's are the multipliers, while the d_{ij}'s are invariants analogous to l in the case $m = 1$.

Hence at least the condition $|d_{ij}| \neq 0$ must be imposed in the case $m > 1$, as analogous to the condition $l \neq 0$ in the case $m = 1$, if an infinitude of nearly periodic motions is to be anticipated in all cases.

Any generalization must of course take proper account of the uniform analytic integrals (such as the energy integral) which exist. In fact, if there are k of these integrals which are independent, the given stable periodic motion will admit of k-fold analytic continuation. Evidently it is not such periodic motions of the same analytic family as the given motion which interest us, but rather nearby periodic motions for the same values of the constants of integration as the given periodic motion, and making many circuits of it in a period. 6

5. The geometric theorem of Poincaré.* Poincaré showed that the existence of an infinite number of periodic orbits in the restricted problem of three bodies and other dynamical problems would follow at once from a certain geometric theorem to which the lemma of section 1 is intimately related.

For convenience we shall first state:

POINCARÉ'S THEOREM. Given a ring $0 < a \leqq r \leqq b$ in the r, θ plane (r, θ being polar coördinates), and a one-to-one, continuous, area-preserving transformation T of the ring into itself, which advances points on $r = a$ and regresses points on $r = b$. Then there will exist at least two points of the ring invariant under T.

* This section is essentially the same as section 34 of my paper, *Dynamical Systems with Two Degrees of Freedom*, Trans. Amer. Math. Soc., vol. 18 (1917).

We will indicate briefly the proof of this theorem.

Let us take $x = \theta$, $y = r^2$ as the rectangular coördinates of a point in the x, y plane. The ring then appears as a strip $a^2 \leqq y \leqq b^2$. The transformation T of this strip advances points of the boundary $y = a^2$ to the right, and moves points on $y = b^2$ to the left. Moreover T is area-preserving in the x, y plane (for we have $2\,r\,dr\,d\theta = dx\,dy$), and displaces any two points which have the same ordinate and whose abscissas differ by a multiple of 2π in the same way.

Let us combine T with a further transformation T_ε which effects a translation of the x, y plane in the direction of the y axis through a distance $\varepsilon > 0$. The transformation T followed by T_ε yields an area-preserving transformation $T\,T_\varepsilon$ which shifts the given strip into the strip $a^2 + \varepsilon \leqq y \leqq b^2 + \varepsilon$.

Suppose if possible that there exists no invariant point of T. There exists then a positive quantity d such that all points are displaced at least a distance d by the transformation T. Choose ε less than d.

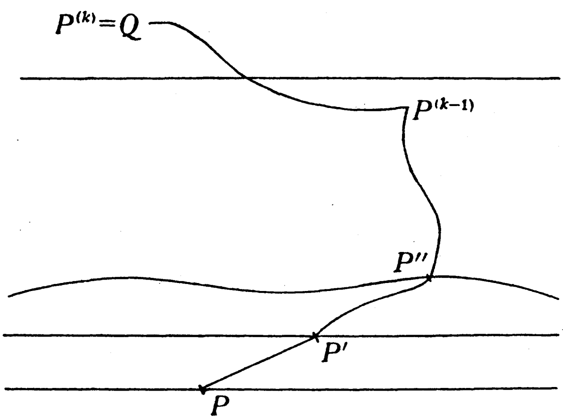
Consider now the narrow strip $a^2 \leqq y \leqq a^2 + \varepsilon$. By the transformation $T\,T_\varepsilon$ the lower edge of this strip is carried into the upper edge and the strip is carried into a second strip lying wholly above the first one save along the common edge. By a repetition of the transformation $T\,T_\varepsilon$ the second strip goes into a third, and so on.

By a continuation of this process, a series of strips is obtained forming consecutive strata. Each of these strata is unaltered by a shift of 2π to the right. This follows from the fact that T and T_ε are single-valued over the ring.

The images of these strata on the ring are a set of closed strata about the ring, all having equal area of course since $T\,T_\varepsilon$ is an area-preserving transformation in the r, θ as well as in the x, y plane. Consequently some one of the strata on the infinite strip, say the k-th. must overlap the upper edge $y = b^2$.

In the x, y plane let Q be a point of the upper edge of the k-th stratum for which y is a maximum. Let P be the point of $y = a^2$ from which Q is derived by k-fold repetition of TT_ε, and let P', P'', $\cdots$, $P^{(k)} = Q$ denote the successive images of P under the iteration of TT_ε. Draw the straight line PP' which will obviously lie on the first stratum. The successive images of this line, PP', $P'P''$, $\cdots$, $P^{(k-1)}P^{(k)}$ will lie in the successive strata, and will have no points in common except that successive arcs have an end-point in common. Thus we get a single arc PQ, made up of all these lines, which is without double points.

Consider now a vector LL' drawn from a point L to its image L' under TT_ε, of which the initial point moves from



P to $P^{(k-1)}$ along the line PQ. The angle which this vector makes with the positive direction of the x axis at the outset may be taken to be a positive acute angle, since the image P' of P lies to the right of and above P. When L has varied to its final position $P^{(k-1)}$, the same angle lies in the second or third quadrant, since $P^{(k)}$ lies to the left of $P^{(k-1)}$, by the hypothesis of the theorem.

Our construction of the successive arcs PP', $P'P''$, $\cdots$, renders it apparent that as L moves from P to $P^{(k-1)}$, its image L' moves along the same curve from P' to Q. Therefore we see at once from the figure that LL' has rotated through the least positive angle from the first direction to the second. If L' be moved further in a vertical direction from Q up to $y = b^2 + \epsilon$ the same statement will still hold of the standard curve described by L provided that ϵ is small, since Q lies at most ϵ above $y = b^2$.

Suppose now that L moves in any manner from a point of $y = a^2$ to a point of $y = b^2$ in the given strip. The transformation TT_ϵ leaves no points of this region invariant, so that the point L will never coincide with L'. In the initial position for L on $y = a^2$ the angle made by LL' lies in the first quadrant. In the final position it lies in the second or third quadrant. But the total variation of angle during the variation of L has been seen to be through the least positive angle in a special case. Since any one path of L from $y = a^2$ to $y = b^2$ can be varied continuously into any other, the same must be true always.

Now let ϵ approach zero. As ϵ becomes smaller the vector LL' continues to have a definite direction, since no invariant points under TT_ϵ are present. By a limiting process we infer that, for the transformation T, the angular variation of LL' is through the least positive angle consistent with its initial and final directions. It should be observed that for L on $y = a^2$ the direction of LL' is that of the positive x axis, while on $y = b^2$ the direction is that of the negative x axis.

Consider now the inverse transformation T^{-1} which is of the same type as T although it moves points on $y = a^2$ to the left, and points on $y = b^2$ to the right. By an entirely analogous argument to that given above we are led to infer that if a vector $LL^{(-1)}$ with end-point $L^{(-1)} = T^{-1}(L)$ has its initial point L varied from a point of $y = a^2$ to a point of $y = b^2$, the total angular variation will be the least negative angle consistent with its initial and final positions.

But the total rotation of $LL^{(-1)}$ is precisely the same as that of the oppositely directed vector $L^{(-1)}L$ which joins a point $L^{(-1)}$ of $y = a^2$ to its image L under T.

Hence by our earlier result the total angular variation of $L^{(-1)}L$ must also be the least positive angle consistent with the two positions. Thus we have been led to a contradiction, so that there must exist at least one invariant point.

To prove that there are at least two invariant points we may adopt the method used by Poincaré.

Let the point L describe the fundamental rectangle

$$0 \leqq x \leqq 2\pi, \qquad a^2 \leqq y \leqq b^2,$$

in the x, y plane, in a positive sense. It is obvious that the total rotation of the vector LL' is 0 over this circuit, since there is no rotation along $y = a^2$ or $y = b^2$, and the rotations along $x = 0$ and $x = 2\pi$ are the negatives of one another. But around a simple invariant point the rotation is $\pm 2\pi$. Hence, according to obvious reasoning, there will either be at least two simple invariant points with rotations $+2\pi$ and -2π, or there will be at least one multiple invariant point. As a matter of fact, there will always be at least two geometrically distinct invariant points.*

6. The billiard ball problem.† In order to see how the theorem of Poincaré and its generalization can be applied, we will consider first a special but highly typical system of this sort, namely that afforded by the motion of a billiard ball upon a convex billiard table. This system is very illuminating for the following reason: Any Lagrangian system with two degrees of freedom is isomorphic with the motion of a particle on a smooth surface rotating uniformly about a fixed axis and carrying a conservative field of force with it.‡ In particular if the surface is not rotating and if the field of force is lacking, the paths of the particle will be geodesics. If the surface is now flattened to the form of a plane convex curve C, the 'billiard ball problem' results.

* See my paper, *An Extension of Poincaré's Last Geometric Theorem*, Acta Mathematica, vol. 47 (1926).

† The sections 6–9 are taken from my paper *On the Periodic Motions of Dynamical Systems*, about to appear in the Acta Mathematica.

‡ See my paper, *Dynamical Systems with Two Degrees of Freedom*, Trans. Amer. Math. Soc., vol. 18 (1917), section 7. It is assumed that the Lagrangian principal function is quadratic in the velocities.

But in this problem the formal side, usually so formidable in dynamics, almost completely disappears, and only the interesting qualitative questions need to be considered. If C is an ellipse an integrable problem results, namely the limiting case of an ellipsoid treated by Jacobi.

In the billiard ball problem one can arrive at the existence of certain periodic motions by direct maximum-minimum methods. As of interest in itself I wish to show how this can be done. Results which are being obtained by Morse (see chapter V, section 8) indicate that the scope of these methods, already developed to some extent by Poincaré, Hadamard, Whittaker and myself, can be further extended. Thus the power of such maximum-minimum considerations in the billiard ball problem is likely to prove typical of the general case.

The longest chord of the boundary C of the billiard table, when traversed in both directions, evidently yields one of the simplest periodic motions. The billiard ball moving along this chord strikes the curved boundary at right angles and recoils along it in the opposite direction. If we seek to vary this chord continuously, while diminishing its length as little as possible, so as finally to interchange its two ends, there will be an intermediate position of least length, which will be the chord crossing C where C is of least breadth. Detailed computation of the slightly perturbed motions indicates that the first of these two periodic motions is unstable, while the second may be stable, or unstable.

Next we ask for the triangle of maximum length inscribed in C. Evidently at least one such triangle will exist, and can have no degenerate side of length 0. At each of its vertices the tangent will of course make equal angles with the two sides passing through the vertex. Hence a 'harmonic triangle' is obtained which will correspond to two distinct motions, one for each of the two possible senses of description.

Moreover, if we seek to vary this triangle continuously, not changing the order of its vertices and diminishing the perimeter as little as possible, so as finally to advance the vertices cyclically, we discover a second harmonic triangle, also corresponding to two periodic motions.

In this way the existence of two harmonic n-sided polygons which make $k \leqq n/2$ circuits of the curve C (k prime to n) can be proved. The two periodic motions corresponding to the polygon of maximum type will be unstable; while the others of minimax type may be stable or unstable.

In the case of a circular boundary the totality of regular inscribed polygons (simple or cross) form the harmonic polygons.

7. The corresponding transformation T. We propose next to set up a ring transformation T associated with the billiard ball problem, and to show how the geometric theorem of Poincaré in its first form leads to the conclusion deduced above. The reduction to a ring transformation is of fundamental theoretic importance, quite aside from the relation to the question of periodic motions. It should be noted also that, in the cases of most interest like the restricted problem of three bodies, the method of reduction to a ring transformation and application of the geometric theorem of Poincaré is available for the treatment of the periodic motions, while the maximum-minimum method has not as yet been shown to be applicable.

To begin with we suppose the length of C to be 2π and to be measured from a fixed point to a variable point P by an angular coördinate φ.

At P, taken as the point of projection of the billiard ball, let θ denote the angle between the positive direction of the tangent and the direction of projection. The variable θ varies between 0 and π only. These coördinates θ, φ suffice to represent all possible states of projection unambiguously. If φ be taken as an angular coördinate in the plane, while θ, augmented by a constant, say π, be taken as a radial coördinate, the set of values (θ, φ) is represented on a ring bounded by concentric circles of radii π and 2π respectively, namely the circles $\theta = 0$ and $\theta = \pi$.

Consider now a definite state of projection at P with given θ, φ. The billiard ball leaves the table at P, to strike it again at P_1, there to be projected in a state (θ_1, φ_1), say, and so forth indefinitely. If C is an analytic curve, as we assume it to be, the correspondence between θ, φ and θ_1, φ_1 is evidently one-to-one and analytic within the ring. When θ is nearly 0 or π, the ball is projected at a slight angle to the edge, and strikes it again at a nearby point with θ nearly 0 or π as the case may be. Hence the points on the bounding circles correspond to themselves with $\theta_1 = \theta$, $\varphi_1 = \varphi$.

One further remark needs to be made about the correspondence along the two boundaries of the ring. If we think of each point (θ, φ) as being carried into (θ_1, φ_1) by a transformation or deformation of the ring, this transformation T will effect a certain number of complete rotations of the inner circle, and also of the outer circle, since the points of these boundaries are invariant as just seen. We may arbitrarily regard the inner circle as having undergone no rotation, but the same will not then be true of the outer circle, which can at once be shown to have undergone a single complete revolution in the positive sense. For let the projection angle θ, for a given point P and corresponding fixed φ, vary from 0 to π. It is obvious that then θ_1 will increase from 0 to π while φ increases by 2π since the point P_1 makes a complete circuit of P in a positive sense. In other words, the transformation T takes radial segments across the ring into curves starting at the same point of the inner circle, but winding around the ring just once while crossing it. Hence the outer boundary has undergone a single positive revolution under the transformation T.

Suppose now that we have a periodic motion, for example that corresponding to one of the harmonic triangles taken in a positive sense. It is evident that the transformation T of the ring takes the point of the ring representing the state of projection at the first vertex into that at the second; and likewise takes the state for the second vertex into that for the third, and that for the third vertex into the first. Thus when T is applied, the triple of points on the ring is cyclically advanced, and each point of the triple is unaltered by the application of the third iterate T^3 of T. Conversely, to any triple with this property, or to any point invariant under T^3, together with its images under T and T^2, corresponds a motion belonging to a harmonic triangle. Evidently then, from considerations advanced earlier, there are at least four such triples. It is obvious that there can be no invariant points under T itself because φ is increased but by less than 2π.

In this way the search for harmonic polygons and the allied periodic motions in the billiard ball problem resolves itself into the determination of sets of distinct points $P_1, \cdots, P_n$ cyclically advanced by T, so that in general $T^n(P_i) = P_i$. More generally, each and every interesting property of the motion of the billiard ball is mirrored in a corresponding property of the transformation T. Thus the dynamical problem is effectively reduced to that of a particular transformation of a circular ring into itself.

8. Area-preserving property of T. There is a further property of the transformation

$$T: \quad \theta_1 = f(\theta, \varphi), \quad \varphi_1 = g(\theta, \varphi),$$

which plays a fundamental part in applying the geometric theorem of Poincaré: the double integral $\int\int \sin\theta \, d\theta \, d\varphi$ taken over any area σ of the ring has the same value as over the image of σ under T and its iterates. This is essentially an area-preserving property in modified coördinates.

Before passing to the entirely elementary proof of this fact, one immediate application may be cited in justification of the earlier statement as to the fundamental theoretic importance of the ring transformation. Since the integral evaluated over $\sigma_1, \sigma_2 \cdots$, has the same value, and since its value over the entire ring is finite, being 4π, some two of the images σ_i and σ_j overlap. Employing the inverse transformation we infer that σ_{i-1} and σ_{j-1} also overlap, and thus finally that σ_{i-j} and σ_0 overlap $(i > j)$. But, interpreted for the billiard ball problem, this means that the ball can be projected very nearly with arbitrary position and direction to return subsequently to nearly the same position and direction. As elaborated by Poincaré,* this chain of reasoning leads to the conclusion that the 'probability' is unity that an arbitrary motion returns infinitely often to the neighborhood of its initial state. He called this property of the dynamical system 'stability in the sense of Poisson'.

The proof that the double integral is invariant depends on an explicit evaluation of the determinant

$$J = \frac{\partial \theta_1}{\partial \theta} \frac{\partial \varphi_1}{\partial \varphi} - \frac{\partial \theta_1}{\partial \varphi} \frac{\partial \varphi_1}{\partial \theta}.$$

In fact, if $\int\int M(\theta, \varphi)\, d\theta\, d\varphi$ is invariant we have

$$\int\int M(\theta_1, \varphi_1)\, d\theta_1\, d\varphi_1 = \int\int M(\theta, \varphi)\, d\theta\, d\varphi$$

where the variables θ_1, φ_1 range over the region σ_1, just as θ, φ do over σ. But according to the fundamental theorem for change of variables, the change of variables T gives the integral on the left the form

$$\int_\sigma\int M(\theta_1, \varphi_1)\, J\, d\theta\, d\varphi.$$

Comparing this expression and the integral on the right, which are both integrals over the same arbitrary region σ, we deduce the functional relation

$$M(\theta_1, \varphi_1)\, J = M(\theta, \varphi)$$

as the necessary and also sufficient condition for invariance. Hence to establish that $\int\int \sin\theta\, d\theta\, d\varphi$ is invariant, we need only prove

$$J = \sin\theta / \sin\theta_1.$$

Let

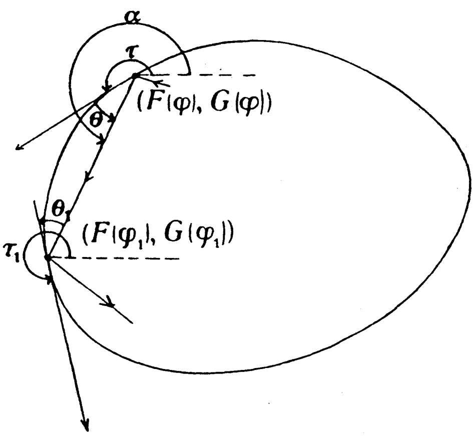
$$x = F(\varphi), \qquad y = G(\varphi)$$

* See his *Méthodes nouvelles de la Mécanique céleste,* vol. 3, chap. 26.

be the equations of C in rectangular coördinates, so that, if τ denotes the angle between the positive x axis and the positive tangential direction at a point of C, we have

$$\tau = \tan^{-1}\frac{G'(\varphi)}{F'(\varphi)}.$$

Similarly let τ_1 denote the like angle at the transformed point, which will be given by the same expression save that φ is replaced by φ_1. Finally let α designate the angle between the positively directed axis and the direction of initial projection (figure).



It is evident that the following two relations will hold

$$\theta = \alpha - \tau,$$

$$\theta_1 = \tau_1 - \alpha.$$

Substituting in the above value for τ and the analogous value for τ_1, and also substituting in for α the value

$$\tan^{-1}\frac{G(\varphi_1) - G(\varphi)}{F(\varphi_1) - F(\varphi)},$$

evident by inspection, we obtain the explicit formulas

$$T: \begin{cases} \theta = \tan^{-1}\dfrac{G(\varphi_1) - G(\varphi)}{F(\varphi_1) - F(\varphi)} - \tan^{-1}\dfrac{G'(\varphi)}{F'(\varphi)} = L(\varphi, \varphi_1), \\ \theta_1 = \tan^{-1}\dfrac{G'(\varphi_1)}{F'(\varphi_1)} - \tan^{-1}\dfrac{G(\varphi_1) - G(\varphi)}{F(\varphi_1) - F(\varphi)} = M(\varphi, \varphi_1). \end{cases}$$

These two equations define the transformation T from (θ, φ) to (θ_1, φ_1). Taking differentials, we find

$$d\theta = L_\varphi\, d\varphi + L_{\varphi_1}\, d\varphi_1, \qquad d\theta_1 = M_\varphi\, d\varphi + M_{\varphi_1}\, d\varphi_1,$$

whence at once

$$d\theta_1 = \frac{M_{\varphi_1}}{L_{\varphi_1}}\, d\theta + \left(M_\varphi - \frac{M_{\varphi_1} L_\varphi}{L_{\varphi_1}}\right) d\varphi,$$

$$d\varphi_1 = \frac{1}{L_{\varphi_1}}\, d\theta - \frac{L_\varphi}{L_{\varphi_1}}\, d\varphi.$$

This gives

$$J = -\frac{M_\varphi}{L_{\varphi_1}} = -\frac{[F(\varphi_1) - F(\varphi)]\, G'(\varphi) - [G(\varphi_1) - G(\varphi)]\, F'(\varphi)}{[F(\varphi_1) - F(\varphi)]\, G'(\varphi_1) - [G(\varphi_1) - G(\varphi)]\, F'(\varphi_1)}.$$

But $F(\varphi_1) - F(\varphi)$, $G(\varphi_1) - G(\varphi)$ are proportional to $\cos \alpha$, $\sin \alpha$ respectively, while also

$$F'(\varphi) = \cos \tau, \quad G'(\varphi) = \sin \tau, \quad F'(\varphi_1) = \cos \tau_1, \quad G'(\varphi_1) = \sin \tau_1,$$

so that finally we obtain

$$J = \frac{\sin(\alpha - \tau)}{\sin(\tau_1 - \alpha)} = \frac{\sin \theta}{\sin \theta_1}$$

as was stated.

9. Applications to billiard ball problem. As has been seen, there are no points of the ring which are invariant under T. On the other hand consider T^2 followed by a rotation of the θ, φ plane through an angle -2π, which we designate by R_{-1}. The resultant transformation of the ring admits the same area integral as T of course, but advances the points of the outer circle by an angle 2π, and those of the inner circle by an angle -2π of opposite sign. These are the two conditions essential for the application of Poincaré's geometric theorem. Hence $T^2 R_{-1}$ (the compound transformation) possesses two invariant points. This means that T^2 has two geometrically distinct invariant points of oppositely signed indices,* although these correspond to an increase of 2π for φ.

* See my paper *An Extension of Poincaré's Last Geometric Theorem*, Acta Mathematica, vol. 47 (1926). By the index of an invariant point is meant the number of positive rotations of a line joining a point P to its image P_1, when P makes a small positive circuit of the invariant point.

To prove this assertion in detail, we may let P be one such invariant point, such that T^n increases φ by $2k\pi$. If the n points

$$P,\ T(P), \cdots, T^{n-1}(P)$$

are not distinct, let $T^m(P) = P$, $(m \leqq n-1)$, and suppose that φ is increased by $2j\pi$. By combination of the two symbolic equations $T^m(P) = P$, $T^n(P) = P$, we obtain $T^d(P) = P$ where d ($\neq 1$ of course) is the greatest common divisor of m and n. Thus P is invariant under T^d. Suppose that under T^d, the φ of P increases by $2f\pi$. From the equation $T^n = T^{qd}$, we see that T^n will then increase the φ of P by $2qf\pi$, so that $k = qf$. Thus k and n would possess a common factor q, contrary to hypothesis.

Not only are the n points distinct but they have the same index. Hence there will be a point φ invariant under T^n and with oppositely signed index. This, with its images under successive powers of T, will necessarily be distinct from the points of the set generated by P, and leads to a second distinct series of n points.

Hence we obtain for every $n > 2$ and every relatively prime $k \leqq n/2$, two geometrically distinct, harmonic polygons with n sides and making k circuits of the curve C. Corresponding to these there will be of course four periodic motions. We shall not attempt to develop here the characteristics as to type of stability and instability, dependent upon the sign of the index.

It is worthy of observation that the method of sections 2, 3 is evidently applicable here to show that there exist infinitely many periodic motions lying in the vicinity of any periodic motion of general stable type if the constant l is not 0.

We shall indicate in particular how this same method seems to apply to the limiting type of periodic motion in which the billiard ball is rolled around the table.

For this purpose it is essential to examine the explicit formulas given for T in the case when θ is small. A direct computation leads to the result

If P is such an invariant point, so is $T(P)$ of course but with the same index. Thus we get two pairs of points, say

$$P,\ T(P),\ Q,\ T(Q),$$

all four distinct. These evidently correspond to the two fundamental periodic motions.

For the application of the theorem of Poincaré to the periodic motions of more complicated type, it is necessary to take account of the fact that every such motion is associated with a distinct second such motion, obtained by reversing the direction of description, although these motions have the same index. However one of these motions increases φ by $2k\pi$, the other increases it by $2(n-k)\pi$. By only considering invariant points of T^n for which φ increases by $2k\pi$, $(k \leqq n/2)$, we clearly obtain each harmonic n-sided polygon only once. It may be noted in passing that this pairing of motions in the billiard ball problem is fully reflected in the fact that T is a product of two involutory transformations; it was the same special property of the ring transformation in the restricted problem of three bodies which enabled me to prove the existence of infinitely many symmetric periodic orbits.*

Now turn to the invariant points of the compound transformation $T^n R_{-k}$ where R_{-k} denotes k-fold rotation through the angle -2π. The rotations on the outer and inner circles are clearly

$$2(n-k)\pi \quad \text{and} \quad -2k\pi,$$

which will be of opposite sign. Thus we can infer the existence of at least two geometrically distinct series of points

$$P,\ T(P), \cdots, T^{n-1}(P), \qquad Q,\ T(Q), \cdots, T^{n-1}(Q)$$

such that $T^n(P) = P$, $T^n(Q) = Q$, while φ has been increased by $2k\pi$; it is assumed that k and n are relatively prime.

* See my paper *The Restricted Problem of Three Bodies,* Rendiconti di Palermo, vol. 39 (1915), in particular, section 14.

$$\theta_1 - \theta = \frac{2}{3}\frac{k'}{k^2}\theta^2 + l\,\theta^3 + \cdots,$$

$$\varphi_1 - \varphi = \frac{2}{k}\theta - \frac{4k'}{3k^3}\theta^2 + m\,\theta^3 + \cdots,$$

where the function $k(\varphi)$ denotes the curvature of C at the point with given φ, and where the functions $l, m, \cdots$ depend on φ only. Proceeding entirely formally and replacing $\theta_1 - \theta$ and $\varphi_1 - \varphi$ by $\Delta\theta$ and $\Delta\varphi$ respectively, we obtain the approximate differential equation

$$\frac{d\theta}{d\varphi} = \frac{1}{3}\frac{k'}{k}\theta,$$

which gives by integration

$$\theta = \theta_0 k^{1/3}(\varphi).$$

Here θ_0 is the value of θ for a point of curvature unity. This result indicates that, to a first approximation, the curve $\theta = \theta_0 k^{1/3}(\varphi)$ near the inner boundary $\theta = 0$ of the ring is nearly invariant under T, and can probably be modified by the inclusion of higher order terms so as to be still more nearly invariant. Evidently the limiting periodic motions formed by C are to be regarded as analogous to stable periodic motions on this account.

Also if the variable n represents the number of iterations, we have the appreximate differential equation

$$\frac{d\varphi}{dn} = 2\,\theta_0 k^{-2/3},$$

whence by integration

$$n = \int k^{2/3}\, d\varphi / 2\,\theta_0.$$

It follows that φ will increase by more than 2π along the approximately invariant curve if $n\,\theta_0$ exceeds $\pi K^{2/3}$ where K denotes the maximum curvature of C.

It thus appears as highly probable that the lemma of section 2 is applicable and that there exist infinitely many periodic motions uniformly near to C.

10. The geodesic problem. Construction of a transformation TT^*. We turn our attention next to the problem afforded by the geodesics on an analytic surface. In order to obtain as concrete results as possible, we shall restrict the surface S to be closed and convex, although it is evident that these limitations are not altogether essential for the argument which follows.

We have already established the existence of at least one closed geodesic g of minimax type (chapter V, section 6), which we shall assume to be without double points. Our first assertion is that there exists a positive quantity L so great that any geodesic arc of length exceeding L intersects g at least once (or falls along g). In the contrary case there would exist a sequence of geodesic arcs g_n of length L_n with $\lim L_n = +\infty$, such that each arc g_n does not meet g. This fact would require that for L_n large enough no part of g_n is nearly coincident with g, since nearby geodesics meet a closed geodesic of minimax type in a succession of points separated by arcs of limited length; more precisely, it may be proved that if P is any point of g and P'' is its second conjugate point in the sense of the calculus of variations, then the arc PP'' constitutes at least one complete circuit of g.†

Hence if we let P_n be the mid-point of q_n, the sequence of points P_n will have a limiting point P not on g, and the geodesic q_n will have at least one limiting direction at P, such that the complete geodesic h through P in this direction fails to meet g and indeed nowhere approaches g.

Consider now that part of the surface S (which is divided by g in two parts) upon which h lies, and in particular the part of S lying between g and h. One boundary of this region s is g, of geodesic curvature 0 everywhere, while the other boundary γ consists of part or all of h, and of its limit points.

Let N_1 and N_2 be two nearby accessible points of the

† For a proof see my paper, *Dynamical Systems with Two Degrees of Freedom*, Trans. Amer. Math. Soc., vol. 18 (1917), in particular section 19.

boundary γ so obtained, so that a short curvilinear arc $N_1 A N_2$ exists in the region s with all of its interior points not on the boundary. Consider also the short geodesic arc $N_1 G N_2$. Then $N_1 A N_2 G N_1$ delimits a region σ.

If any interior point of $N_1 G N_2$ is part of the boundary γ, all of $N_1 G N_2$ must be part of it; otherwise the curve h would cut across $N_1 G N_2$ and so certainly intersect $N_1 A N_2$, contrary to hypothesis. In this case $N_1 G N_2$ is a part of the boundary γ and the region σ lies entirely in s. If there is no such interior point, the geodesic arc $N_1 G N_2$ lies inside of s except at N_1 and N_2. Hence if we surround the boundary γ by a cyclical chain of nearby points

$$N_1, N_2, \cdots, N_n,$$

the short distinct geodesic arcs

$$N_1 N_2, N_2 N_3, \cdots, N_n N_1$$

lie inside of s or coincide with γ. We assume that this chain encircles γ in a positive angular sense: in this event, the part toward γ will lie everywhere to the left.

Now in the geodesic polygon so formed, the angle in s at any vertex will be less than or equal to π. For if the angle at N_{i+1}, say, exceeds π and we consider the short geodesic arc $N_i N_{i+2}$, it is apparent that γ lies to the left of $N_i N_{i+2}$ also, so that N_{i+1} is completely encircled by points not on γ which is impossible.

But the integral curvature of the part of S bounded by this polygon will be precisely the sum of these n interior angles diminished by $(n-2)\pi$ by a well known formula, and so will be less than 2π, which is impossible since the integral curvature of either part of S bounded by g is exactly 2π by the same formula.

Thus a contradiction has been obtained, so that the original assertion must be true.

We now introduce parameters as follows. Let an arbitrary directed geodesic f cut the directed geodesic g in a point P.

The position of P may be measured by means of the arc length θ from a fixed point P_0 of g; if the total length of g be taken as 2π by an appropriate choice of a unit of length, the variable θ is a periodic variable of period 2π. Furthermore let φ denote the angle between the positive directions of g and f at the point P so that $0 < \varphi < \pi$.

The next crossing of g by h will be in the opposite sense, and the corresponding θ_1^* and φ_1^* will vary analytically with θ and φ. Thus a transformation T is defined

$$T\colon \quad \varphi_1^* = f(\varphi, \theta), \qquad \theta_1^* = g(\varphi, \theta),$$

which takes the ring

$$R\colon \quad 0 < \varphi < \pi, \qquad 0 \leqq \theta < 2\pi$$

(φ, θ, polar coördinates) into itself.

Similarly a crossing in the opposite sense at φ^*, θ^*, will be followed by a crossing φ_1, θ_1, thus defining a second one-to-one, analytic transformation

$$T^*\colon \quad \varphi_1 = f^*(\varphi^*, \theta^*), \qquad \theta_1 = g^*(\varphi^*, \theta^*)$$

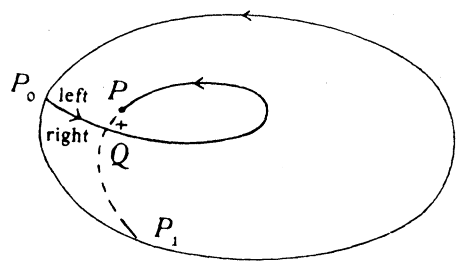
of R into itself.

Along and near the boundaries of R, T and T^* are to be regarded as continuous. This fact may be seen as follows: If a geodesic f near to g intersects g at a small angle φ with given θ, then of course f will intersect later at a small angle φ_1^* with coördinate θ_1^* nearly that of the conjugate point to the first point of intersection, as was observed above. Obviously this ensures the specified continuity. Furthermore, we know that three successive conjugate points correspond to an arc of more than one complete cycle of g.

If we have $\varphi = 0$ or π we have respectively $\varphi_1^* = 0$ or π, while if $\varphi^* = 0$ or π we have likewise $\varphi_1 = 0$ or π. Furthermore as θ or θ^* increases along a boundary so will θ_1^* or θ_1; in fact it has been noted that if θ is the coördinate of P, then θ_1^* is the coördinate of the conjugate point P'. Hence

T and T^* are direct transformations leaving the outer and inner boundaries of R invariant.

It is now desirable to consider somewhat more in detail the nature of T and T^* along the boundaries. These transformations are evidently not determined up to a complete rotation, and it is desirable to make a convention which eliminates this arbitrary factor, and enables us to compare the transformations along the inner and outer boundaries. Let us consider any directed geodesic arc $P_0 P_1$ of f between two successive crossings P_0 and P_1 of g by f, and let us define a double point Q of $P_0 P_1$ as positive if a moving point in describing the arc $P_0 P_1$ passes over the arc $P_0 Q$ already described at Q. from the left to the right side; and as negative in the contrary case. The 'index' of $P_0 P_1$ may be defined as the difference between the number of positive and the number of negative crossings within $P_0 P_1$. Let us define the value θ_1^* of θ at P_1 as that corresponding to the positively taken arc $P_0 P_1$ increased by $2\pi i$ where i is the index.



We can now show that, as thus defined, θ_1^* varies continuously with θ. In fact as θ and φ change, no new crossings can be introduced within $P_0 P_1$ inasmuch as a geodesic arc cannot be tangent to itself. If a positive crossing is introduced at P_1, evidently θ_1^* increases through an exact multiple of 2π as it should. If a negative crossing is introduced, θ_1^* decreases similarly. Thus the convention is appropriate to all cases.

Now it is also evident that if a geodesic arc $P_0 P_1$ near to g be continuously deformed near g while not maintaining its geodesic character, with P_0 and P_1 fixed, and only simple interior tangency allowed, then positive and negative crossings appear in or disappear in associated pairs. Thus if the arc $P_0 P_1$ be deformed into a spiral so as to eliminate all negative crossings, it is plain that the index will equal the number of apparent circuits of g made by the spiral. In this case the increase in θ, namely $\theta_1^* - \theta$, will be measured by the arc length along the curve g, according to our convention. However if the positive crossings are eliminated, the convention gives $2\pi i$ ($i \leqq 0$) increased by the angular value of the short positive arc $P_0 P_1$ as the difference $\theta_1^* - \theta$, and hence this difference exceeds the arc length $P_0 P_1$ by 2π.

Consequently it is seen that $\theta_1^* - \theta$ is measured in the sense of the convention by the actual increase in θ along g for $\varphi = 0$, while, for $\varphi = \pi$, $\theta_1^* - \theta$ is measured by the algebraic increase along the arc (actually a decrease) augmented by 2π since $P_0 P_1$ is to be taken positively.

Now in going from a point P to its second conjugate point P'' with $\varphi = 0$, θ increases by an amount α,

$$2k\pi < \alpha < 2(k+1)\pi,$$

where $k \geqq 1$ is independent of the position of P; in fact the one-to-one, direct transformation of the points of g from P to P'' defines a rotation number† which lies between $2k\pi$ and $2(k+1)\pi$ with $k \geqq 1$, because of a property of the conjugate points along a geodesic of minimax type already specified. Hence we infer

$$2k\pi \leqq g(0, \theta) - \theta < 2(k+1)\pi.$$

Similarly if we consider $\varphi = \pi$, then $\theta_1^* - 2\pi$ is diminished by a like amount so that we find

$$-2k\pi < g(\pi, \theta) - \theta \leqq -2(k-1)\pi.$$

It follows that the transformation T (or T^*) advances points in opposite angular directions along the two boundaries of R,

† For definition and short discussion of the Poincaré rotation numbers, see my paper, *Surface Transformations and their Dynamical Applications*, Acta Mathematica, vol. 43 (1922), section 45.

at least if the exceptional case of a second conjugate point coincident with the initial point after a single circuit be excluded.

11. Application of Poincaré's theorem to problem. The geodesic problem is of Hamiltonian form of course, with principal function H given by the squared velocity. In the four-dimensional manifold of states of motion, the quadruple integral

$$\iiiint dp_1\, dq_1\, dp_2\, dq_2$$

is an invariant integral. In a particular invariant sub-manifold $H =$ const. there must then be an invariant volume integral, namely

$$\iiint dq_1\, dp_2\, dq_2,$$

provided that $H = h, q_1, p_2, q_2$ are taken as coördinates (section 3). The restriction $H =$ const. merely fixes the constant velocity.

The ring R of states of motion crossing g positively is evidently represented in this manifold by a ring $\bar{R}$ bounded by the two closed curves representing g traversed in the two possible senses. The transformed point $TT^*(P)$ of a point P of this ring by TT^* is obviously obtained by following the corresponding stream line until it meets $\bar{R}$ a first time. Further consideration shows that the ring $\bar{R}$ so obtained is an analytic surface.

Now if we consider a tube of stream lines with two bases of areas $\Delta\sigma_1$, $\Delta\sigma_2$, in $\bar{R}$, with q_1 as independent variable, and if α_1, α_2 be the angles which the stream lines make with $\bar{R}$ at these two surface elements respectively, then the loss of 'volume' at one base in time Δq_1 is nearly

$$(\sin \alpha_1\, \Delta\sigma_1)\, \Delta q_1$$

while the gain at the other is nearly

$$(\sin \alpha_2\, \Delta\sigma_2)\, \Delta q_1 .$$

Thus $\int \sin \alpha \, d\sigma$ yields the positive invariant area integral necessary for the application of Poincaré's geometric theorem.

Hence there exist two points of R invariant under TT^*, and this gives immediately the following conclusion:

Let there be given a convex analytic surface on which the closed geodesic of minimax type known to exist is without double points, and suppose that the second conjugate point of no point on this geodesic arises on precisely a single complete circuit. Then there will exist a second closed geodesic which intersects the known minimax geodesic only twice.

Of course a single closed geodesic yields two invariant points.

Under the same conditions there must be two distinct closed geodesics which meet the geodesic of minimax type only twice.

To prove this we proceed as follows. We have for the points of R

$$T(\theta, \varphi) = (\theta_1^*, \varphi_1^*)$$

by definition. On the other hand the same geodesic may be taken in the opposite sense so that

$$T(\theta_1^*, \pi - \varphi_1^*) = (\theta, \pi - \varphi).$$

If then we define the 'reflection' U in such wise that

$$U(\theta, \varphi) = (\theta, \pi - \varphi),$$

we obtain

$$TU(\theta, \varphi) = (\theta_1^*, \pi - \varphi_1^*)$$

and thence

$$TUTU = I$$

where I is the identity. Hence $TU = V$ is a transformation of period 2, as is U, and we find $T = VU$, i. e. T is a product of two involutory transformations. Similarly we find $T^* = V^*U$ where V^* is also involutory. Hence we infer that TT^* has the form VUV^*U. Suppose now that there is an invariant point P under TT^* so that

$$VUV^*U(P) = P.$$

We obtain then, by inverting,

$$UV^*UV(P) = P,$$

whence

$$TT^*V(P) = V(P).$$

Hence if P is an invariant point under TT^*, so is $V(P)$.

But $V(P)$ must be a geometrically distinct point from P itself. Otherwise we should have

$$TU(P) = P$$

or, more explicitly, for the corresponding (θ, φ)

$$\theta_1^* = \theta, \qquad \pi - \varphi_1^* = \varphi.$$

But this would mean that the first intersection of the geodesic with the minimax geodesic g crosses it at the same point and in the opposite direction, which is manifestly impossible.

Now the indices of the invariant point P and of the associated invariant point $V(P)$ under the transformation TT^* are equal. In fact make the change of variables corresponding to the symbolic equation

$$Q = V(P),$$

by which any point P is taken into $V(P)$. The modified transformation less is of course

$$VTT^*V = (TT^*)^{-1}$$

as is at once verified by substitution of the factored form of TT^* derived earlier. Hence the transformation TT^* in the neighborhood of one invariant point is equivalent to the inverse transformation in the neighborhood of the associated invariant point. But the index of an invariant point is unaltered by a change of variables, and is the same as for the inverse transformation. Consequently the indices of the two associated invariant points are necessarily equal in all cases.

It is obvious geometrically that the association of pairs of points arises from the circumstance that every geodesic may be traversed in two opposite senses.

Since the extended theorem of Poincaré allows us to infer the existence of two invariant points with oppositely signed indices, we infer that there exist two geometrically distinct closed geodesics which intersect the closed geodesic only twice.

This completes the proof of the previous italicized statement.

By applying the same theorem to higher powers of TT^* the existence of other types of closed geodesics could be inferred. Moreover the methods of section 1 are applicable and show that there will be in general infinitely many closed geodesics in the immediate vicinity of a closed geodesic of stable type.

There are two further remarks concerning the geodesic problem which I will make in conclusion. In the first place the conclusion that there will exist at least two other distinct closed geodesics meeting the geodesic of minimax type only twice is no doubt valid for surfaces of much more general type.

Secondly, if the convex surface is symmetric in space about a plane containing a geodesic g, there will be various closed geodesics which intersect g twice at right angles. Methods for dealing with these symmetric closed geodesics can be used analogous to those which I employed in dealing with certain symmetric periodic orbits in the restricted problem of three bodies (loc. cit.). In fact if g is of minimax type, the transformations T and T^* become identical, and TT^* appears as the square of a product of two involutory transformations, as is the case for the fundamental transformation in the restricted problem of three bodies.