CHAPTER V

EXISTENCE OF PERIODIC MOTIONS

1. Role of the periodic motions. The periodic motions, inclusive of equilibrium, form a very important class of motions of dynamical systems. It will be our principal aim in this chapter to consider various general methods by which the existence of periodic motions may be established. In the next chapter a deeper insight into the distribution of the periodic motions is obtained for dynamical systems with two degrees of freedom. Such systems are of the simplest non-integrable type.

The case of a single equation of the first order presents no interest. If the equation is

$$dx/dt = X(x),$$

there will clearly be equilibrium for roots of $X = 0$, while for all other motions there will either be asymptotic approach to one of the equilibrium positions or indefinite increase of $|x|$. Here the equilibrium positions play the central role.

For the case next in order of simplicity there are a pair of equations of the first order. Here the geometric methods of Poincaré* yield the qualitative characteristics of all possible motions, and it is found that the equilibrium positions and periodic motions play the central role. The next following section will be devoted to an example of this type.

When we restrict attention, however, to a pair of equations of Hamiltonian or Pfaffian type (corresponding to a single degree of freedom), there is the special circumstance of an energy integral. Here we assume that the time t does not

* See his paper *Sur les courbes définies par une équation différentielle*, Journal de Mathématiques, ser. 3, vol. 7 (1881), vol. 8 (1882), ser. 4, vol. 1 (1885), vol. 2 (1886).

explicitly appear in the equations; the case when t is involved as a periodic variable is properly to be regarded as of the same degree of generality as that of two degrees of freedom. If we represent the two variables p, q by points in the plane,* the curves of motion will pass one through each point and can only form closed curves or branches open toward infinity in both senses. The corresponding families of periodic motions and unstable motions constitute the totality of motions, except that certain of these curves may contain one or more equilibrium points, in which case there is asymptotic approach to equilibrium in one or both senses.

In the case of dynamical systems of more complicated type it is not clear that the periodic motions take an equally important part. For dynamical systems with only two degrees of freedom, such as are considered in the next chapter, it is, however, almost certain that they continue to play a dominant role. In more complicated cases of still more degrees of freedom, the recurrent motions, introduced in chapter VII, are perhaps to be regarded as the appropriate generalization of the periodic motions, and so likely to become of considerable theoretic importance.

2. An example. The example alluded to in the preceding section is concerned with the rectilinear motion of a particle in a field of force of general type.

More exactly stated, we consider the motion of a particle P of unit mass which moves in a line subject to a force $f(x, v)$ dependent on its space coördinate x and its velocity v.† For the sake of definiteness we shall assume that there is one and only one possible position of equilibrium in the line of motion, and that the motion under consideration is stable in the sense that, for $t > 0$, x and v remain limited in absolute magnitude.

* It is conceivable that p, q are coördinates on a more complicated surface, but we will refer only to the simplest case here.

† For an outline of the treatment here given see my paper, *Stabilità e Periodicità nella Dinamica*, Periodico di Matematiche, ser. 4, vol. 6 (1926).

The usual form for the equation of motion is the single equation

$$d^2x/dt^2 = f(x, dx/dt),$$

where f is a known function which we shall assume to be analytic in the two variables involved. On taking the point of equilibrium at $x = 0$ we have further

$$f(0, 0) = 0; \qquad f(x, 0) \neq 0 \text{ if } x \neq 0.$$

Write $dx/dt = y$ and replace the above equation of the second order by the equivalent pair of equations of the first order

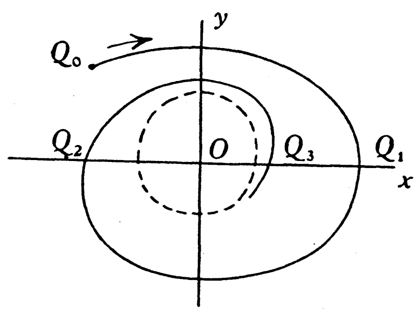
$$dx/dt = y, \qquad dy/dt = f(x, y).$$

These are clearly a pair of equations of the type to which the existence and uniqueness theorems apply.

If we take x, y as the rectangular coördinates of a point Q in the plane, the possible motions of the particle correspond to analytic integral curves, filling the x, y plane, for which

$$dy/dx = f(x, y)/y.$$

The only point curve is the origin O which corresponds to equilibrium. The other curves have everywhere a continuously turning tangent, since not both dx/dt and dy/dt vanish except at the origin. Moreover it is only along the x axis that the slope is infinite.



Consider now the particular integral curve corresponding to the given stable motion. Let Q_0 (see figure) be the point of this curve which corresponds to $t = 0$. For definiteness we assume Q_0 to lie in the upper half-plane; the modifications in our statements necessary to make them apply if the point Q_0 lies in the lower half-plane are obvious. Since

$dx/dt = y$ is positive as long as the point Q on the integral curve fails to cross the x axis, Q moves continually towards the right as t increases from 0.

But by hypothesis the point Q lies in a square of sides $2M$ with center at O and sides parallel to the axes.

Hence, while t increases but remains less than any time corresponding to a crossing of the x axis, x, which is also increasing and bounded, must approach a limit $\bar{x}$.

If t approaches a finite limit $\bar{t}$ as x approaches $\bar{x}$, then by the existence theorem the motion can be continued further, but of course not with $y > 0$ according to the definition of $\bar{t}$. Hence in this case $\bar{y}$ must be 0, and the curve crosses the x axis at $(\bar{x}, 0)$. It is to be noted that $\bar{x}$ cannot be zero. Else we should have two solutions, namely the given solution $x(t)$, $y(t)$ and also $x = 0$, $y = 0$, both of which satisfy the initial conditions $x = 0$, $y = 0$ at $t = \bar{t}$, and this would contradict the uniqueness theorem.

If t approaches an infinite limit as x approaches $\bar{x}$, we may show that (x, y) approaches $(0, 0)$. If possible suppose the contrary to be true. It is apparent that, inasmuch as x continually increases but remains less than M in absolute value while y also remains less than M in absolute value, the point (x, y) either approaches a point $(\bar{x}, \bar{y})$, or a strip $(\bar{x}, y)$ where $y_0 \leqq y \leqq y_1$.

But this second possibility would clearly require indefinitely large curvature $\varkappa$ of the integral curve near $(\bar{x}, \bar{y}_0)$ and $(\bar{x}, \bar{y}_1)$, given by

$$\varkappa = \frac{y(f_x y + f_y f) - f^2}{y^2 + f^2}$$

(where f_x, f_y denote partial derivatives), and an indefinitely small value of $y^2 + f^2$. This necessitates of course that $\bar{x}$, $\bar{y}_0$, $\bar{y}_1$ vanish, contrary to hypothesis.

Hence y must approach a value $\bar{y}$ as t approaches infinity.

Now clearly the arc length

$$\int_0^t \sqrt{f^2 + y^2}\, dt$$

will become infinite with t unless (x, y) approaches $(0, 0)$, and indefinitely large arc length would also imply indefinitely large curvature.

It follows then that either Q approaches the origin O from the left as t becomes infinite, or that Q crosses the axis at some point $Q_1 = (x_1, 0)$.

In this second case it is obvious that the point Q, on crossing the x axis at Q_1, starts to move to the left. By the argument employed above, Q will either continue to move to the left and approach O as t becomes infinite, or it must cross the x axis again at $Q_2 = (x_2, 0)$ for the second time.

The point Q_2 must, however, lie on the opposite side of O from Q_1 in this case; otherwise $f(\overline{x}_1, 0)$ and $f(\overline{x}_2, 0)$ have the same sign, and the point Q would move downwards at Q_2 as well as at Q_1. It is seen then that Q_1 falls to the right of O, and Q_2 to the left of O.

In fact it follows always that Q_1 is to the right of O, for if Q lies to the left of O, the particle cannot approach O after passing Q_1 since it then moves to the left. Thus a point Q_2 would fall on the same side of O as Q_1.

By repeating this argument indefinitely, we either arrive at a finite number of crossings $Q_1, Q_2, \cdots, Q_n$, alternately to right and left of O, after the last of which Q approaches O, or we find an indefinite set of points $Q_1, Q_2, \cdots$.

From the *analysis situs* of the figure it is apparent that the integral curve must either expand away from O towards a limiting oval about O; or must form such an oval; or must contract towards such an oval (like that indicated in the figure); or contract towards the point O itself. Of course the curve cannot touch or cross itself, by the elementary existence and uniqueness theorems.

Hence we have the following types as the only possible ones for a stable rectilinear motion of a particle in a force field with single equilibrium position:

(a) The particle oscillates indefinitely often back and forth past the equilibrium position with expanding amplitude, and approaches a periodic motion asymptotically.

(b) The particle oscillates periodically about the equilibrium position.

(c) The particle oscillates indefinitely often back and forth with decreasing amplitude and approaches a periodic motion.

(d) The particle oscillates back and forth a finite or an infinite number of times and approaches the equilibrium position.

(e) The particle is in the equilibrium position.

A very clear idea of all the possible motions of the particle under an arbitrary law of force of this type is readily obtained, on the basis of the above discussion.

Let us consider the ordered set of distinct closed curves in the x, y plane corresponding to the periodic motions. All of these must evidently enclose the origin which can properly be regarded as the innermost curve of the set.

Any other curve of motion may lie between a pair of the periodic curves in which case it will continue between them always and so be stable. The particle then approaches one of these enclosing periodic motions asymptotically as t increases indefinitely and the other as t decreases indefinitely.

In the only alternative case the curve will lie outside of the outermost curve of the set of periodic motions, and will clearly be stable in one and only one sense of time t and will approach the corresponding periodic motion asymptotically in this sense.

3. **The minimum method.** Suppose now that we have a Lagrangian dynamical problem,

$$\delta I = \delta \int_{t_0}^{t_1} L\,dt = 0,$$

where L is a function of the space coördinates $q_1, \cdots, q_m$ and of the velocities $q_1', \cdots, q_m'$, being quadratic in these last variables. There is an integral of energy, namely

$$L_2 - L_0 = \text{const.}$$

By adding an appropriate constant to L_0, we may assume that the arbitrary constant vanishes for the given motion.

We propose to simplify the given problem by making use of the integral of energy which may then be written

$$L_2 - L_0 = 0.$$

As has already been seen, it is possible to give an alternative variational formulation of the problem

$$\delta \int_{t_0}^{t_1} 2 \sqrt{L_0 L_2} + L_1)\, dt = 0$$

(see chapter II, section 3), in which the integrand is positively homogeneous of dimensions unity, so that the value of the integral is independent of the parameter t employed along the path of integration.

Now the underlying variables here are $q_1, \cdots, q_m$. But it is not important which particular set of variables is employed, since an arbitrary one-to-one, analytic transformation does not affect the validity of the variational principle. It is important, however, to require that the sets of coördinate values $q_1, \cdots, q_m$ form a certain analytic manifold M of known connectivity. Accordingly we assume that the coefficients in L are analytic in $q_1, \cdots, q_m$ over this manifold when $q_1, \cdots, q_m$ are properly chosen. Let us furthermore assume that $4 L_0 L_2 - L_1^2$, which is a homogeneous quadratic form in the velocities, is a positive definite form. The expression $ds^2 = L_0 L_2 \, dt^2$ may be considered as the element of arc on the 'characteristic surface' M.

Let l denote any closed curve in M which cannot be continuously deformed to a point. Here M is taken to be multiply connected in the sense of linear connectivity.

Suppose further that under such deformation the integral I taken around the curve l increases indefinitely if l does not remain wholly on the finite part of the manifold, and also that I exceeds a certain positive constant I_0 for any choice of l. There will then be a positive lower bound for I along these curves.

It is intuitively apparent that this bound will be attained, and will yield a closed curve corresponding to a periodic motion. We shall not attempt here to enter any of the logical details but merely to state the result.*

Given a Lagrangian dynamical problem of this type with

$$L = L_0 + L_1 + L_2,$$

and such a circuit l not deformable to a point on the characteristic surface, then, for the prescribed value c of the energy constant, i. e., of

$$L_2 - L_0 = c,$$

there exists a periodic motion of the same type as l for which

$$I = \int_l (2\sqrt{(L_0 + c)\, L_2} + L_1)\, dt$$

is an absolute minimum.

If $L_1 = 0$, so that the problem is reversible, the integral I becomes the arc length s on the characteristic surface, and the periodic motion corresponds to a closed geodesic of the given type.

In the case of two degrees of freedom ($m = 2$), this minimum method yields only those periodic motions of unstable type for which the two non-vanishing multipliers are real.† Doubtless analogous results hold for any number of degrees of freedom, and can be obtained by means of classical methods in the calculus of variations.

If the energy constant c varies, it is almost obvious that these periodic motions of minimum type vary analytically. Here is an instance then of the analytic continuation of a periodic motion with variation of a parameter (see section 9).

4. Application to symmetric case. There is one case in which the direct application of the minimum method fails, namely that in which there is no circuit l on the characteristic

* Cf. my paper, *Dynamical Systems with Two Degrees of Freedom*, Trans. Amer. Math. Soc., vol. 18 (1917), where the minimum method is developed more fully, and the important antecedent papers by Hadamard, Whittaker, Hilbert and Signorini are referred to.

† Cf. my paper (loc. cit), section 14.

surface not reducible to a point. It is interesting to observe that even here a slight modification of the minimum method may be applicable. This will be possible in those cases when the dynamical problem is 'symmetric', in the sense that the points of the characteristic surface may be grouped in distinct symmetric pairs, such that I has the same value along any curve and along its symmetric image. When this property is satisfied and the coördinates $q_1, \cdots, q_m$ of each point of a pair are taken to be locally the same, then L_0, L_1, L_2 are also the same at symmetric points.

To illustrate the idea involved, let the surface M be thought of as lying in ordinary space, and symmetric in the origin (but not passing through it), so that if x, y, z are the coördinates of a point of M then $-x, -y, -z$ are the coördinates of the symmetric point of M. Of course M is taken to be connected and as having the properties previously specified; in particular, M may be a convex surface symmetrical in the origin. The integral I may be thought of as ordinary arc length along a curve in the surface M.

Now suppose any closed curve $ABCDA$ is drawn in M such that ABC is the image of CDA, with A and C as symmetrical points. Let this curve l be continuously deformed in any manner, but with the condition that it is always to be composed of two symmetrical arcs ABC, CDA.

The integral I along l will then possess an absolute minimum which will be attained along some curve of this type. In fact we need only regard symmetric points as identical and consider I along the closed curve ABC on the modified manifold M' so defined.

If a Lagrangian problem of this type admits of symmetry in the sense specified, and if l is any symmetric circuit on the characteristic surface M, there will exist a symmetric periodic motion of the same type as l, for which I over any l is an absolute minimum.

In particular let there be a closed m-dimensional analytic surface with the connectivity of a hypersphere, which lies in $(m+1)$-dimensional space and is symmetric in the origin.

The result above is at once applicable and indicates the existence of at least one closed geodesic without multiple points.

More generally, if the Lagrangian problem of this type admits of an analytic transformation T into itself, whose k-th iterate is the identical transformation, and if l is a circuit invariant under T and not deformable to a point on M, there will exist a periodic motion of the same type as l.

5. **Whittaker's criterion and analogous results.** Hitherto we have dealt with Lagrangian dynamical problems possessing characteristic surfaces M without any boundary save at infinity. It is, however, very simple to extend our results to the case when M has one or more analytic $(m-1)$-dimensional boundaries, provided that the unique short geodesic arc joining any ordered pair of nearly points within M also lies within M. A boundary will be termed 'convex' in case it has this property.

In fact the minimizing curve in M is then either a closed extremal, in which case it is obviously not tangent to any of these boundaries and so lies within M, or it is composed of a finite or infinite number of extremal arcs whose vertices must lie on these boundaries of course. But, by the very definition of the convex boundary, boundary vertices cannot occur on the minimizing curve. In fact a short arc AVB of such a minimizing curve with vertex at V can be replaced by the shorter single extremal arc AB, which must lie wholly within M. This is absurd.

The surface M, defined in the preceding section, may be assumed to possess any number of finite convex boundaries in addition to the possible boundaries at infinity, without affecting the existence theorems there obtained.

The original criterion of Whittaker referred to the reversible case of two degrees of freedom when M was a ring. Here the result obtained is that there is a periodci motion of minimum type which makes a single circuit of the ring.*

* Cf. my paper, loc. cit., sections 10–13.

6. The minimax method. By means of the 'minimax' method it is possible to establish the existence of still further periodic motions.

Perhaps the simplest illustration of the method is obtained through the consideration of the geodesics on a torus-shaped surface in ordinary three-dimensional space. Evidently the minimum method explained above yields at least one closed geodesic having the type of any closed curve on the torus not deformable to a point. Now let a closed curve l be moved in any way from this minimizing geodesic back to the same position, while at least one of the two angular coördinates increases by $2k\pi$. It will be necessary certainly to increase the length of l during this movement of l, and there will be a least upper bound of length l^*, necessary in order that the movement be possible. At some intermediate stage then the varying curve will actually be of this length and will be taut. This position of l corresponds to a closed geodesic. Clearly it is not possible to deform all nearby curves of length less than l^* into one another or l^* would not be the minimax specified. This property is characteristic of geodesics of minimax type.

The above treatment is intuitive. However a rigorous treatment can be given.†

More generally we are led to the following conclusion: *A Lagrangian problem restricted as in section 4 and with $k > 1$ periodic motions of minimum type associated with a closed curve l, will necessarily possess at least $k-1$ further periodic motions of minimax type associated with this same circuit.*

If we omit consideration of all exceptional cases and use an intuitive form of reasoning, this more general principle may be made plausible as follows. Let $I_1, \cdots, I_k$ be the values of the integral I along the k periodic motions of minimum type which are known to exist, and let I^* be so large that it is possible to deform a curve l from any one of

† Cf. my paper, loc. cit., sections 15–19, for a development of the minimax method in the case of two degrees of freedom.

the corresponding curves $l_1, \cdots, l_k$ into any other without I becoming as large as I^*. For definiteness let us suppose that $I_1, \cdots, I_k$ are in increasing order of magnitude.

Let u be a variable parameter and consider the closed curves l of the given type for which $I < u$. For $u < I_1$ (the absolute minimum) there are no curves but, as u increases through I_1, closed curves not differing much from the curve for I_1 are admitted, and the more u increases the more extensive the variation of l about the curve for I_1 may be with $I < u$. Similarly as u increases through I_2 a new isolated set of curves l in the neighborhood of the curve for I_2 come into existence, for which $I < u$. And as u increases finally through I_k a last k-th set of curves comes into existence about the curve for I_k.

Now as u increases some two of these k sets of curves may unite, i. e. it becomes possible to deform a curve l from the curve for one I_α into a curve for I_β. There will then be a least value of u for which this is possible, and a corresponding periodic motion of minimax type. Each time that such a union takes place the number of sets of curves l with $I < u$ is diminishes by 1.

But when $u = I^*$ there is only one common group, so that $k-1$ unions have taken place. This is what we desired to prove.

It is not difficult to show that, unless the periodic motion of minimax type is multiple, only two sets of curves can coalesce at one of these critical values of u.

When the characteristic surface admits of discrete transformations into itself, an exceptional case arises in which the periodic motions of minimum type are to be counted more than once. This is the case of the torus alluded to above.

It is also of interest to observe that when a curve l is regarded as described k times, $k > 1$, the motions of minimum type remain so, but the motions of minimax type associated with them will not be the same as in the case $k = 1$, but will be distinct from these.

The general situation here requires further study.

7. Application to exceptional case. The case of the m-dimensional Lagrangian problem when the characteristic surface can be set into one-to-one analytic correspondence with the hypersphere is of exceptional interest, but the minimax method outlined above is not directly applicable since there are no closed curves l, not deformable to a point, from which to start. Nevertheless the existence of a periodic motion of minimax type may be established.

In order to make the reasoning as concrete as possible we shall direct attention to the reversible geodesic problem, although it is clear that the reasoning applies equally well for a Lagrangian problem of the kind treated in the preceding sections, with characteristic surface homeomorphic with the hypersphere.

Our first step will be to define what is meant by a 'covering' of the surface. In the case of a two-dimensional surface let the surface of the sphere be set into one-to-one analytic correspondence with the given surface. The small circles on the sphere in planes perpendicular to some axis are evidently carried into a set of closed analytic curves covering the given surface, two of these being point-curves. Thus we may conceive of the spherical surface as being distorted analytically to form a covering of the given surface M by means of this set of closed curves. The points of the covering can then be specified by two angular coördinate functions θ, φ on the surface where θ and φ represent colatitude with respect to the given axis and longitude respectively. The given closed curves correspond to $\theta =$ const., while φ varies from 0 to 2π. The coördinate θ ranges from 0 to π only, with the two extreme values corresponding to the point curves.

Now conceive of this covering as continuously deformed. This means that each point of the covering is carried by continuous variation into nearby points, while the curves of the covering go over into new curves. It is obvious that such a covering will always actually cover each point at least once and cannot reduce to a point.*

* The suggestion for a proof may be found in a footnote, p. 246 of my article, loc. cit., and this proof extends readily to the m-dimensional case.

Similarly for the m-dimensional case, we introduce a system of small circles on the hypersphere

$$x_1^2 + \cdots + x_{m+1}^2 = 1$$

($x_1, \cdots, x_{m+1}$ rectangular coördinates) with equations

$$x_3 = x_3^0, \cdots, x_{m+1} = x_{m+1}^0, \quad x_1^2 + x_2^2 = 1 - x_3^{0^2} \cdots - x_{m+1}^{0^2}.$$

Here the null circles of the set are in one-to-one, continuous correspondence with an $(m-1)$-dimensional hypersphere.

The image of this system of circles leads to an analytic covering of the given characteristic surface M. The points of the covering can then be specified by suitable coördinates, and we may conceive of the covering as continuously varied. It is obvious that such a covering will always cover each point of M at least once.

Now there is a maximum length L^* for any image of a circle, and there can be selected a distance d such that two points at geodesic distance not greater than d from each other in M are connected by a unique minimizing geodesic of length $\delta < d$. Let n be the positive integer such that

$$\frac{L^*}{n} < d \leqq \frac{L^*}{n-1}.$$

On the image of any circle select the point P_1 as the point which corresponds to $\varphi = 0$ and divide the curve into n arcs

$$P_1 P_2, P_2 P_3, \cdots, P_n P_1,$$

of equal length $< d$. Let a point Q_i move from P_i to P_{i+1} $(P_{n+1} = P_1)$ on each such arc in such wise that each arc is constantly divided proportionately. Consider the short geodesic arc $P_i Q_i$ and the arc $Q_i P_{i+1}$ of $P_i P_{i+1}$ during this process. At the outset the combined arc is $P_i P_{i+1}$ while at the conclusion it has been varied continuously to the geodesic arc $P_i P_{i+1}$.

In this way we see that it is possible by continuous variation to replace the covering with the given images of the circles by a covering of closed curves made up of geodesic arcs $P_1 \cdots P_n P_1$. Furthermore it is clear that the maximum length of any curve of the new covering cannot exceed L^*, while the maximum length of any component geodesic arc is less than d.

This constitutes our first step in a sequence of continuous variations of the given covering. The second step is to divide each curve of geodesic arcs in n equal parts starting with the midpoint of $P_1 P_2$. Thus points $Q_1, \cdots, Q_n$ are obtained, and a second modified covering made up of geodesic arcs $Q_1 Q_2, \cdots, Q_n Q_1$, each of length less than d, while the closed curves are of length less than L^*. The modification can be affected by essentially the same device as before.

The process of successive n-section and variation thus commenced can be indefinitely continued. At each stage the individual arcs are of length less than d, and the total length of each curve is less than L^*. Furthermore the effect of each step is to diminish (or at least not to increase) the length of the curve.

It is conceivable that some of the intermediate curves reduce to points during the process, but this would in no way invalidate the reasoning here made. However it is not possible that the maximum length of every curve becomes less than d. This may be seen readily. In the contrary case the curves $P_1 \cdots P_n P_1$ of geodesic arcs of the corresponding covering can all be reduced to point-curves as follows: let each point Q move toward P_1 along the unique short geodesic joining it to P_1, in such wise that proportional parts of this distance are simultaneously described by every point P. In this way the m-dimensional covering becomes $(m-1)$-dimensional at most, and so cannot pass through all points of M, which is absurd.

In connection with this last step it should be noted that the process adopted leaves the set of points P_i an analytic $(m-1)$-dimensional manifold at each stage, so that no difficult point-set questions connected with dimensionality arise.

We infer that the maximum length L_p at the p-th stage diminishes as p increases, and approaches a positive lower limit $L > d$.

It is now easy to prove that any corresponding sequence of geodesic arcs

$$P_1^{(p)} \cdots P_n^{(p)} P_1^{(p)} \qquad (p = 1, 2, \cdots)$$

for which the length is this maximum L_p will contain a closed limiting geodesic of length L. In fact we shall be able to establish this fact at once on the basis of the following lemma:

LEMMA. If any closed curve of n equal arcs $P_1 P_2, \cdots, P_n P_1$ each of length $\leqq d$, and of total length $\geqq d$, is modified to the curve of geodesic arcs $P_1 \cdots P_n P_1$, and then to the series of geodesic arcs $Q_1 \cdots Q_n Q_1$ joining the midpoint Q_1 of $P_1 P_2$ with the points $Q_2, \cdots, Q_n$ of n-section, and if the exterior angle between the geodesics $P_i P_{i+1}$ at some vertex exceeds $\delta > 0$, then the difference between the length of the initial and final curve exceeds a specifiable positive constant dependent only on δ.

In the first place we note that the two steps each decrease total length. Hence the proposition will certainly be true unless the first step changes the total length only very slightly. Since n is fixed one for all, this means that each geodesic arc $P_i P_{i+1}$ is substantially the same in length as the original arc $P_i P_{i+1}$. By Osgood's theorem in the calculus of variations, the original arc must be very near to the modified geodesic arc, and these latter arcs are nearly equal also. Hence the points of Q_i of n-section of the geodesic arcs fall very near to the points half way between the points P_i of n-section on the original curve. In consequence it appears that if the exterior angle at any P_{i+1} exceeds a specified δ the sum of the geodesic arcs $Q_i P_{i+1}$ and $P_{i+1} Q_{i+1}$ will exceed the geodesic arc $Q_i Q_{i+1}$ by a specifiable positive quantity. This would lead to the desired conclusion.

Thus an outline of the proof has been given. Obviously the proof is of such a character that a full statement of all the questions of uniformity involved would be lengthy, although the outline sufficiently indicates the general direction of procedure.

On the basis of the lemma our discussion is at once completed. In fact the lengths of the sucessive curves specified would decrease indefinitely often by a specifiable amount unless the exterior angles at the vertices of the geodesic arcs approach 0 uniformly. Hence these exterior angles approach 0. But the points P_1^p have at least one limit point P, and the directions of the geodesics have a limiting direction, so that there is a limiting geodesic which is clearly closed and of length L precisely.

If the m-dimensional characteristic surface M is homeomorphic with the m-dimensional hypersphere, there exists at least one periodic motion obtained by the process specified above.

It is natural to expect that such a motion is of minimax type, but we shall not attempt to establish this conjecture here. In the simplest case of two degrees of freedom this conjecture holds true.

8. The extensions by Morse. The methods of minimum and minimax suffice only to give certain of the periodic motions. Remarkable recent work by Morse* indicates with a high degree of probability that all of the types of periodic motions can be discovered by suitable extension of these methods, based on a deeper use of the principles of *analysis situs.* Moreover the number of periodic motions of the various types (the minimum and minimax types being the simplest of a series) are characterized by certain interrelations discovered by Morse, although so far only explicitly developed by him as to apply to dynamical systems with two degrees of freedom in the neighborhood of a periodic motion.

9. The method of analytic continuation. The method of analytic continuation of Hill and Poincaré starts with a

* See his paper *Relations Between the Critical Points of a Function of n Independent Variables*, Trans. Amer. Math. Soc., vol. 27 (1925), and a forthcoming paper, *A Theory of the Ordinary Problem of the Calculus of Variations in the Large.*

known periodic motion and then obtains an analytic continuation of it with variation of a parameter c.

For the sake of definiteness we consider a system of Hamiltonian equations

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

in which H is analytic in $p_1, \cdots, q_m, t, c$, and periodic in t of period 2π. Furthermore we suppose that the origin is a point of generalized equilibrium for $c = 0$.

By proper preliminary linear change of variables of the type made in chapter III, section 6, we may take H in the normal form

$$H = -\sum_{j=1}^{m} \lambda_j p_j q_j + H_3 + \cdots$$

for $c = 0$, in case $\lambda_1, \cdots, \lambda_m$ are distinct.

Now the general solution of the given system may be written

$$p_i = p_i(p_1^0, \cdots, q_m^0, t, c), \qquad q_i = q_i(p_1^0, \cdots, q_m^0, t, c)$$

for $i = 1, \cdots, m$, where $p_1^0, \cdots, q_m^0$ denote the values of $p_1, \cdots, q_m$ respectively at $t = 0$.

The condition for periodicity is then contained in the system of $2m$ simultaneous equations

$$p_i(p_1^0, \cdots, q_m^0, 2\pi, c) = p_i^0, \qquad q_i(p_1^0, \cdots, q_m^0, 2\pi, c) = q_i^0 \qquad (i = 1, \cdots, m).$$

Here all the variables are involved analytically as was established in chapter I. But this system of equations admits of the solution

$$p_1^0 = \cdots = q_m^0 = 0$$

for $c = 0$, by hypothesis. Hence there will be a unique solution $p_1^0, \cdots, q_m^0$, analytic in c, provided that the functional determinant

$$\begin{vmatrix} \frac{\partial p_1}{\partial p_1^0}-1, & \frac{\partial p_1}{\partial p_2^0}, & \cdots, & \frac{\partial p_1}{\partial q_m^0} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial q_m}{\partial p_1^0}, & \frac{\partial q_m}{\partial p_2^0}, & \cdots, & \frac{\partial q_m}{\partial q_m^0}-1 \end{vmatrix}$$

does not vanish for $t = 2\pi$, $c = 0$. But the $2m$ functions

$$\frac{\partial p_1}{\partial p_i^0}, \cdots, \frac{\partial q_m}{\partial p_i^0}$$

$(i = 1, \cdots, m)$ form m solutions of the equations of variation, as do also

$$\frac{\partial p_1}{\partial q_i^0}, \cdots, \frac{\partial q_m}{\partial q_i^0}$$

$(i = 1, \cdots, m)$. In addition these reduce to 0 at $t = 0$, except for $\partial p_i/\partial p_i^0$ and $\partial q_i/\partial q_i^0$ which are 1 for $i = 1, \cdots, m$.

The known nature of the terms in H for $c = 0$ yields as equations of variation

$$dy_i/dt = \lambda_i y_i, \quad dz_i/dt = -\lambda_i z_i \quad (i = 1, \cdots, m),$$

where y_i, z_i correspond to p_i, q_i. Hence the solutions above are of the following explicit forms

$$\begin{matrix} e^{\lambda_1 t}, & 0, & \cdots, & 0, \\ 0, & e^{\lambda_2 t}, & \cdots, & 0, \\ \cdot & \cdot & \cdot & \cdot \\ 0, & 0, & \cdots, & e^{-\lambda_m t}, \end{matrix}$$

where the only non-zero elements are the diagonal terms, namely

$$e^{\lambda_1 t}, \cdots, e^{\lambda_m t}, e^{-\lambda_1 t}, \cdots, e^{-\lambda_m t}.$$

Consequently the determinant written above reduces to

$$\prod_{j=1}^{m} (e^{2\pi\lambda_j} - 1)(e^{-2\pi\lambda_j} - 1),$$

and cannot vanish unless some λ_i is an integral multiple of $\sqrt{-1}$.

Under these circumstances then we have an analytic family of solutions

$$p_1(t, c), \cdots, q_m(t, c),$$

periodic of period 2π in t, as we desired to prove.

These restrictions may be notably lightened. In the first place, when not all the multipliers $\lambda_1, \cdots, \lambda_m$ are distinct, a similar normal form of solution exists. For example, take $\lambda_1 = \lambda_2$, while taking $\lambda_2, \cdots, \lambda_m$ distinct from each other and λ_1. In general the first and second solutions now have the forms

$$0, e^{\lambda_1 t}, 0, \cdots, 0,$$
$$e^{\lambda_1 t}, te^{\lambda_1 t}, 0, \cdots, 0,$$

and the $(m+1)$-th and $(m+2)$-th solutions likewise

$$0, e^{-\lambda_1 t}, 0, \cdots, 0,$$
$$e^{-\lambda_1 t}, te^{-\lambda_1 t}, 0, \cdots, 0$$

respectively. If the first and second rows are interchanged as well as the $(m+1)$-th and $(m+2)$-th rows, it appears that the final determinant will have vanishing elements on the lower side of the diagonal and will not itself vanish unless some λ_i is an integral multiple of $\sqrt{-1}$.

This is an entirely general result, namely that analytic continuation is possible so long as there is not a multiplier λ_i which is an integral multiple of $\sqrt{-1}$.

But such a multiplier indicates neither more nor less than the presence of a solution of the equations of variation with the period 2π of the given motion.

Let us define the generalized equilibrium point as 'simple' when there is no solution of the equations of variation with the period of the generalized equilibrium point, and otherwise as 'multiple'.

Analytic continuation of the generalized equilibrium is possible so long the equilibrium is simple.

By a change of variables

$$p_i = p_i(t, c) + P_i, \qquad q_i = q_i(t, c) + Q_i \quad (i = 1, \cdots, m),$$

we obtain modified Hamiltonian equations

$$\frac{dP_i}{dt} = -\frac{\partial H^*}{\partial q_i}, \qquad \frac{dQ_i}{dt} = \frac{\partial H^*}{\partial p_i} \quad (i = 1, \cdots, m)$$

where

$$H^* = H + \sum_{j=1}^{m} (p_j' Q_j - q_j' P_j).$$

These are of the same type as before, but have a generalized equilibrium point at the origin *for all small values of* c. We have written p_i', q_i' for the time derivatives of $p_i(t, c)$, $q_i(t, c)$ respectively.

The formal series solutions of a system of this type will of course also involve the parameter c. It is formal series of this type which are often useful in the applications; and the vanishing of the parameters such as c may correspond to a special integrable case of the dynamical problem when the periodic motion from which we start admits of explicit determination.

10. The transformation method of Poincaré. Sometimes a dynamical problem can be given a striking change of form. In fact the solutions of n differential equations of the first order, with right-hand members independent of the time t, can be represented by the steady motion of an n-dimensional fluid, of which a moving point has the dependent variables as coördinates. Now suppose that a closed, $(n-1)$-dimensional, analytic surface S can be constructed in this 'manifold of states of motion', such that every stream line cuts S at least once within any fixed interval of time τ, sufficiently large, and always in the same sense. Then S may be called a 'surface of section'. If a point P of S be followed along the stream line through P in the sense of increasing time, it intersects S again at a first subsequent point P_1. Thus there is defined a one-to-one, direct, analytic transformation of the surface of section into itself, namely the transformation T which takes each point P into the point P_1.

In this manner it may be possible to associate the given dynamical problem with a discrete transformation T of a closed $(n-1)$-dimensional surface into itself. Properties of the motions are then mirrored in properties of T. For example, the periodicity of a motion represented by a closed curve in the manifold of states of motion meeting S in $P, P_1, \cdots, P_{k-1}$ is reflected in the symbolic equations

$$P_1 = T(P),\ P_2 = T(P_1), \cdots, P = T(P_{k-1}),$$

so that $P, P_1, \cdots, P_k$ are all invariant points of S under the k-th iterate of T. Conversely, if P is so invariant, there is a corresponding periodic motion, represented by a closed curve meeting S in the k points $P, T(P), \cdots, T^{k-1}(P)$.

A surface of section in this sense will only exist if there is an angular variable φ in the manifold of states of motion which may be so defined as to constantly increase along each stream line. The necessity of this condition may be seen as follows. If a surface of section S exists, let φ be defined as 0 on this surface and as $2\pi t/\tau$ at any other point P, where τ is the complete time interval necessary to pass from S to S along the stream line on which P lies. Evidently φ is an analytic function of position which increases along every stream line by exactly 2π between successive intersections with S. Hence an angular variable φ exists. Conversely if such a variable φ exists, then $\varphi = 0$ will yield a surface of section.

A necessary and sufficient condition for a closed surface of section is the existence of an angular variable φ in the manifold of states of motion which constantly increases along every stream line.

More explicitly stated, there must exist a differential inequality

$$\frac{d\varphi}{dt} = \sum_{j=1}^{n} \Phi_j X_j > 0$$

in which Φ_i satisfy integrability conditions,

$$\frac{\partial \Phi_i}{\partial x_j} = \frac{\partial \Phi_j}{\partial x_i} \qquad (i, j = 1, \cdots, n),$$

and $X_1, \cdots, X_n$ denote the right-hand members of the differential equations as usual.

An extremely interesting type of surface of section possessing boundaries can be found in certain dynamical problems. Here the boundaries of S are closed analytic $(n-2)$-dimensional manifolds of stream lines, and every stream line not on these boundaries of S cuts the interior of S at least once within any interval of time τ, sufficiently large, and always in the same sense.

In the case $n = 3$, the surface of section is two-dimensional, and its boundaries may then be the closed curves corresponding to a single periodic motion. Now in the case of a Hamiltonian or Pfaffian dynamical problem with two degrees of freedom, the use of the energy integral reduces the order to $n = 3$. For such problems there seem in general to exist surfaces of section, as will appear in the next chapter.* The example of the next section illustrates the possibility of such a surface of section when there are more than two degrees of freedom, In such a case, however, it is necessary to make use of an $(n-2)$-dimensional, analytic, closed surface made up of stream lines, such as do not appear in general to exist.

Another case of very decided interest is that of an open analytic surface of section such as Koopman† has obtained in the exterior case of the restricted problem of three bodies.

It is evident in all of these cases that by the reduction to a transformation problem, the determination of the periodic motion is made to hinge upon that of the invariant points of a surface of section S under a transformation T and its iterates.

* Cf. my paper, loc. cit., sections 22–29.

† *On Rejection to Infinity and Exterior Motion in the Restricted Problem of Three Bodies,* Trans. Amer. Math. Soc., vol. 29 (1927).

11. An example. An example showing that such surfaces of section may exist for Hamiltonian problems of more than two degrees of freedom is afforded by the following dynamical problem:

A particle P in a conservative field of force in space moves in such wise that the force has always a positive component towards a fixed plane for points outside of that plane.

Here the equations of motion form a system of the sixth order and may be written

$$dx/dt = x', \qquad dy/dt = y', \qquad dz/dt = z',$$
$$dx'/dt = -\partial U/\partial x, \quad dy'/dt = -\partial U/\partial y, \quad dz'/dt = -\partial U/\partial z,$$

where x, y, z are the rectangular coördinates of P in space, and where $z = 0$ may be taken as the fixed plane. Also $\partial U/\partial z$ has the same sign as z, so that

$$\partial U/\partial z = \lambda z$$

where λ is a positive analytic function of x, y, z. The integral of energy may be written

$$\frac{1}{2}(x'^2 + y'^2 + z'^2) + U = 0,$$

provided that we absorb a suitable constant into U. Thus we restrict attention to the totality of motions satisfying this last relation, thereby reducing the system from the sixth to the fifth order. We shall consider only the case in which the surface $U = 0$ in space constitutes a closed simply-connected surface intersecting $z = 0$ in an oval, with $U < 0$ within the surface. The particle is then necessarily restricted to lie in the region $U \leqq 0$.

The five-dimensional manifold M of states of motion consists of the sets x, y, z, x', y', z', subject to the integral relation. The selected surface of section S will then be the four-dimensional part of M for which z vanishes with $z' = dz/dt \geqq 0$. The three-dimensional boundary $z = z' = 0$

of S is evidently made up of stream lines, since a point $z = z' = 0$ for any value t_0 of t remains always of this type.

Now the differential equation

$$\frac{d^2 z}{d t^2} + \lambda z = 0$$

shows at once that z vanishes at least once in an interval of time τ sufficiently large, but cannot vanish twice in an arbitrarily small interval of time. Hence a point of S followed along its stream line in the manifold of states of motion will cut S again within an interval 2τ of time, and always in the same sense, since for $z = 0$ we have $dz/dt > 0$ within S.

Evidently there is thus set up a one-to-one, analytic transformation of points within S into themselves. Furthermore for z, z' small, we have nearly

$$\frac{d^2 z}{d t^2} + \lambda(x, y, 0)\, z = 0,$$

where x, y are the coördinates of the motion in the plane near the given motion. But the particular solution of this equation for which $z = 0$ at $x = x_0$, $y = y_0$, $x = x_0'$, $y = y_0'$, with

$$\frac{1}{2}(x_0'^2 + y_0'^2) + U(x_0, y_0, 0) = 0$$

of course, vanishes at x_1, y_1, x_1', y_1', with

$$\frac{1}{2}(x_1'^2 + y_1'^2) + U(x_1, y_1, 0) = 0,$$

where x_1, y_1, x_1', y_1' are evidently analytic in x_0, y_0, x_0', y_0'. It is thus seen that the transformation T of S into itself can be regarded as one-to-one and continuous even along the boundary of S, provided we define T as taking the point

$$x_0,\ y_0,\ 0,\ x_0',\ y_0',\ 0$$

of S into the point

$$x_1, y_1, 0, x_1', y_1', 0.$$

We propose to show by means of this reduction to a transformation problem that there always exists a periodic motion intersecting the plane $z=0$ twice, unless there is a multiple periodic motion lying in the plane $z=0$.

In order to do this, let us consider the connectivity of the surface of section S. Evidently we may change the variables x, y to $\bar{x}, \bar{y}$ so that the oval $z=0$, $U \leqq 0$ becomes the circle

$$\bar{x}^2+\bar{y}^2 = 1.$$

If then we write

$$U = p(\bar{x}, \bar{y})\,(\bar{x}^2+\bar{y}^2-1)$$

with $p>0$ within this circle, and if we write further

$$x' = \sqrt{p}\,\bar{x}', \qquad y' = \sqrt{p}\,\bar{y}', \qquad z' = \sqrt{p}\,\bar{z}',$$

the equation for S takes the form

$$\bar{x}'^2+\bar{y}'^2+\bar{z}'^2+\bar{x}^2+\bar{y}^2 = 1 \qquad (\bar{z}' \geqq 0)$$

which may be written

$$\bar{z}' = (1-\bar{x}^2-\bar{y}^2-\bar{x}'^2-\bar{y}'^2)^{1/2}.$$

Hence the interior and boundary of S are in one-to-one, continuous correspondence with the interior and boundary of the four-dimensional hypersphere

$$\bar{x}^2+\bar{y}^2+\bar{x}'^2+\bar{y}'^2 = 1.$$

The transformation T defines a one-to-one, continuous, direct transformation of this hypersphere into itself.

But, by a well-known theorem due to Brouwer, such a transformation leaves some point invariant. In the problem at hand, we conclude that a periodic motion exists which intersects $z = 0$ twice, (case of an interior invariant point), or else a periodic motion $z = 0$ exists (case of an invariant boundary point). But this last case is that in which the equations $x_1 = x_0$, $y_1 = y_0$, $x_1' = x_0'$, $y_1' = y_0'$ obtain. This clearly means that the equations of variation possess a periodic solution along this plane periodic motion in which the z component is not 0. Hence the periodic motion is multiple, and, in a certain sense there is still a periodic motion in the infinitesimal vicinity of $z = 0$, intersecting $z = 0$ twice. It seems highly probable that an actual periodic motion intersecting $z = 0$ twice must exist in all cases.