UNIVERSAL
LIBRARY

SURVEYS IN APPLIED MATHEMATICS

II

DYNAMICS AND NONLINEAR MECHANICS

SURVEYS IN APPLIED MATHEMATICS

The books in this series were written as a joint project of the Office of Naval Research and *Applied Mechanics Reviews.*

I ELASTICITY AND PLASTICITY

by J. N. Goodier and P. G. Hodge, Jr.

II DYNAMICS AND NONLINEAR MECHANICS

by E. Leimanis and N. Minorsky

III MATHEMATICAL ASPECTS OF SUBSONIC AND TRANSONIC GAS DYNAMICS

by Lipman Bers

IV SOME ASPECTS OF ANALYSIS AND PROBABILITY

by Irving Kaplansky, Edwin Hewitt, Marshall Hall, Jr., and Robert Fortet

V NUMERICAL ANALYSIS AND PARTIAL DIFFERENTIAL EQUATIONS

by George E. Forsythe and Paul C. Rosenbloom

NEW YORK · JOHN WILEY & SONS, INC.

London · Chapman & Hall, Limited

DYNAMICS
AND NONLINEAR MECHANICS

II

SOME RECENT ADVANCES IN THE DYNAMICS OF RIGID BODIES AND CELESTIAL MECHANICS

E. LEIMANIS
The University of British Columbia

THE THEORY OF OSCILLATIONS

N. MINORSKY
Stanford University

Library of Congress Catalog Card Number: 58–10869
Printed in the United States of America

Preface

These survey articles, as any of their kind, have the primary objective of accounting in a summary fashion for the state of the fields which they cover, as determined by contributing developments at various times and in many places. Considering the sustained growth of mathematics itself, as well as the intensive use of an ever-increasing number of its branches in diversified applied contexts, the availability of such surveys in selected areas is believed to fill a real need. The present articles are therefore aimed not so much at research specialists, actively contributing to the subjects discussed, as they are aimed at a broader, mathematically literate audience, looking for contemporary information on the important problems and results in these disciplines, whether it be for use in classroom and seminar, or for the sake of possible application to problems in other fields of science and engineering, or simply for reasons of personal interest.

The selection of the areas surveyed, as well as their coverage, was further guided by giving first consideration to developments of whose current state no comprehensive picture could be obtained by going only to the readily accessible literature in familiar languages. It is a unique distinction of the mathematical community to have never been taken in by the myth, now generally shattered, that Russian science merely followed the lead of the West—at a respectful distance. Mathematicians remained aware of the vigorous development during the postwar years of research in their field also in the Communist countries, and they knew of the steady stream of important results which it produced. An early start was made to overcome language and other communication difficulties. The American Mathematical Society was the first of this country's scientific societies to institute, with support from the

Office of Naval Research, the systematic selection and translation of significant articles which appeared in inaccessible journals or unfamiliar languages. A further step was taken two years ago, when the Editorial Office of *Applied Mechanics Reviews*, again with Office of Naval Research support, initiated work on these surveys. It had become apparent by then that there existed major areas of modern mathematics and theoretical mechanics of whose current scope and fabric the Western literature alone—quite apart from missing individual recent results—conveyed only an incomplete and therefore inadequate picture. This is to be remedied in the areas covered by the present *Surveys in Applied Mathematics*.

Clearly, the success of this venture required that the surveys should be written by authorities in the respective fields, fully conversant with current research and abreast of the international literature. Special thanks must therefore be extended to the distinguished authors of these articles for having given their thought and time to the purpose at hand. The enlistment of their co-operation is due in large measure to the indefatigable leadership of the Editor of *Applied Mechanics Reviews* and his associates, as well as to the effective way in which the Midwest Research Institute and the Southwest Research Institute (after *Applied Mechanics Reviews* had moved there) jointly arranged for the conduct of the editorial work.

F. Joachim Weyl, Director
Mathematical Sciences Division
Office of Naval Research

Contents

SOME RECENT ADVANCES IN THE DYNAMICS OF RIGID BODIES AND CELESTIAL MECHANICS

Contents

THE THEORY OF OSCILLATIONS

SOME RECENT ADVANCES IN THE DYNAMICS OF RIGID BODIES AND CELESTIAL MECHANICS

BY E. LEIMANIS

The University of British Columbia

INTRODUCTION

After the problem of motion of a particle the simplest problem of dynamics is that of the motion of a rigid body. This last problem reduces to the following two problems: (i) the motion of the center of mass G of the body, and (ii) the motion of the body with respect to its center of mass G, considered as a fixed point. Consequently, every problem of the motion of a rigid body, for instance, the motion of a planet or an artillery projectile, contains the problem of motion of a rigid body about a fixed point as one of its component problems. We are also faced with the same problem in the theory of gyroscopic phenomena which underlies the construction of gyroscopic instruments, for example, the aircraft gyroscopic flight instruments. Therefore the problem of motion of a rigid body about a fixed point is not only of theoretical interest but also of great practical value. This problem and the problem of three bodies are the most celebrated of all unsolved classical dynamical problems.

Although the problem of motion of a heavy rigid body about a fixed point is completely or partially solved in certain special cases such as those of Euler, Lagrange, Kovalevskaya, Hess, and others, very little is known about the motion of a heavy rigid body when the mass distribution or the initial conditions of motion are arbitrary. The situation is similar in the three-body problem.

After the cases considered by Euler (1758) and Lagrange (1788), there was no progress for 100 years towards the solution of the problem of motion of a heavy rigid body about a fixed point. Obviously, a new approach to the problem was needed.

The success of Kovalevskaya (1888) lies in her new and more general reformulation of the problem in terms of the concepts of the theory of analytic functions. Considering the time t—the independent variable—

as a complex variable, Kovalevskaya proposed to find all cases of motion for which the parameters defining the motion can be expressed as meromorphic functions of t containing five arbitrary constants. As a consequence of her proposition, she arrived at the conclusion that, besides the cases of Euler and Lagrange, there is only one new case possible which at present is known as the case of Kovalevskaya.

Similarly, by application of the theory of analytic functions, Sundman (1912) obtained the general solution of the three-body problem in the sense proposed by Painlevé (1897). By extension of the classical representation of the elliptic motion he expressed the nine cartesian coordinates of the three bodies and the time, varying from $-\infty$ to $+\infty$, in terms of uniformly convergent power series of a single complex variable. The series obtained by Sundman for the coordinates remain valid if two of the three bodies collide. In this case they define an analytic continuation of the motion beyond the instant of collision. A more recent attempt by Vernić (1952) to extend the validity of the Sundman solution to triple collisions has not been successful. He tried to apply the theory of uniformization of multiple-valued functions to the three-body problem, but his principal theorems turned out to be erroneous.

The success in solving completely the problem of motion of a heavy rigid body about a fixed point in the previously mentioned three cases of Euler, Lagrange, and Kovalevskaya lies in the fact that it is possible to find besides the three classical integrals a new independent fourth algebraic integral. According to Jacobi's theory of the last multiplier, which is equal to unity in our problem, the existence of such an integral implies complete integrability of the problem. Hence, there arises the problem of finding all cases when there exists an additional fourth integral, algebraic or transcendental but single-valued in the parameters. Investigations by Poincaré, Husson, and others have led to the conclusion that such a fourth single-valued integral exists only in the cases of Euler, Lagrange, and Kovalevskaya, i.e., when the integrals of the equations of motion are meromorphic functions of t. However there is no general theorem to the effect that, if the general solution of a system of differential equations is meromorphic, this system assumes algebraic first integrals.

Similar efforts were made to obtain other algebraic integrals in the problem of three bodies, independent of the ten classical integrals, but without success. The theorems of Bruns (1887) and Poincaré (1890) on the nonexistence of further single-valued integrals are well known. Later, however, Cherry (1924) showed that the theorem of Poincaré no longer holds in its original formulation if the restriction that the integral is to be developable in powers of the perturbing mass μ is dropped.

In characterizing the motions of a dynamical system, those of periodic and asymptotic type are of the greatest importance. The investigations of Hill, Poincaré, Darwin, Moulton, Strömgren, and others on the periodic and asymptotic solutions in the restricted and in the general three-body problem have opened new vistas into the nature of this problem. However, so far it has proved impossible to arrive at the general solution of the problem by means of these special solutions.

Mettler [1] has dealt with periodic and asymptotic solutions of the equations of motion of an asymmetric heavy rigid body with a fixed point. He takes the differential equations not in the Eulerian form, because of their too special character, but in the more general Hamiltonian form which has to its advantage certain well-established methods of integration.

Although the literature on the motion of a heavy rigid body about a fixed point has grown very extensively during the last 65 years, it is concerned almost entirely with consideration of special cases. In conclusion we note that Klein's and Sommerfeld's hopes that "by finding enough special cases, we may some day be able to know more about the general solution of the problem" have not yet come true.

This report is an attempt to present from the mathematical point of view the development and the current state of the two classical dynamical problems and to give certain generalizations of them together with references to some of the more important papers in the two fields. Personal preferences of the author have been influential in the selection of the topics covered. Because of existing difficulties in securing the literature of the world, the survey does not intend to be complete.

It is in the nature of such a survey that proofs have to be omitted or can only be given in outline. Similarly, references to older results used are almost exclusively references to standard works on differential equations, dynamics of rigid bodies, and celestial mechanics.

THE DYNAMICS OF RIGID BODIES AND MATHEMATICAL EXTERIOR BALLISTICS

SECTION 1

CHAPTER 1

MOTION OF A RIGID BODY ABOUT A FIXED POINT

THE EULER AND POISSON EQUATIONS

1. The Euler angles

Let O be a fixed point of a rigid body about which the rotation of the body takes place, and let $OXYZ$ be a right-hand rectangular trihedral fixed in space. Let $Oxyz$ be a right-hand rectangular trihedral fixed in the body and moving with it. Furthermore, let the coordinate planes XOY and xOy intersect along the line of nodes ON, which is perpendicular to the plane through the axes OZ and Oz. Choose the orientation along ON in such a way that the trihedral $ONZz$ is right-handed. Denote the angles ZOz, XON, NOx by ϑ, ψ, φ, respectively. These are known as the three Euler angles, defining the position of the trihedral $Oxyz$ with respect to the trihedral $OXYZ$.

2. The Euler and Poisson equations

Let the principal axes of the momental ellipsoid of inertia of a rigid body at the fixed point O be taken as the axes of the moving trihedral $Oxyz$, and let A, B, C be the principal moments of inertia at O along these axes.

The Euler vector equation of motion of a rigid body about the fixed point O is then

$$\dot{\mathbf{I}} + \boldsymbol{\omega} \times \mathbf{I} = \mathbf{M} \tag{1}$$

where a dot denotes differentiation with respect to the time t, $\boldsymbol{\omega}(p, q, r)$ the instantaneous angular velocity vector, $\mathbf{I}(Ap, Bq, Cr)$ the angular momentum vector, and $\mathbf{M}$ the instantaneous moment vector of the external forces.

If the only force applied is gravity $-mg\mathbf{K}$, the point of application being the center of mass $G(x_0, y_0, z_0)$, then $\mathbf{M} = -w(\mathbf{r}_0 \times \mathbf{K})$, where $w = mg$ is the weight of the body, $\mathbf{K}(\alpha, \beta, \gamma)$ the unit vector along the OZ-axis and $\mathbf{r}_0(x_0, y_0, z_0)$ the position vector of the center of mass G. Without any loss of generality, the weight w can be assumed to be equal to unity. Equation (1) then assumes the form

$$\dot{\mathbf{I}} + \boldsymbol{\omega} \times \mathbf{I} + \mathbf{r}_0 \times \mathbf{K} = 0 \tag{2}$$

The Poisson vector equation is

$$\dot{\mathbf{K}} + \boldsymbol{\omega} \times \mathbf{K} = 0 \tag{3}$$

Scalar multiplication of equation (2) by $\boldsymbol{\omega}$ and $\mathbf{K}$ respectively and the use of equation (3) gives the two first integrals of motion

$$\mathbf{r}_0 \cdot \mathbf{K} = h - T \tag{4}$$

and

$$\mathbf{I} \cdot \mathbf{K} = c \tag{5}$$

where

$$T = \tfrac{1}{2}\boldsymbol{\omega} \cdot \mathbf{I}$$

is the kinetic energy of the body and h and c are constants of integration. The first integral is the so-called energy integral whereas the second integral expresses the law of conservation of the angular momentum about the vertical.

The third integral

$$K^2 = 1$$

which is obtained by scalar multiplication of equation (3) by $\mathbf{K}$, expresses the fact that the sum of the squares of the direction cosines of $\mathbf{K}$ is unity.

Sbrana [1] showed that the Euler and Poisson equations can be replaced by equations which contain only the components p, q, r of the angular velocity vector $\boldsymbol{\omega}$ but do not contain any direction cosines.

Bottema and Beth [2] have deduced in n-dimensional space a certain system of $n(n - 1)/2$ equations as the appropriate generalization of the Euler equations. For the special case of $n = 4$, the same authors [3] have studied the stationary motions of a rigid body, assuming that there are no forces other than the reactions at the fixed point O.

3. Particular integrable cases

The two vector equations (2) and (3) are equivalent to the following two sets of scalar equations

$$A\dot{p} + (C - B)qr = \beta z_0 - \gamma y_0 \tag{6}$$
$$B\dot{q} + (A - C)rp = \gamma x_0 - \alpha z_0$$
$$C\dot{r} + (B - A)pq = \alpha y_0 - \beta x_0$$

and

$$\dot{\alpha} = r\beta - q\gamma, \qquad \dot{\beta} = p\gamma - r\alpha, \qquad \dot{\gamma} = q\alpha - p\beta \tag{7}$$

where

$$\alpha = \sin\vartheta \sin\varphi, \qquad \beta = \sin\vartheta \cos\varphi \qquad \gamma = \cos\vartheta$$

It is well known that the problem of motion of a heavy rigid body about a fixed point can be reduced to quadratures if it is possible to find besides the three classical integrals of the Euler and Poisson equations an additional fourth independent integral. Stankevich [1–4], for instance, started with the Jacobi partial-differential equation and used contact transformations for finding a fourth integral of the equations of motion.

The three classical cases when complete integration of the Euler and Poisson equations is possible are the following:

Case of Euler and Poinsot. Euler (1758) considered the case for which the external forces are zero. This case is realized if the fixed point O coincides with the center of mass G of the body, i.e., $\mathbf{r}_0 = 0$, and the resultant of the external forces passes through G. Scalar multiplication of the equation (2) by $\mathbf{I}$ and integration give then the fourth integral I^2 = constant, which expresses the fact that the square of the modulus of the angular momentum vector is constant.

Later, Jacobi (1849) and Somov (1851) expressed the components p, q, r and the Euler angles φ, ψ, ϑ in terms of elliptic functions of t.

The motion of a rigid body under no forces admits the elegant geometric interpretation of Poinsot (1834) which, however, does not give a satisfactory picture of the motion with respect to t. Although p, q, r and the two Euler angles ϑ and ψ are given in terms of t by means of the Jacobi elliptic functions *cn*, *sn*, and *dn*, the third angle φ is given by an integral which is not suitable for computational purposes. In a recent paper Grammel [2] has modified this solution in such a way as to obtain approximate formulas for numerical calculations.

The Poinsot motion is an immediate generalization of the regular precession which is the most general motion of a symmetric body under no external forces. As in the symmetric case, the Poinsot motion can be represented by a superposition of a rotation about the figure axis f, which is fixed in the body, and a rotation about the precessional axis p, fixed in space. However, in the case of an unsymmetric body, these two rotations are no longer uniform but oscillate about certain mean values, the periods of these oscillations being equal to the period of the oscillations of the nutation angle ϑ. Grammel decomposes the "proper rotation" φ and the "precessional rotation" ψ of the Poinsot motion into their mean values and their oscillations about these values. He also gives the Fourier series for the latter, discarding all but first terms in his series expansions of the Jacobian functions in terms of the theta functions.

Case of Lagrange and Poisson. Lagrange (1788) and, independently, Poisson (1815) considered the case $A = B$, $x_0 = y_0 = 0$, i.e., when two of the principal moments of inertia at O are equal and the center of mass G of the body lies on the third axis of inertia of the body.

In this case the last equation in system (6) gives the fourth integral $r = \text{const}$, i.e., the body rotates about its axis of symmetry Oz with constant angular velocity. Jacobi (1849), Somov (1856), and recently Chrapan [1] have shown that, if Euler's angles are used to describe the position of the body, the equations of motion can be integrated in terms of elliptic functions of t. The stability of rotation in the Lagrange case has been discussed recently by Chetaev [1], using a criterion of Lyapunov. Previously Routh solved the stability problem for uniform precession about the vertical, the moments of inertia and the position of G being unrestricted.

Pignedoli [1] examines problems differing but slightly from those of the two cases above. He is concerned (i) with the motions of a heavy rigid body about a fixed point O the distance of which from the center of mass G of the body is infinitesimal, or, in other words, with motions which are close to those of Poinsot (when G coincides with O); and (ii) with the motions of a heavy rigid body about a fixed point O when G is different from O and is slightly off the z-axis, and when the structure of the body with respect to this axis differs slightly from a gyroscopic structure.

Case of Kovalevskaya. In both preceding cases the general solution of the systems of equations (6) and (7) is expressible in terms of elliptic functions of t, and the fourth integral of these systems is a polynomial in p, q, r, i.e., it is algebraic in these variables.

The elliptic functions are single-valued analytic functions for all finite t, except for certain points in the complex t-plane at which these functions have poles of the first order.

These two particular results suggested the following two general problems:

1. To find all cases when the general solution of the systems (6) and (7) is expressible in terms of meromorphic functions of the time t (t being a complex variable).

2. To find all cases when these systems possess a fourth integral which is algebraic in p, q, r.

The first problem was solved by Kovalevskaya (1888), using the analytic theory of differential equations. She showed that only the following four cases are possible:

(a) $A = B = C$ (case of kinetic symmetry)
(b) $x_0 = y_0 = z_0 = 0$ (case of Euler and Poinsot)
(c) $A = B$, $x_0 = y_0 = 0$ (case of Lagrange and Poisson)
(d) $A = B = 2C$, $z_0 = 0$ (case of Kovalevskaya)

Kovalevskaya sought solutions of equations (6) and (7) in the form of power series

$$(8) \qquad p = \frac{1}{\tau^{n_1}} \sum_{n=0}^{\infty} p_n \tau^n, \qquad q = \frac{1}{\tau^{n_2}} \sum_{n=0}^{\infty} q_n \tau^n, \qquad r = \frac{1}{\tau^{n_3}} \sum_{n=0}^{\infty} r_n \tau^n$$

$$\alpha = \frac{1}{\tau^{m_1}} \sum_{n=0}^{\infty} a_n \tau^n, \qquad \beta = \frac{1}{\tau^{m_2}} \sum_{n=0}^{\infty} b_n \tau^n, \qquad \gamma = \frac{1}{\tau^{m_3}} \sum_{n=0}^{\infty} c_n \tau^n$$

where $\tau = t - t_0$, t_0 being the initial time.

Substituting (8) in (6) and (7) we find that

$$n_1 = n_2 = n_3 = 1, \qquad m_1 = m_2 = m_3 = 2$$

The coefficients of the series (8) are determined by corresponding systems of equations obtained upon the above-mentioned substitution.

The results of Kovalevskaya were extended by Lyapunov (1894) who showed that, if A, B, C; x_0, y_0, z_0 are real and A, B, C are all different from zero, the cases of Euler and Poinsot, Lagrange and Poisson, and Kovalevskaya are the only ones in which the parameters p, q, r; α, β, γ are single-valued functions of the time t for arbitrary initial values of these parameters. In all other cases real initial values of the parameters p, q, r; α, β, γ can be chosen in such a way that at least some of them will not be single-valued functions of t.

Appel'rot showed that if $A \neq B \neq C$, there is neither the general solution nor are there particular solutions of the systems (6) and (7), having for p, q, r poles of order higher than one and for α, β, γ poles of order higher than two. If, however, $A = B \neq C$, $y_0 = 0$, $z_0 \neq 0$, there may exist particular solutions, having for p and q poles of order three.

Furthermore, Appel'rot and Lyapunov pointed out a particular case, overlooked by Kovalevskaya, when $a_0 = b_0 = c_0 = 0$, i.e., when all six parameters $p, q, r; \alpha, \beta, \gamma$ have poles of the first order. This is the so-called Hess (1890) case. However, these conditions are only necessary for the existence of a single-valued solution.

The second general problem concerning the existence of a fourth algebraic integral of systems (6) and (7) was investigated by Poincaré, Husson, Burgatti, and others. Poincaré showed that if μ, being the product of the weight of the body and the distance between the fixed point and the center of mass, is small and the initial conditions are arbitrary, then, in order that a single-valued first integral of systems (6) and (7) exists which is not an algebraic combination of the classical integrals and which does not contain implicitly the time t, it is necessary for the momental ellipsoid of inertia of the body at the fixed point O to be an ellipsoid of revolution. Husson arrived at the same conclusion as Poincaré but, for arbitrary μ, requiring the existence of an algebraic first integral. Obviously an algebraic integral is single-valued. Burgatti showed that, if the momental ellipsoid of inertia at the fixed point O is an ellipsoid of revolution, a fourth algebraic integral exists only in the cases of Euler and Poinsot, Lagrange and Poisson, and Kovalevskaya. Hence a fourth single-valued first integral of the systems (6) and (7) exists only in the cases in which the general solution of these equations is expressible in terms of single-valued functions of t for arbitrary initial conditions.

In the case of Kovalevskaya, when $A = B = 2C$, and the center of mass G of the body lies in the plane of equal moments of inertia, the fourth algebraic integral is

$$(p^2 - q^2 - a\alpha)^2 + (2pq - a\beta)^2 = \text{const}$$

where $a = x_0/C$.

Whereas in the cases of Euler and Poinsot, Lagrange and Poisson the general solution can be expressed in terms of elliptic functions of t, in the case of Kovalevskaya the solution is given by means of hyperelliptic functions of t.

In the Kovalevskaya case the body can rotate uniformly about the vertical through O, if G remains on it. Using Lyapunov's "second method" Rumyancev [1] showed that this motion is stable if, and only if, G is below O.

4. Particular solutions

So far we have dealt with cases in which the general solution of the systems of equations (6) and (7) can be found for completely arbitrary initial conditions. Since the above-mentioned theorems of Poincaré, Husson, and Burgatti exhaust all such possible cases, let us consider now certain cases where a solution can be found for a particular choice of the initial conditions or for a particular mass distribution. In other words, we shall list certain cases when, corresponding to special initial conditions or mass distributions, or combinations of both, a fourth algebraic integral exists independent of the three classical integrals.

Case of Hess and Appel'rot. The most important and interesting case is that of Hess (1890) and Appel'rot (1894). Assuming that $A > B > C$, we have

$$y_0 = 0, \qquad x_0^2 A(B - C) - z_0^2 C(A - B) = 0 \tag{9}$$

and the Euler equations admit the fourth integral

$$Apx_0 + Crz_0 = 0 \tag{10}$$

Zhukovskiĭ (1894) showed that the conditions (9) imply that the center of mass G of the body is situated on the axis of one of the cyclic planes of the reciprocal momental ellipsoid of inertia at the fixed point O, and that at the beginning of motion the impulse vector is in this circular cross-section plane. This vector, as shown by Zhukovskiĭ, then remains in this cross-section plane during the whole motion, i.e., this property is characteristic for the Hess and Appel'rot case. Chaplígin (1895) showed, on the other hand, that if one considers the central reciprocal ellipsoid of inertia, constructs the two circular cross-section planes through its center G, erects to each of these planes a perpendicular through G, and chooses any point on one of these perpendiculars for the fixed point O, then the conditions (9) are satisfied.

Whereas in the cases of Euler and Poinsot, Lagrange and Poisson, and Kovalevskaya the existence of the fourth algebraic integral imposes restrictions on the mass distribution of the body, in the case of Hess and Appel'rot the integral (10) restricts the initial angular velocity.

If we put $A = B$ in (9), then for $x_0 = 0$ the case of Lagrange and Poisson follows. $A = B = C$ gives the case of kinetic symmetry. Finally, putting $x_0 = z_0 = 0$, the case of Euler and Poinsot is obtained.

Manacorda [3], seeking to characterize structurally a rigid body which, acted upon by a system of forces the moment of which is perpendicular to OG, has a moment of momentum also perpendicular to OG, arrived at the Hess and Appel'rot conditions for a heavy rigid body.

Case when the center of mass is on the axis of a circular section of the inertia ellipsoid at the fixed point. We have just seen that, when the center of mass G of a heavy rigid body lies on the axis of one of the cyclic planes of the reciprocal momental ellipsoid of inertia at the fixed point O, a fourth algebraic integral (10) exists which is linear in the components of the angular velocity vector $\boldsymbol{\omega}$. In analogy with this case, in recent years several Italian mathematicians such as Grioli, Agostinelli, Zeuli, and Miss Griseri have studied the case when the center of mass G is located on the axis of a circular section of the inertia ellipsoid at O.

Grioli [1], seeking for regular undegenerate precessions among the possible motions of a heavy asymmetric body, obtained a particular solution containing one arbitrary constant. The corresponding motions are uniform rotations of the body about the barycentric axis which, in turn, rotates uniformly and with the same angular velocity about the axis of precession which is orthogonal to the first axis. It is interesting to note that the initial conditions of motion alone determine these precessional motions, provided that O has been properly chosen. The structure of the body can be arbitrary. The same author [2] also investigated the linear stability of the above-mentioned motions.

Colombo [1] has derived the general expression for a moment vector acting on a gyroscope in order that merostatic motions be possible, i.e., regular precessions or uniform rotations about axes different from the gyro axis. From this expression he then derived a general stability criterion and applied it to another case studied by Grioli [3]. The latter author is concerned with the motion of a gyroscope σ about its center of mass G, subject to a force $\mathbf{F} = \mu\mathbf{H} \times \mathbf{v}\, d\sigma$, where $d\sigma$ denotes an element around a point P of the gyroscope, $\mathbf{v}$ the velocity of P, $\mathbf{H}$ a constant vector, and μ a function of P such that $\int_\sigma \mu\, d\sigma$ is different from zero. The ellipsoid of inertia at G, having a mass distribution of density μ is supposed to be an ellipsoid of revolution with axis of rotation coinciding with that of the gyroscope. Examples of the above-mentioned type of forces are (i) the compound centrifugal force due to the rotation of the Earth and (ii) a force to which a gyroscope is subjected if it carries an electric charge distribution μ and is immersed in a uniform magnetic field $\mathbf{H}$.

Agostinelli [1, 2] proved the existence of a solution, depending upon five arbitrary constants. The integrals of motion are given by uniformly convergent power series in terms of a parameter s, which is proportional to the component of the angular momentum along the barycentric axis. Recursion formulas are given for the coefficients of these series in terms of four arbitrary constants. If no conditions are imposed upon the moments of inertia of the body, the only polynomial solution in s is obtained for ω_b = const, where ω_b is the component of the angular velocity $\boldsymbol{\omega}$ along the barycentric axis. The corresponding motions are the regular precessions obtained by Grioli. Another polynomial solution in s can be obtained under the assumption that ω_b is a linear function of s. The moments of inertia, then, have to satisfy a certain relationship.

Zeuli [1] studied the motions of a heavy rigid body about a fixed point O under the assumption that the structure of the body is very close to that of a symmetric body. He expands the unknown functions p, q, r; α, β, γ in power series of a parameter, the coefficients being functions of the time. For the zero value of the parameter the problem reduces to that of a heavy symmetric body.

Miss Griseri [1] studied solutions which are meromorphic in the plane of the complex variable t (time) with a pole at the origin. The existence of three such solutions is proved; although two of them do not impose any restrictions on the moments of inertia of the body, the third solution exists only if these structural constants satisfy a certain relation.

Case of Staude. Staude (1894) showed that the motion under gravity of an asymmetric rigid body, one point O of which is fixed, can be a uniform rotation about each of a system of ∞^1 axes through O, when such an axis is put in a vertical position. In such a case the angular velocity vector remains constant in direction and magnitude with respect to the space and the body. The axes of rotation form a cone of the second order (fixed within the body) the equation of which is

$$\boldsymbol{\omega}\cdot(\mathbf{r}_0 \times \mathbf{I}) = 0$$

Van der Woude showed that each axis a' of permanent rotation is a principal axis of inertia of the body at one point of a' which is different from O, the x-, y-, and z-axes being principal axes of inertia at O. A simple geometric proof of this fact was given by Manarini [1, 2]. He also showed that the period of rotation is given by the well-known Galileian formula for the period of oscillations of a simple pendulum, if, instead of taking the length of the pendulum, the distance between the fixed point O and that particular point with respect to which the axis of rotation is a principal axis of inertia of the body is taken.

The stability of the Staude rotations has been investigated by Hadamard, Grammel [1], Stoewa [1], and Bottema [1].* Hadamard tried to obtain explicit conditions of stability for the given mass distribution. Such a problem, however, turned out to be hopeless. Grammel considered the point O, the axis of rotation, and the principal moments of

* Recently, Rumyancev, V. V., Stability of permanent rotations of a heavy solid, *Prikl. Mat. Mekh*, **20**, 1, 51-66 (1956), using Chetaev's method for finding Lyapunov's stability function, has obtained conditions for determining the stability zone on the cone of permanent axes.

inertia at O as given, and determined the region in which the center of mass must lie in order that the uniform rotation be stable. Stoewa discussed the stability of the Staude rotations, using three criteria. The criterion of Lagrange and Lejeune-Dirichlet for the stability of motion reads

$$(11) \qquad L_0 = U + \frac{\omega^2}{2} C = \max \qquad (\omega \text{ being constant})$$

where $U = -(x_0 x + y_0 y + z_0 z)$ is the force function and C is the moment of inertia with respect to the axis of rotation.

The Lord Kelvin and Routh criterion of stability requires

$$(12) \qquad L_0{}^* = U + \frac{I^2}{2C} = \max \qquad (I \text{ being constant})$$

where $I = \omega C$ is the impulse moment. The third criterion of stability is derived from the point of view of first-order variations from known permanent rotations.

Stoewa shows that if condition (11) is satisfied, condition (12) also holds, i.e., from $L_0 = \max$ follows $L_0{}^* = \max$; and if condition (12) is satisfied, i.e., if $L_0{}^* = \max$ holds, then the third criterion also holds. However, it is also possible that the third criterion of stability is satisfied when condition (12) shows instability of rotation, i.e., when $L_0{}^* = \min$. Since Grammel used the third criterion to determine the stability of the Staude rotations, Stoewa concludes that there will not be stability of rotation for all values of ω as given by Grammel.

Bottema considered a more direct question than Grammel; namely, given a rigid body with a fixed point O, what axes of Staude are stable? Assuming that the center of mass lies on one of the principal axes of inertia at O, he solved the problem completely.

Cases of Steklov and Bobylev. Steklov (1896) and Bobylev (1896) showed that when $B = 2A$, and the center of mass of the body is on the y-axis of the ellipsoid of inertia at 0, i.e., when $x_0 = z_0 = 0$, the systems of equations (6) and (7) can be integrated and the solution depends upon three arbitrary constants. In fact, assuming that

$$r = 0, \qquad q = q_0 = \text{const}, \qquad p = n\alpha \qquad \left(n = \frac{y_0}{Aq_0}\right)$$

is a particular solution of these two systems of equations, system (7) assumes the form

$$\dot{\alpha} = -q_0\gamma, \qquad \dot{\beta} = n\alpha\gamma, \qquad \dot{\gamma} = \alpha(q_0 - n\beta)$$

The integrals of this last system together with the above-mentioned particular solution constitute the solution of the problem.

If $B = C$, the case of Kovalevskaya is obtained.

Steklov (1899) dealt also with another integrable case, namely $B > A > 2C$, $y_0 = z_0 = 0$. He discussed in detail the subcase $A < B$ and showed that p, q, and r are proportional to cn kt, sn kt, dn kt respectively, where k is determined by A, B, C and the distance between the points O and G. Recently Kuz′min [1] carried over Steklov's results to the subcase $A > B$ and obtained the same expressions for p, q, and r but with a different value of k. In the case of Steklov as well as that of Kuz′min the solution depends upon a single arbitrary constant.

Cases of Goryachev and Chaplígin. Goryachev (1900) showed that when $A = B = 4C$, the center of mass of the body is in the plane of equal moments of inertia at O, i.e., $z_0 = 0$ (by a rotation of the trihedral $Oxyz$ about the z-axis we can make the center of mass of the body lie on the x-axis, i.e., also $y_0 = 0$) and the initial angular momentum of the body about the vertical through O is zero, i.e., $c = 0$, then a fourth integral

$$p^2 + q^2 = C_4 p^{2/3} \qquad (C_4 = \text{constant of integration}) \tag{13}$$

can be found and the solution depends upon three arbitrary constants.

Chaplígin (1901) improved this result of Goryachev and obtained a solution which depends upon four arbitrary constants. Let

$$s = (p^2 + q^2)r - ap\gamma$$

and calculate ds/dt. Then, due to the Euler and Poisson equations (6) and (7), we obtain

$$\frac{ds}{dt} = -\frac{aq}{4}[4(p\alpha + q\beta) + r\gamma]$$

The angular momentum integral in this case reads

$$4(p\alpha + q\beta) + r\gamma = \frac{c}{C}$$

Hence,

$$\frac{ds}{dt} = -\frac{aq}{4}\frac{c}{C}$$

If $c = 0$, then $ds/dt = 0$, and $s =$ constant. Hence

$$(p^2 + q^2)r - ap\gamma = C_4' \qquad (C_4' = \text{constant})$$

is the fourth integral, the left-hand side of which is of degree three with respect to the components of the angular velocity $\boldsymbol{\omega}$ of the body.

Using Chaplígin's formulas and results, Sretenskiĭ [1] investigated at length the motion of the Goryachev and Chaplígin gyroscope, assuming that initially the body is rotating with a large angular velocity about the axis through the fixed point and the center of mass of the body. The motions obtained are analogous to the quasi-regular precessions of the Lagrange gyroscope.

In the case of Goryachev one of the classical first integrals is free from arbitrary constants, since we assumed that $c = 0$. Several authors, such as Goryachev (1897), Steklov (1899), and Chaplígin (1904), set out to find new integrable cases, assuming that one algebraic first integral, independent of t and of arbitrary constants and similar to the classical first integrals, exists. The most general results obtained by these authors are those of Chaplígin.

Assuming that $y_0 = z_0 = 0$, he sought conditions for the existence of particular integrals of the form:

$$x_0\beta = \delta pq + \lambda p^n q, \qquad x_0\gamma = \epsilon pr + \mu p^n r$$

where δ, ϵ, λ, and μ are undetermined constants. He showed that only three cases are possible:

(*a*) $\lambda = \mu = 0$ (case of Steklov)
(*b*) $\lambda = 0, n = 3$ (case of Goryachev)
(*c*) $n = -\frac{1}{3}$, $0.5965 < C/A < 0.6$, $1.5 < B/A < 1.5965$ (new case)

Case of Mercalov. Under the same first two assumptions as by Goryachev, i.e., (i) $A = B = 4C$, and (ii) the center of mass of the body is in the equatorial plane of the ellipsoid of inertia at the fixed point O, Mercalov [1] obtained a fourth integral of the equations of motion when $c \neq 0$. Discarding the case $q = q_0 = \text{const}$ (Steklov-Bobylev) and assuming that the center of mass of the body lies on the y-axis of the ellipsoid of inertia at O, i.e., $x_0 = z_0 = 0$, he introduces a new variable ζ where

$$\frac{d\zeta}{dt} = p$$

and his fourth integral reads

$$(p^2 + q^2)\frac{dq}{d\zeta} - \frac{3}{2}q\frac{d}{d\zeta}(p^2 + q^2) = -\frac{3}{4}\frac{n_1\zeta}{C} + C_4' \qquad \left(n_1 = \frac{cy_0}{A}\right)$$

For $c = 0$ this last integral reduces to that of Goryachev (13) (after interchanging the variables p and q). Hence a fourth integral of the equations of motion exists also in the case when $c \neq 0$, i.e., when the initial angular momentum of the body about the vertical through O is different from zero.

Further, assuming that the initial values $p_0 = q_0 = 0$, Mercalov gives an approximate expression for the component q of the angular velocity on the barycentric axis.

Cases of Kowalewski, Corliss, and others. Kowalewski (1908) assumed that the center of mass G of the body lies on one of the principal axes of inertia at O, say, on the z-axis, i.e., $x_0 = y_0 = 0$, and sought all cases for which p^2 and q^2 can be expressed as polynomials of degree three in r. In doing so he arrived at the three particular solutions of Goryachev, Steklov, and Chaplígin and, in addition, at one new case. Whereas in the first three cases the solution is given in terms of elliptic functions, in the new case hyperelliptic functions are needed to represent the motion.

Field [1] considered a case related to that of Kowalewski and characterized by the conditions $C^2 = 2AB$, $c = 0$, the constant of energy h having a well-determined value. The solution is given in terms of elliptic integrals.

The investigations of Kowalewski have been continued by Corliss [1], Fabbri [1–3] and De Angeli [1]. In analogy with the Kowalewski case, Corliss assumes that (i) the center of mass G of the body lies on the z-axis, i.e., $x_0 = y_0 = 0$, and (ii) the projection of the angular momentum upon the vertical is zero, i.e., $c = 0$. Further, he introduces two new variables u, and v, which are closely related to the Hess invariants T, and U (Art. 6), and requires that the new variables u, v be quadratic functions in r. As a consequence of the assumptions made, he is led to two new particular solutions, the second being a special case of that of Kovalevskaya. In a second paper, Corliss [2] proposed to find all cases in which T and U are expressible as polynomials in r. For this purpose he derived two symmetric differential equations of the second order for his variables u and v, and sought to solve them by polynomials of degree n in r. It turns out that n cannot be greater than 4. The cases $n = 1$ and $n = 2$ lead to the two cases given in his first paper. The case of Field is a special case of that for $n = 2$. The case $n = 3$ corresponds to the case of Kowalewski. For $n = 4$ a new case is obtained.

Center of mass lies in the characteristic plane. Myasnikov [1] proposed a new method for finding integrable cases, assuming that the center of mass G of the body lies in the so-called characteristic plane determined by the angular velocity vector and the angular momentum vector. This assumption yields a fourth algebraic integral

$$Apx_0 + Bqy_0 + Crz_0 = m$$

where m is a constant.

If G lies simultaneously in the characteristic plane and on one of the principal axes of inertia at O, say, on the z-axis, i.e., $x_0 = y_0 = 0$, the cases of Euler and Poinsot, Lagrange and Poisson, and Steklov and Bobylev are obtained. When G is in the characteristic plane and in a principal plane of inertia at O, say, in $x_0 = 0$, then, if $p \neq 0$, uniform rotations of Staude about the vertical, or the pendular motions of Mlodzeevskiĭ about the horizontal x-axis, or the Euler and Poinsot case, or the spherical gyroscope is obtained. When G lies in the plane $z_0 = 0$, the same cases are possible. If, however, G is in the plane $y_0 = 0$, a loxodromic pendulum is possible. Hence, Myasnikov's method yields all the known integrable cases from a unified point of view.

In order to obtain new possible cases he assumes that G is in the plane $x_0 = 0$ and $p = 0$. The Euler and Poisson equations then reduce to

$$(6a) \qquad (C - B)qr = \beta z_0 - \gamma y_0, \qquad B\dot{q} = -\alpha z_0, \qquad C\dot{r} = \alpha y_0$$

$$(7a) \qquad \dot{\alpha} = r\beta - q\gamma, \qquad \dot{\beta} = -r\alpha, \qquad \dot{\gamma} = q\alpha$$

In his review of the Myasnikov paper [*Math. Rev.* **15,** 996 (1954)] Wundheiler pointed out, however, that in the case $m \neq 0$, $B - C \neq 0$ the Staude permanent rotations about the vertical follow, whereas in the case $m = 0$, $B - C = 0$ the pendular motions of Mlodzeevskiĭ about the horizontal y-axis are obtained. Hence all the cases obtained were known previously, contrary to the opinion of Myasnikov who, by his rather advanced treatment of the equations (6a) and (7a) thought he had arrived at a new integrable particular case.

Case when the center of mass lies on one of the principal planes of inertia. Investigations which in a certain sense are parallel to those of Kowalewski have been carried out by Manacorda [2] and Nadile [1], who assumed that the center of mass G lies in one of the principal planes of inertia of the body relative to the fixed point O. In the terminology of Stäckel this is the so-called planar case of the motion of a rigid body.

Manacorda sets up two invariant relations in terms of the components of the angular velocity vector as power series in s, s being a certain linear combination of these components. For only the zero value of a certain exponent does he arrive at the general solution depending upon five arbitrary constants. He also shows how the cases of Hess ($s = 0$), Grioli, and Kowalewski fit into his theory.

Nadile finds two types of uniform rotations and certain other motions involving only quadratures and integration of a Riccati equation. These latter motions depend upon five arbitrary constants of integration.

Motion under forces different from gravity. The problem of motion of a rigid body about a fixed point under the influence of forces different from gravity has been discussed by Kolosov, Goryachev [1, 2], Tallqvist [1], De Simoni [1], Udeschini [1], Arrighi [1], Koshlyakov [1], Alfieri [1], Braunbeck [1], and others.

Goryachev solved the problem of finding all cases when the Euler equations admit an integral which is a polynomial of degree three or four in p, q, r, and determined in each of these cases the corresponding force function U as a function of the position of the body. Under the further assumption that the projection of the angular momentum on the vertical is zero, i.e., $c = 0$, he obtained in each of the two cases mentioned above the solution in terms of four arbitrary constants.

Tallqvist pointed out several new integrable cases of the Lagrange and of the Kovalevskaya gyroscopes, subjected to forces distinct from gravity.

De Simoni studied the motion of a rigid body about a fixed point O the distance d of which from the center of mass G is small enough so that $d^n (n \geq 2)$ can be neglected (quasi-barycentric suspension). He considered, in particular, the case when the moment of momentum is parallel to a constant resultant force applied to the body.

Udeschini discusses the motion of a heavy top spinning on a horizontal plane and which is subject to forces (in addition to those of gravity and reaction of the plane) consisting of a couple with respect to the center of mass G and of a horizontal force applied at G and intersecting a fixed vertical line.

Arrighi considers the motion of a symmetric rigid body suspended frictionlessly from its center of mass G and subject to magnetostatic forces due to two dipoles: one fixed in space with its axis in line with G, the other located along the axis of symmetry of the body and rigidly connected with it. Such a problem may be of some interest in atomic physics. The problem of finding the Eulerian angles which determine the position of the body is reduced to the integration of a single first-order differential equation followed by two quadratures. The question of regular precessions is discussed in detail, examining various possible combinations of cases with elongated and oblate central ellipsoid of inertia and cases with acute or obtuse angle of nutation.

Koshlyakov assumes that the equatorial moments of inertia A and B ($B > A$) of a body differ only slightly from each other and that the axial moment C is greater than A. Furthermore, he assumes that the components of the moment generated by the resistance forces are $-\lambda Ap$, $-\lambda Bq$, $-\lambda Cr$, where λ is a certain coefficient of proportionality. A solution of the Euler equations is sought in the form of power series in terms of the small parameter $\epsilon = (B - A)/C$. The method of successive approximations is applied in such a way that the zero approximation corresponds to the symmetric case $A = B$. In the latter case and under the further assumption that the moments of the resisting forces are of certain particular forms, the Euler equations are integrable in terms of Bessel functions and a degenerate hypergeometric function.

Alfieri solved a problem which includes that of Staude, namely, to determine all possible uniform rotations of a rigid body about a fixed point O different from its center of mass G, when the system of acting forces is equivalent to a constant force applied to a point distinct, in general, from both O and G.

Braunbeck considers the motion of a gyroscope suspended at the center of mass G and subject to a torque

$$\mathfrak{M} = \mu[\mathbf{a} \times (\mathbf{H}_0 + \mathbf{H}_1)]$$

generated as follows. A magnetic bar with magnetic moment μ is situated along the figure axis of a gyroscope, subject to a homogeneous magnetic field $\mathbf{H} = \mathbf{H}_0 + \mathbf{H}_1$, where $\mathbf{H}_0$ is a constant field and $\mathbf{H}_1$ is an alternating field which varies periodically with respect to the time. $\mathbf{a}$ denotes the unit vector in the direction of the figure axis of the gyroscope. Subject to the constant field $\mathbf{H}_0$ alone, the motion of the gyroscope is identical to that of a gyroscope under the influence of gravity. The general problem with $\mathbf{H} = \mathbf{H}_0 + \mathbf{H}_1$ is of interest in astronomy (perturbation of the rotational motion of the Earth around its axis under the influence of forces arising from the planetary system) and in atomic physics (electrons and nuclei in a high-frequency magnetic field may be considered as atomic gyroscopes subject to their mechanical and magnetic moments). Braunbeck studies extensively two cases: (i) the alternating field $\mathbf{H}_1$ is parallel to the constant field $\mathbf{H}_0$, and (ii) $\mathbf{H}_1$ is orthogonal to $\mathbf{H}_0$. The nonlinear differential equations of motion are replaced by certain physically plausible approximate equations from which many interesting phenomena, such as resonances and subharmonic resonances, are deduced.

5. Self-excited asymmetric rigid body

From the mathematical point of view Euler considered the simplest case of motion of a rigid body about a fixed point; namely, when no external forces act on the body.

The next simplest case we can imagine is that of the motion of a rigid body about a fixed point subject to forces with a fixed torque direction in the body. Before the modern use of jet propulsion such a problem seemed to lack a physical meaning. With reference to the Eulerian equations, constancy of the torque direction in the body means that the torque changes its direction as the body continues to rotate about the fixed point. However, since the introduction of apparatus with internal reactions such a problem is of the greatest importance.

Grammel [4] defines a rigid body which is free to rotate about a point fixed in the body and in the space as "self-excited" when it is acted on by a torque vector $\mathfrak{M}$ arising from internal reactions. Two cases are possible, depending on whether $\mathfrak{M}$ is fixed in the body or moves in it in a prescribed manner. We shall refer to these two cases as self-excitements with fixed or moving direction within the body respectively. In the first case two subcases are to be distinguished, depending on whether the modulus of $\mathfrak{M}$ changes with the time or not. These are the time-dependent or time-independent self-excitements with a fixed direction in the body.

Bödewadt [1] showed that the integration of the Euler equations of motion of a symmetric rigid body ($A = B$) with time-independent self-excitement in a fixed direction in the body can be reduced to quadratures. The solution is given in terms of integrals of the Fresnel type. In a somewhat similar way the solution of the Euler equations of motion of a symmetric rigid body can be found for the case of a time-dependent self-excitement in a fixed direction in the body. Very little, however, is known concerning the motion of an asymmetric self-excited rigid body.

Grammel [3], considering the general case of a rigid body with a fixed torque vector $\mathfrak{M}$ in it, is first concerned with motions of fixed angular velocity vector $\boldsymbol{\omega}$, i.e., with motions for which $\dot{p} = \dot{q} = \dot{r} = 0$, and with the discussion of their stability. He shows that in the case of a symmetric body there is a family of ∞^3 permanent rotations that are "almost all" stable, whereas in the case of an asymmetric body there is a family of ∞^3 permanent rotations that are "almost all" unstable. He also clarifies the apparent discrepancy between the last statement and the results on the stability of the Staude permanent rotations. The Staude cone contains besides an infinity of unstable axes of permanent rotation also an infinity of axes that are stable. In the case of the Staude permanent rotation the torque vector, fixed within the body, is the moment vector $\mathfrak{M}_g$ of gravity about the fixed point O. Such a torque vector, however, can also be generated by the internal reactions instead of by gravity. A disturbance of a permanent rotation changes the vector $\mathfrak{M}_g$ into a vector of variable direction in the body. This is so because of the displacement of the center of mass G with respect to O. However, a torque vector which by its nature is of fixed direction within the body, in general, remains also fixed after the disturbance.

Grammel also shows that instability of permanent rotations can be avoided by making the torque vector not strictly constant but depending on $\boldsymbol{\omega}$. Furthermore he proves that the only steady motions which an asymmetric rigid body with a fixed torque vector in the body can exhibit are the permanent rotations.

Given the angular velocity vector $\boldsymbol{\omega}$, fixed or variable within the body, the torque $\mathfrak{M}$ of the required force is given by the Euler equations. Grammel proposed to attack the inverse problem; i.e., given the torque vector $\mathfrak{M}$ with respect to the body, to find the corresponding angular velocity vector $\boldsymbol{\omega}$ and, finally, the Eulerian angles. This problem presented considerable difficulties inasmuch as $\boldsymbol{\omega}$ enters nonlinearly in the Euler equations. This circumstance forced Grammel to use some nonlinear procedure such as the iteration method of approximations for solving the Euler equations. Except for certain special cases Grammel was not able to obtain the general solution for an asymmetric rigid body with time-independent or time-dependent self-excitement $\mathfrak{M}$ with fixed or variable direction in the body, but he came close to such a solution, successively improving the degree of approximation, and, within certain domains of convergence, even reaching it. In other words, Grammel has generalized the Poinsot motions (under no external forces) to a variety of new motions corresponding to various kinds of torques due to internal reactions.

THE HESS AND SCHIFF EQUATIONS

6. The Hess and Schiff equations

Hess suggested that instead of p, q, r the invariants S, T, U be introduced as the new variables, where

(14) $$S = \mathbf{r}_0 \cdot \mathbf{I}$$

(15) $$2T = \boldsymbol{\omega} \cdot \mathbf{I}$$

(16) $$2U = \mathbf{I} \cdot \mathbf{I}$$

These three variables have a simple physical meaning: S is the spin, T is the kinetic energy, and U is one half of the square of the angular momentum vector of a rigid body.

Schiff modified the Hess equations and reduced them to a particularly convenient form.

Consider the system of equations

(17) $$\mathbf{r}_0 \cdot \mathbf{K} = h - T$$
$$\mathbf{I} \cdot \mathbf{K} = c$$
$$(\mathbf{r}_0 \times \mathbf{I}) \cdot \mathbf{K} = \dot{U}$$

where the first two equations are (4) and (5) respectively, and the last equation is obtained from (16) by differentiation with respect to t. The system (17) gives the components of $\mathbf{K}$ in a right-hand trihedral, determined by the vectors $\mathbf{r}_0$, $\mathbf{I}$ and $\mathbf{r}_0 \times \mathbf{I}$ constructed at O.

Solving this system of equations with respect to $\mathbf{K}$, we obtain

$$(18) \quad \Delta\mathbf{K} = [2(h - T)U - cS]\mathbf{r}_0 + [cr_0^2 - (h - T)S]\mathbf{I} + \dot{U}(\mathbf{r}_0 \times \mathbf{I})$$

where

$$\Delta = (\mathbf{r}_0 \times \mathbf{I})^2 = 2r_0^2U - S^2$$

and Δ is positive unless

$$\mathbf{r}_0 \times \mathbf{I} = 0$$

i.e., unless (1) $\mathbf{r}_0 = 0$ (case of Euler and Poinsot), (2) $\mathbf{I} = 0$ (case of rest), or (3) $\mathbf{r}_0$ is parallel to the angular momentum vector $\mathbf{I}$.

Substituting (18) in (2) we obtain the equation of motion in the form of

$$(19) \quad \dot{\mathbf{I}} + \boldsymbol{\omega} \times \mathbf{I} = \frac{1}{\Delta}\{-S\dot{U}\mathbf{r}_0 + r_0^2\dot{U}\mathbf{I} - [cr_0^2 - (h - T)S](\mathbf{r}_0 \times \mathbf{I})\}$$

Scalar multiplication of equation (19) by $\mathbf{r}_0$ and $\boldsymbol{\omega}$ and of equation (18) by $\mathbf{K}$ yields the Schiff scalar equations

$$(20) \quad \dot{S} = \boldsymbol{\omega}\cdot(\mathbf{r}_0 \times \mathbf{I})$$

$$(21) \quad \Delta\dot{T} = [(h - T)S - cr_0^2]\dot{S} + [2r_0^2T - S(\mathbf{r}_0\cdot\boldsymbol{\omega})]\dot{U}$$

$$(22) \quad \dot{U}^2 = r_0^2(2U - c^2) - 2(h - T)^2U + 2cS(h - T) - S^2$$

This last system of equations does not contain the time t explicitly. Therefore, substituting for $\dot{U}$ in (21) its value obtained from equation (22), we obtain, together with equation (20), a system of two first-order differential equations with respect to S and T the coefficients of which are algebraic.

7. Equivalence between the Euler-Poisson equations and the Hess-Schiff equations

Stäckel (1909) and Lazzarino [1–4] discussed at length the Hess and Schiff equations with reference to their equivalence to the Euler and Poisson equations. In general there is no equivalence between these two systems of equations. Lazzarino obtained criteria for the equivalence or nonequivalence in terms of the nonvanishing or vanishing of the time derivatives of the Hess invariants S, T, and U. He proved the following theorems:

Theorem 1. If the invariants S and U are functions of the time t, the Hess and Schiff equations are equivalent to those of Euler and Poisson.

Theorem 2. If any two of the Hess invariants are independent of t, then also the third invariant is independent of t. The corresponding motion of the rigid body is a permanent rotation of Staude about the vertical.

Theorem 3. There are two singular cases: U = const and S = const for which the equivalence mentioned in Theorem 1 does not exist.

The case $U = U_0$ = const, previously analyzed incorrectly by Schiff and Stäckel, has been discussed by Lazzarino and Hamel [1]. In this case, one of the Hess and Schiff equations is a consequence of the remaining two equations. Therefore, to re-establish the equivalence of the Hess and Schiff equations and those of Euler and Poisson, it is necessary to obtain an auxiliary equation, independent of the remaining two Hess and Schiff equations. This auxiliary equation is

$$[2U_0(h - T) - cS][(\boldsymbol{\omega}\cdot\mathbf{r}_0)S - 2r_0^2T] - [S(h - T) - cr_0^2]^2 = 0$$

The only possible motions of the rigid body are permanent rotations. If, in particular, the barycentric axis OG is vertical, the body rotates uniformly about this axis. If the ellipsoid of inertia at the fixed point O is an ellipsoid of revolution, the axis of symmetry or any axis perpendicular to that axis may be taken for the axis of uniform rotation. If the ellipsoid of inertia at O reduces to a sphere with center at O, every diameter of this sphere, fixed in the body, is an axis of uniform rotation.

In the singular case $S = S_0 \neq 0$ (S_0 = const), two subcases are to be distinguished: (i) The plane $S = S_0$ cuts the Staude cone, and (ii) the Staude cone degenerates into a pair of planes.

Subcase (i). In this case one of the Hess and Schiff equations is a consequence of the remaining two equations, but the above-mentioned equivalence can be re-established by introduction of the auxiliary equation

$$\mathbf{r}_0\cdot\left(\frac{d\mathbf{I}}{dt}\times\mathbf{I}\right) = 2U(\mathbf{r}_0\cdot\boldsymbol{\omega}) - (h - T)S_0 + cr_0^2$$

obtained by Lazzarino and Hamel on multiplying equation (19) by $\mathbf{r}_0 \times \mathbf{I}$. The only possible motions are permanent rotations of Staude.

Subcase (ii). In order that the Staude cone degenerates, it is necessary and sufficient that the ellipsoid of inertia at O be an ellipsoid of revolution or that the barycentric axis moves in one of the principal planes of inertia at O. In this last case the motion of the rigid body is said to be "planar," and the principal plane of inertia at O which contains the barycentric axis OG is called the first Staude plane, the other plane being called the second Staude plane.

If $y_0 = 0$ is the equation of the first Staude plane, the equation

$$p(A - B)z_0 + q(B - C)x_0 = 0$$

represents the second Staude plane. The condition that the second Staude plane coincide with the plane $S = S_0$ reads

$$A(B - C)x_0^2 - C(A - B)z_0^2 = 0$$

A rigid body, the motion of which is "planar," is called the Hess or the Mlodzeevskiĭ body, depending upon whether its second Staude plane coincides or does not coincide with the plane $S_0 = 0$ respectively.

For $S_0 = 0$, equations (14) and (20) become

$$\mathbf{r}_0 \cdot \mathbf{I} = 0 \tag{23}$$

$$\boldsymbol{\omega} \cdot (\mathbf{r}_0 \times \mathbf{I}) = 0 \tag{24}$$

and represent a plane through O perpendicular to the barycentric axis and the Staude cone respectively.

If the equations (23) and (24) are independent of each other, the plane $S_0 = 0$ can intersect the Staude cone at (*a*) its vertex O only, or (*b*) along two of its elements, or (*c*) touch it along an element. In the case (*a*) the body rests, in the case (*b*) the motion of the body is "planar," and in the case (*c*) the axis of rotation coincides with an element of the Staude cone (element of contact) and is of fixed direction in the body and in the space. In this last case, according to whether the angular velocity of the body is a constant or a function of the time t, we have permanent rotations of Staude around the vertical, or pendular motions of Mlodzeevskiĭ around the third principal axis of inertia which is horizontal. In the last case the center of mass of the body oscillates in a vertical plane.

If equations (23) and (24) are not independent of each other, the plane $S_0 = 0$ coincides with one of the planes into which the Staude cone degenerates. During the whole course of motion the other plane contains the barycentric axis and coincides with one of the principal planes of inertia at O, i.e., we have the Hess rigid body.

THE EULER SPACE

8. The configuration space. The Euler space

The position of a heavy rigid body with a fixed point is determined by three parameters, for example, the Eulerian angles φ, ψ, ϑ. Considering these parameters as coordinates of a point in a three-dimensional space, the motion of a rigid body with a fixed point can be interpreted by the motion of a point $(\varphi, \psi, \vartheta)$ in a three-dimensional space, called the configuration space of the rigid body. It is convenient to introduce instead of three independent parameters, such as φ, ψ, ϑ, four parameters of Cayley ξ_1, ξ_2, ξ_3, ξ_4, connected by the relation

$$\xi_1^2 + \xi_2^2 + \xi_3^2 + \xi_4^2 = 1 \tag{25}$$

The correspondence between the positions of a rigid body and the sets of values of the four parameters ξ_i $(i = 1, 2, 3, 4)$, satisfying the relation (25), is not a one-to-one correspondence, since the two sets of values ξ_i and $-\xi_i$ determine one and the same position of the rigid body. Interpreting the Cayley parameters ξ_i as the rectangular coordinates of a point in a four-dimensional Euclidean space E_4, equation (25) determines a hypersphere in this space. Hence, the configuration space of a rigid body with a fixed point is homeomorphic to a hypersphere of E_4, diametrically opposite points of which are considered to be identical. In other words, the configuration space of a rigid body with a fixed point is homeomorphic to a three-dimensional projective space.

The components ω_i $(i = 1, 2, 3)$ of the angular velocity $\boldsymbol{\omega}$ (with respect to a moving trihedral fixed in the body) are given in terms of the Cayley parameters and their derivatives by the expressions

$$\begin{aligned} \omega_1 &= 2(\xi_4\dot{\xi}_1 + \xi_3\dot{\xi}_2 - \xi_2\dot{\xi}_3 - \xi_1\dot{\xi}_4) \\ \omega_2 &= 2(-\xi_3\dot{\xi}_1 + \xi_4\dot{\xi}_2 + \xi_1\dot{\xi}_3 - \xi_2\dot{\xi}_4) \\ \omega_3 &= 2(\xi_2\dot{\xi}_1 - \xi_1\dot{\xi}_2 + \xi_4\dot{\xi}_3 - \xi_3\dot{\xi}_4) \end{aligned} \tag{26}$$

Consider the Cayley parameters ξ_j $(j = 1, 2, 3, 4)$ as the components of a radius vector $\boldsymbol{\xi}$ of a point on the unit hypersphere (25) in E_4. Further introduce three mutually orthogonal unit vectors

$$\underset{1}{\mathbf{e}}(\xi_4, \xi_3, -\xi_2, -\xi_1), \quad \underset{2}{\mathbf{e}}(-\xi_3, \xi_4, \xi_1, -\xi_2), \quad \underset{3}{\mathbf{e}}(\xi_2, -\xi_1, \xi_4, -\xi_3) \tag{27}$$

lying in the three-dimensional tangent plane to the hypersphere at the point ξ_j. Then it follows from (26) and (27) that

$$\omega_i = 2\underset{i}{\mathbf{e}}\cdot\dot{\boldsymbol{\xi}} \qquad (i = 1, 2, 3) \tag{28}$$

If one introduces on the hypersphere an arbitrary curvilinear system of coordinates x^α, the vectors $\underset{i}{\mathbf{e}}$ are determined by their coordinates $\underset{i}{e}^\alpha$ with respect to this system, and (28) can be rewritten in the form

(29) $$\omega_i = 2\underset{i}{e}_\alpha \dot{x}^\alpha$$

The kinetic energy of a rigid body moving about a fixed point O is given by the formula

(30) $$T = \tfrac{1}{2}\sum_1^3 I_i\omega_i^2$$

where I_i $(i = 1, 2, 3)$ denote the principal moments of inertia of the body at O.

Introduce a metric in the configuration space by means of the kinetic line element, defined by

(31) $$ds^2 = 2T\,dt^2 = \sum_1^3 4I_i(\underset{i}{e}_\alpha\,dx^\alpha)^2$$

Such a three-dimensional Riemannian space, being the configuration space of a rigid body moving about a fixed point under inertia, is called an Euler space. Obviously, a space of constant positive curvature of the elliptic type is a particular case of an Euler space (when the ellipsoid of inertia at O is a sphere).

Vagner [1] has shown that necessary and sufficient conditions for a three-dimensional Riemannian space to be an Euler space are that the roots r_i $(i = 1, 2, 3)$ of the characteristic equation of the Ricci tensor be constant and negative and that the Ricci principal directions determine an orthogonal geodesic net.

If the conditions mentioned are satisfied, the principal moments of inertia of a rigid body are given by the formulas

$$I_i = -\frac{2r_i}{r_i^2 + \underset{2}{R}}$$

where

$$\underset{2}{R} = r_1r_2 + r_1r_3 + r_2r_3$$

Vagner also gives an invariant characteristic of the Euler space: the invariants $\underset{1}{R}$, $\underset{2}{R}$, $\underset{3}{R}$ of the Ricci tensor must be constant and satisfy the inequalities

$$\underset{1}{R} < 0, \qquad \underset{2}{R} > 0, \qquad \underset{3}{R} < 0$$

and the Ricci tensor must satisfy the condition

$$R^{\alpha\delta}\,\nabla_\alpha R_{\delta\gamma} = 0$$

To the motion of a rigid body about a fixed point under the influence of no external forces corresponds the motion of a point along a geodesic in the corresponding Euler space.

Consider a rigid body with a fixed point moving in a conservative field of forces. In order that this motion can also be interpreted as the motion of a point along a geodesic in the corresponding configuration space, the metric of this space is to be defined by means of the action line element:

$$ds^2 = 2(h - V)T\,dt^2 \tag{32}$$

where h is the energy constant and V the potential energy.

Comparison of (32) with the fundamental quadratic form (31) of the Euler space shows that the new configuration space is a three-dimensional Riemannian space which can be mapped conformally on the Euler space. Hence the spaces, conformal to the Euler spaces, are the most general three-dimensional Riemannian spaces of interest, if one studies from a geometric point of view the motion of a rigid body around a fixed point in a conservative field of forces.

Vagner [1] also gives a necessary and sufficient condition for a given three-dimensional Riemannian space with a given metric tensor to be mapped conformally on an Euler space, different from a space of constant curvature.

9. Motion about a fixed point subject to nonholonomic constraint

Finally Vagner [2] treats a new case of the motion of a rigid body the center of mass of which is fixed and the motion of which is subject to a nonholonomic constraint. This constraint is realized by fastening to the body two diametrically opposite coplanar integraph wheels which are made to roll on the inner surface of a fixed, hollow sphere the diameter of which is equal to the distance between the extreme points of the two wheels. The mass of the wheels is supposed to be negligible. The resulting motion of the body may be interpreted in the Euler space by the motion of a point along an admissible curve of a certain nonholonomic manifold V_3^2. Analytically, the problem is reducible to the integration of the so-called Riemann P-equation which can be solved in terms of hypergeometric functions.

10. Manifold of configurations and time

In the manifold of configurations and time it is more difficult to pick out a line element because many possibilities are available. For example, one may define ds by

$$ds = 2L\,dt$$

where L denotes the kinetic potential. For this line element the natural trajectories are geodesics by virtue of Hamilton's principle. However, since ds is not the square root of a homogeneous quadratic form in the components of the velocity vector, the corresponding geometry is not a Riemannian one. The geometry of this more general type of metric manifolds has been developed by Finsler, Synge, Dirac, Cartan, and more recently by Lichnerowicz [1] and Rund [1–3].

BIBLIOGRAPHY

Agostinelli, C. [1], Sul moto di un corpo rigido pesante asimmetrico col baricentro appartenente all'asse di uno dei piani ciclici dell'ellissoide d'inerzia, *Atti Seminario Mat. e Fis. Univ. Modena,* **3,** 248–260 (1949).

——— [2], Sul moto intorno a un punto fisso di un corpo rigido pesante il cui baricentro appartiene all'asse di uno dei piani ciclici dell'ellissoide d'inerzia, *Ann. Mat. Pura Appl.* (4), **30,** 211–224 (1949).

Alfieri, L. [1], Risoluzione di un problema comprendente quello di Staude, *Univ. Roma Ist. Naz. Alta Mat. Rend. Mat. e Appl.* (5), **12** (1953), 367–372 (1954).

Arrighi, G. [1], Sul giroscopio magnetico, *Atti Accad. Naz. Lincei Mem. Cl. Sci. Mat. Nat.* (8), **1,** 195–204 (1947).

Bödewadt, U. T. [1], Der symmetrische Kreisel bei zeitfester Drehkraft, *Math. Z.,* **55,** 310–320 (1952).

Bottema, O. [1], The stability of Staude's top motion, *Koninkl. Nederland. Akad. Wetenschap., Proc.,* **48,** 316–325 (1945).

Bottema, O., and Beth, H. J. E. [2], Euler's equations for the motion of a rigid body in n-dimensional space, *Koninkl. Nederland. Akad. Wetenschap., Proc.,* Ser. *A,* **54,** 106–108 (1951).

——— [3], The stationary motions of a rigid body under no forces in four-dimensional space, *Koninkl. Nederland. Akad. Wetenschap., Proc.,* Ser. *A,* **54,** 123–129 (1951).

Braunbeck, W. [1], Der symmetrische Kreisel mit zeitlich periodischem Richtmoment, *Z. Angew. Math. Mech.,* **33,** 174–188 (1953).

Chetaev, N. G. [1], On stability of rotation of a rigid body with a fixed point in Lagrange's case (Russian), *Prikl. Mat. Mekh.,* **18,** 123–124 (1954).

Chrapan, J. [1], The Lagrangian rigid body (Slovak.), *Mat.-Fyz. Sborn. Slovensk. Akad. Vied Umeni,* **2,** 23–51 (1952).

Colombo, G. [1], Osservazioni sulla stabilità dei moti merostatici di un giroscopio ed applicazioni ad un caso notevoli, *Rend. Sem. Mat. Univ. Padova,* **20,** 59–77 (1951).

Corliss, J. J. [1], On the unsymmetric top, *Acta Math.,* **59,** 423–441 (1932).

——— [2], On the unsymmetric top, *Acta Math.,* **62,** 301–312 (1934).

De Angeli, E. [1], Su alcuni casi integrabili di movimento di un giroscopio asimmetrico pesante, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (6), **25,** 699–703 (1937).

De Simoni, F. [1], Sul moto dei corpi rigidi con sospensione quasi-baricentrica, *Ann. Scuola Norm. Sup. Pisa Sci. Fis. Mat.* (2), **11,** 197–210 (1942).

Fabbri, M. R. [1], Sopra una soluzione particolare delle equazioni del moto di un solido pesante intorno ad un punto fisso, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (6) **19,** 407–415 (1934).

——— [2], Sopra un particolare movimento di un solido pesante intorno ad un punto fisso, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (6), **19,** 495–502 (1934).

——— [3], Sui coni di Poinsot in una particolare rotazione dei solidi pesanti, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (6), **19,** 872–873 (1934).

Field, P. [1], On the unsymmetrical top, *Acta Math.*, **56,** 355–362 (1931).

——— [2], On the unsymmetrical top, *Acta Math.*, **62,** 313–316 (1934).

Goryachev, D. N. [1], New cases of motion of a rigid body about a fixed point (Russian), *Varshav. Univ. Izvest.*, no. **3,** 1–11 (1915).

——— [2], New case of integrability of the Euler dynamical equations (Russian), *Rostov-Don. Varshav. Univ. Izvest.*, no. **3,** 1–13 (1916).

Grammel, R. [1], Die Stabilität der Staudeschen Kreiselbewegungen, *Math. Z.*, **6,** 124–142 (1920).

——— [2], Zur Berechnung der Poinsotbewegung, *Ingen.-Arch.*, **18,** 53–59 (1950).

——— [3], Die stationaren Bewegungen des selbsterregten Kreisels und ihre Stabilität, *Ingen.-Arch.*, **21,** 149–156 (1953).

——— [4], Der selbsterregte unsymmetrische Kreisel, *Ingen.-Arch.* **22,** 73–97 (1954).

Grioli, G. [1], Esistenza e determinazione delle precessioni regolari dinamicamente possibili per un solido pesante asimmetrico, *Ann. Mat. Pura Appl.* (4), **26,** 271–281 (1947).

——— [2], Questioni di stabilità riguardanti le precessioni regolari del solido pesante asimmetrico, *Ann. Scuola Norm. Sup. Pisa* (3), **1** (1947), 43–74 (1949).

——— [3], Moto attorno al baricentro di un giroscopio soggetto a forze di potenza nulla, *Univ. Roma Ist. Naz. Alta Mat. Rend. Mat. e Appl.* (5), **6,** 439–463 (1947).

Griseri B. [1], Soluzioni meromorfe delle equazioni del moto di un solido pesante intorno a un punto fisso col baricentro situato sull'asse di una delle sezioni circolari dell'ellissoide d'inerzia relativo al punto fisso, *Univ. e Polit. Torino Rend. Sem. Mat.*, **9,** 225–243 (1950).

Hamel, G. [1], Über den allgemeinen schweren Kreisel, *Z. Angew. Math. Mech.*, **25/27,** 159–160 (1947).

Koshlyakov, V. N. [1], On certain particular cases of integration of Euler's equations, connected with the motion of a gyroscope in a resisting medium (Russian), *Prikl. Mat. Mekh.*, **17,** 137–148 (1953).

Kuz'min, P. A. [1], Supplement to V. A. Steklov's case of motion of a heavy rigid body about a fixed point (Russian), *Prikl. Mat. Mekh.*, **16,** 243–245 (1952).

Lazzarino, O. [1], Sull'equivalenza fra le equazioni differenziali di Hess-Schiff e quelle di Euler-Poisson nella teoria dei giroscopi asimmetrici pesanti, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (5), $\mathbf{28_1}$, 325–331, 341–346 (1919).

——— [2], Sopra alcuni casi singolari nella teoria dei giroscopi asimmetrici pesanti, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (5), $\mathbf{28_2}$, 9–14, 259–263, 329–333 (1919).

——— [3], Sul moto dei giroscopi asimmetrici pesanti nel caso in cui l'invariante principale S è costantemente nullo, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (5), $\mathbf{28_2}$, 422–426, 489–493 (1919).

——— [4], Teoria sintetica dell'equivalenza fra i sistemi di equazioni differenziali di Euler-Poisson e di Hess-Schiff nella teoria dei giroscopi rigidi pesanti e studio

dei casi singolari per i quali l'equivalenza non sussiste I, II, *Ann. Scuola Norm. Sup. Pisa* (2), **7**, 91–107 (1938); **7**, 109–136 (1938).
Lichnerowicz, A. [1], Les espaces variationels généralisés, *Ann. Sci. École Norm. Sup.* (3), **62**, 339–384 (1945).
Manacorda, T. [1], Sul moto di un solido asimmetrico attorno ad un punto fisso, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (8), **6**, 711–714 (1949).
——— [2], Sul moto di un solido planare intorno ad un punto fisso, *Riv. Mat. Univ. Parma*, **1**, 383–391 (1950).
——— [3], Sul moto di un solido attorno ad un punto fisso, *Ricerca Sci.*, **20**, 487–490 (1950).
Manarini, M. [1], Sopra un teorema di Staude-van der Woude relativo al moto di un corpo pesante intorno ad un punto fisso, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (6), **14**, 572–577 (1931).
——— [2], Alcune notevoli proprietà del centro d'inerzia relative alle rotazioni permanenti nella dinamica del corpo rigido con un punto fisso, *Boll. Un. Mat. Ital.* (3), **3**, 214–219 (1948).
Mercalov, N. I. [1], The problem of motion of a rigid body, having a fixed point, with $A = B = 4C$ and the area integral $\neq 0$ (Russian), *Bull. Acad. Sci. U.R.S.S. Cl. Sci. Techn.* [*Izvest. Akad. Nauk S.S.S.R., Otdel. Tekhn. Nauk*], 697–701 (1946).
Mettler, E. [1], Periodische und asymptotische Bewegungen des unsymmetrischen schweren Kreisels, *Math. Z.*, **43**, 59–100 (1938).
Myasnikov, P. V. [1], On a new method of singling out integrable cases of motion from the general problem of rotation of a heavy rigid body about a fixed point; a new particular case of motion (Russian), *Vestnik Moskov. Univ. Ser. Fiz. Mat. Est. Nauk*, **9**, no. 3, 47–55 (1954).
Nadile, A. [1], Sul moto intorno a un punto fisso di un corpo rigido pesante il cui baricentro appartiene a uno dei piani principali di inerzia, *Matematiche Catania*, **5**, 68–82 (1950).
Pignedoli, A. [1], Sul moto di un corpo rigido pesante intorno ad un punto fisso, prossimo al baricentro o poco differente della struttura giroscopica, *Atti Soc. Nat. Mat. Modena*, **6** (76), 115–143 (1945).
Rumyancev, V. V. [1], On stability of rotation of a heavy rigid body with one fixed point in Kovalevskaya's case (Russian), *Prikl. Mat. Mekh.*, **18**, 457–458 (1954).
Rund, H. [1], Die Hamiltonsche Funktion bei allgemeinen dynamischen Systemen, *Arch. Math.*, **3**, 207–215 (1952).
——— [2], Application des méthodes de la géométrie métrique généralisée à la dynamique théorique, *Geom. Diff. Colloq. Int. Cent. Nat. Rech. Sci.*, Strassbourg, 41–51, (1953).
——— [3], Über nicht-holonome allgemeine metrische Geometrie, *Math. Nachr.*, **11**, 61–80 (1954).
Sbrana, F. [1], Un osservazione sul moto di un solido pesante attorno ad un punto, *Boll. Un. Mat. Ital.*, **13**, 269–271 (1934).
Sretenskiĭ, L. N. [1], Motion of the Goryachev-Chaplígin gyroscope (Russian), *Izvest. Akad. Nauk S.S.S.R., Otdel. Tekhn. Nauk*, no. 1, 109–119 (1953).
Stankevich, I. V. [1], Symmetric coordinates for the motion of a rigid body about a fixed point (Russian), *Trudy Moskov. Stanko-Instrum. Inst.*, **1**, 135–152 (1937).
——— [2], Rotation of a heavy rigid body about a fixed point. Application of the Jacobi method (Russian), *Trudy Moskov. Stanko-Instrum. Inst.*, **2**, 17–42 (1938).
——— [3], Application of the Jacobi method to the investigation of the motion of a heavy rigid body about a fixed point (Russian), *Trudy Moskov. Stanko-Instrum. Inst.*, **5**, 5–26 (1939).

Stankevich, I. V. [4], Rotation of a heavy rigid body about a fixed point. Application of the Jacobi method (Russian), *Trudy Moskov. Stanko-Instrum. Inst.*, **6**, 5–56 (1940).

Stoewa, P. [1], Die Stabilität der Staudeschen Kreiselbewegungen, *Jb. Phil. Fak.*, Leipzig 1922, II Halbjahr, pp. 131–132.

Tallqvist, H. [1], Über die Drehung eines starren Körpers um einen festen Punkt., *Acta Soc. Sci. Fennicae*, **50**, no. 14, 1–27 (1926).

Udeschini, P. [1], Giroscopio rotolante sopra un piano orizzontale in presenza di un polo magnetico, *Atti Accad. Sci. Torino, Cl. Sci. Fis. Mat. Nat.*, **77**, 384–398 (1942).

Vagner, V. V. [1], Geometry of the configuration space of a rigid body rotating about a fixed point (Russian), *Uchen. Zap. Saratov. Gos. Univ. im. Chernyshevskogo, Ser. Fiz. Mat. Inst.*, **1**, no. 2, 34–57 (1938).

——— [2], Geometric interpretation of the motion of non-holonomic dynamical systems, *Trudy Sem. Vektor. Tenz. Anal.*, **5**, 301–327 (1941).

Zeuli, T. [1], Sul moto di un corpo rigido pesante asimmetrico col baricentro appartenente all'asse di una delle sezioni circolari dell'ellissoide d'inerzia, *Univ. e Polit. Torino Rend. Sem. Mat.*, **9**, 263–295 (1950).

CHAPTER 2

MATHEMATICAL EXTERIOR BALLISTICS

INTEGRABILITY AND INTEGRATION OF THE DIFFERENTIAL EQUATIONS OF MOTION OF A PROJECTILE CONSIDERED AS A PARTICLE

1. Introduction

An analytic solution of the problem of determining the motion of a spinning projectile through the air has never been found. The aerodynamic forces acting on the projectile in its motion through the air depend upon the characteristics of the air, the velocity, size and shape of the projectile, and the orientation of the projectile with respect to the direction of motion. An exact physical law which governs the aerodynamic forces has not yet been found. Attempts have been made to describe the motion of the projectile by introducing various simplifying assumptions.

Consider the projectile as a material particle of mass m moving in still air. Suppose that it starts from the origin in a vertical xy-plane with initial velocity v_0, and is acted upon only by the force of gravity $m\mathbf{g}$ and the drag $m\mathbf{R}$ of the air in the direction opposite to the velocity $\mathbf{v}$ of the particle. If $\mathbf{r}$ denotes the radius vector which specifies the position of the particle at any instant t, then

$$\dot{\mathbf{v}} = \mathbf{R} + \mathbf{g}, \qquad \dot{\mathbf{r}} = \mathbf{v}$$

are the Newtonian equations of motion of the projectile.

In rectangular coordinates the equations of motion are

$$(1) \qquad \dot{x} = -\frac{R}{v}\dot{x}, \qquad \ddot{y} = -g - \frac{R}{v}\dot{y}$$

For a projectile symmetric about an axis and moving with its axis always in the direction of the tangent to the trajectory, equations (1) describe the motion accurately. These equations, with properly chosen drag function R, have been found most useful as first approximations to the trajectories of projectiles. They need corrections to account for various disturbances due to wind, abnormal density of the air, and other causes. Investigations of Prandtl, Lorenz and Darrieus on the form of the drag function have led to the expression

$$R = c\,\delta(y)v^2K(v\sqrt{T_0/T(y)}\,) \tag{a}$$

where $\delta(y)$ is the density of the air at the altitude y, $T(y)$ the absolute temperature of the air as a function of the altitude y, T_0 the so-called standard value of the absolute temperature, $K(v, y)$ the so-called Siacci function which is characteristic for the drag, and c the ballistic coefficient, dependent on the diameter of the projectile, form factor, standard density of the air on the ground, and the weight of the projectile.

When φ, the angle between the tangent to the trajectory and the horizontal, is introduced as the independent variable, the system (1) reduces to

$$\frac{d(v\cos\varphi)}{d\varphi} = \frac{c}{g}\,\delta(y)v^3K(v\sqrt{T_0/T(y)}\,) \tag{2}$$

$$\frac{dy}{d\varphi} = -\frac{v^2}{g}\tan\varphi$$

$$\frac{dx}{d\varphi} = -\frac{v^2}{g}$$

$$\frac{dt}{d\varphi} = -\frac{v}{g\cos\varphi}$$

The first two equations form a simultaneous system for v and y. If one can solve these two equations, all desired information about the trajectory can be obtained by quadratures. The solution in which we are interested corresponds to the initial conditions

$$\varphi = \varphi_0, \qquad t = 0, \qquad x = y = 0, \qquad v = v_0$$

φ_0 being the angle of projection of the projectile.

2. The work of Drach and Leimanis

If the drag R depends on the velocity v only, so that

$$R = cv^2K(v) = cf(v) \tag{b}$$

the system (2) reduces to

$$(3) \qquad \frac{d(v \cos \varphi)}{d\varphi} = \frac{c}{g} vf(v), \qquad \frac{dx}{d\varphi} = -\frac{v^2}{g}$$

$$\frac{dy}{d\varphi} = -\frac{v^2}{g} \tan \varphi, \qquad \frac{dt}{d\varphi} = -\frac{v}{g \cos \varphi}$$

The first equation of this system (denoted by (3_1)) is the so-called equation of the hodograph. If one can solve (3_1), the integration of the system (3) is reduced to quadratures. If the function $f(v)$ is arbitrary, the integration of (3_1) cannot even be reduced to quadratures. The problem of determining the forms of the function $f(v)$ for which the differential equation (3_1) is integrable by quadratures was completely solved by Drach [1] who used the elegant analytic theory of integration of differential equations developed by himself.

Only a few of the forms for $f(v)$ obtained by Drach can be considered as approximations to actual circumstances, the others have no relation to the actual drag functions and have only a theoretical interest.

Up to now we have not been in possession of a theory analogous to that of Drach for the general case of variable density and temperature of the air. In the case of constant temperature, i.e., under the assumption that the drag experienced by a given projectile depends only on its velocity and on the density of the air, the drag function has the form

$$(c) \qquad R = c\, \delta(y) v^2 K(v) = c\, \delta(y) f(v)$$

Even in this case the integration of the corresponding system of equations (2′), obtained from (2) by replacing the first equation by

$$\frac{d(v \cos \varphi)}{d\varphi} = \frac{c}{g} \delta(y) v^3 K(v) = \frac{c}{g} \delta(y) vf(v)$$

involves great mathematical difficulties. Two particular integrable cases were given by Legendre and Cavalli, namely

$$\delta(y) = 1/(1 + ky), \qquad f(v) = v^2 \qquad (k = \text{constant})$$

and $\quad \delta(y) = e^{-ky}, \qquad f(v) = v$

respectively.

Assuming that the density and the temperature of the air are variable quantities, Leimanis [1, 2] has shown that by application of infinitesimal transformations all the known integrable cases, and, in addition, new integrable cases of a more general character can be found. His method led to the following cases integrable by quadratures.

Case A. The density of the air is variable, but the temperature is constant, i.e., $\delta = \delta(y)$, $T(y) = T_0 = \text{const.}$ The drag function then has the form $R = c\,\delta(y)f(v)$, and the integrable cases are:

(i) $f(v) = v, \qquad \delta(y) = 1$ (see de Jong)

(ii) $f(v) = v, \qquad \delta(y) = e^{-ky}$ (see Cavalli)

(iii) $f(v) = v, \qquad \delta(y) = 1/(1 + ky)^{1/2}$

(iv) $f(v) = v, \qquad \delta(y) = 1 + ky$

(v) $f(v) = v, \qquad \delta(y) = 1/(1 + ky)^2$

where y may also be replaced by a linear expression in y,

(vi) $f(v) = v^2, \qquad \delta(y) = 1/(1 + ky)$ (see Legendre)

(vii) $f(v) = v^{-2}e^{-kv^2/2g}, \qquad \delta(y) = e^{-ky}$ (k = constant)

Case B. The density and the temperature of the air vary according to the laws

$$\delta(y) = \delta_0(1 - \lambda y/T_0)^{(1/\alpha\lambda)-1}, \qquad T(y) = T_0 - \lambda y$$

where $\lambda \sim 0.005$ and α is the so-called gas constant. These laws correspond to the actual conditions in nature. In this case the drag function is

$$R = c\,\delta_0(1 - \lambda y/T_0)^{1/\alpha\lambda}f(w)$$

where $w = v\sqrt{T_0/T(y)}$. The integrable cases are:

(viii) $f(w) = w$

(ix) $f(w) = (1/w^2)(1 - \lambda w^2/2gT_0)^{(1/\alpha\lambda)+1}$

(x) $f(w) = w(1 - \lambda w^2/2gT_0)^{3/2}$

3. The work of Popoff

Take the mouth of a cannon as the origin O of a coordinate system the x-axis of which coincides with the direction and sense of the initial velocity $\mathbf{v}_0$ at the origin O and the y-axis of which is directed downward. Decomposing the resistance function $cF(v)$ in the direction of the velocity $\mathbf{v}$ of the projectile and in the directions of the x- and y-axes, the three sides of the force triangle are proportional to v, x' and y' respectively. Therefore the equations of motion of the center of mass of the projectile become

$$\frac{dx}{dt} = x', \quad \frac{dx'}{dt} = -x'\frac{cF(v)}{v} = -cx'f(v) \tag{4}$$

$$\frac{dy}{dt} = y', \quad \frac{dy'}{dt} = g - y'\frac{cF(v)}{v} = g - cy'f(v)$$

where

$$f(v) = \frac{F(v)}{v}$$

and the velocity v is given by the formula

$$v^2 = x'^2 + y'^2 - 2x'y' \sin \varphi_0$$

φ_0 being the angle made by the initial velocity $\mathbf{v}_0$ with the horizontal line through O. The initial conditions of motion are:

$$t = 0, \quad x(0) = y(0) = 0, \quad x'(0) = v_0, \quad y'(0) = 0$$

The system (4) has been studied extensively by Popoff [1–3], using a classical theorem of Poincaré on the analyticity of the solution (corresponding to the given initial conditions) with respect to a parameter, the right-hand sides of the given equations being analytic functions of the parameter. In the system (4) such a parameter is $\sin \varphi_0$ and the series solution obtained in terms of this parameter converges very rapidly.

Popoff assumes that the function $F(v)$ satisfies the following conditions:

1. $F(0) = 0$.
2. If v is real and positive, then also $F(v)$ is real and positive.
3. $F'(v) \geq 0$ for real and positive v.
4. $F(v)$, considered as a function of a complex variable v, is an integral

function (i.e., it is transcendental or algebraic, in the last case being a polynomial) for which

$$\lim_{n \to \infty} \left[\frac{F^{(n)}(0)}{n!} \right]^{1/n} = 0$$

This last assumption is, evidently, rather restrictive.

4. The work of Biggeri

Recently Biggeri [1] has criticized the work of Popoff, claiming that it suffers from two shortcomings of fundamental importance. First, he points out that Popoff has not shown that his differential equations (4) satisfy the Cauchy-Lipschitz conditions at every point of the trajectory corresponding to the given initial conditions. That is to say, he has not shown that the integrals of the system (4) are limited in every finite interval on the real and positive semiaxis of t. In other words, Popoff has not shown that the moving singularities of the system (4) which are at a finite distance from O do not belong to the real and positive semiaxis of t. He should have at least referred to the fundamental property that the inferior limit of the distances of the moving singularities of the system (4) from the real and positive semiaxis of t is always greater than zero, equality being excluded. A second shortcoming of Popoff, according to Biggeri, lies in the fact that analytical continuation should be carried out, since the above-mentioned theorem of Poincaré is of a local character only. Therefore, he concludes that the Popoff solution of the principal problem of exterior ballistics is far from being a global solution.

Furthermore, Biggeri points out that the various forms of the differential equations used in exterior ballistics are inadequate for the following reasons:

1. the resistance function does not depend only upon the velocity v of the projectile but also upon certain other variables such as, for example, the altitude y. In addition it may depend on several parameters. Therefore, there does not exist a resistance function of the air but only a resistance function of the air and the projectile.

2. The ballistic coefficient c is not a constant, but it depends upon the velocity v of the projectile and the ordinate y, and, in addition, upon certain parameters coupled with the projectile.

3. Actually, in the equations of the hodograph we are not allowed to separate the ballistic coefficient c from the function $F(v)$ in the form $cF(v)$. It is even too restrictive to assume that the resistance function of the air and the projectile contains the ballistic coefficient as a factor $c(y)$ which depends only upon the altitude y.

A system which approximates to a great extent the actual circumstances in nature is the following one (thermodynamical statement of Göttingen):

$$(5) \qquad \frac{d(v\cos\varphi)}{d\varphi} = -\frac{S - R(v, y)}{mg} v$$

$$\frac{dy}{d\varphi} = -\frac{v^2}{g}\tan\varphi$$

$$\frac{dx}{d\varphi} = -\frac{v^2}{g}$$

$$\frac{dt}{d\varphi} = -\frac{v}{g\cos\varphi}$$

where R = resistance function of the air (kg)

$$= c_w \frac{v}{(KgR')^{1/2}(T_0 - \theta y/1000)^{1/2}} \frac{\rho_0}{2}\left(1 - \frac{\theta y}{273 + t_0}\right)^{5.83} v^2 A$$

c_w = resistance coefficient of the air, which is a function of the Mach number $M = v/c$

c = velocity of sound, meter/sec

θ = gradient of temperature with respect to the ordinate, 6.5°/1000 meter

T_0 = absolute temperature on the Earth

R' = air constant $\sim$ 29.27 kg-m

$K = \dfrac{c_p}{c_v} = \dfrac{\text{specific heat at constant pressure}}{\text{specific heat at constant volume}}$

ρ_0 = air density on the Earth

t_0 = air temperature on the Earth, °C

A = surface area of the projectile to which refers the resistance coefficient c_w, meter^2

m = mass of the projectile, $\dfrac{\text{kg-sec}^2}{\text{meter}}$

S = force (thrust), kg

and the other variables v, x, y, t, φ have the same meaning as before.

Fundamental theorem 1. Assume that a system of equations (G) (differential, integral, or integro-differential) of exterior ballistics be given in a form which satisfies the following conditions:

(a) The resistance function G of the air and the projectile (which depends not only on v but also on a finite number of variables and parameters) is lipschitzian with respect to its variables in certain neighborhoods of the initial values of these variables.

(*b*) The resistance function G vanishes for $v = 0$.

(*c*) The resistance function G increases as v increases through real and positive values.

Further, let ϵ be an arbitrary positive number, no matter how small, and T a given real positive number, no matter how large.

The problem is to determine a function F, which will be called a resistance function of the air and the projectile, containing the same number of variables and parameters as G and satisfying the following conditions:

(*d*) The function F is lipschitzian with respect to its variables in certain neighborhoods of the initial values of these variables. The initial values are supposed to be the same as for the system of equations (G).

(*e*) The function F vanishes for $v = 0$.

(*f*) The function F is an increasing function of v for real and positive values of v.

(*g*) Given an arbitrary finite interval of the semiaxis $0 \leq v < +\infty$, there exists a region in the plane of the complex variable v, containing this interval, such that in its interior F is an analytic function of v, assuming that F depends only upon v. If, in addition, F depends upon several other variables besides v, the generalization of the property (*g*) is obvious.

Denote by (F) the system of equations obtained from (G) by replacing the function G by the function F. Then the moduli of the differences of the solutions of the two systems (G) and (F), corresponding to the same initial conditions, remain less than ϵ for every value of t in the interval $0 \leq t \leq T$.

Biggeri calls this theorem a fundamental theorem of global approximation to the solution of the principal problem of exterior ballistics.

Given a system of equations (G) with resistance function G, satisfying the conditions (*a*), (*b*), and (*c*), its transformation into a system (F) with F as the resistance function is called its holomorphization.

According to Biggeri's opinion the greatest difficulty in solving the principal problem of exterior ballistics consists in its holomorphization. The system (5), for instance, can be holomorphized.

Without any restriction of generality, assume that the resistance function depends only upon the velocity v of the projectile and its altitude y above the horizon. In addition, assume that the weight of the projectile is a function of y. Choosing the positive direction of the

y-axis upward, consider a system of the equations of motion of the projectile in the form:

(6)
$$X' = \frac{dX}{dt}, \qquad Y' = \frac{dY}{dt}$$
$$\frac{dY'}{dt} = -g - Y'\frac{F(v, Y)}{v} = -g - Y'f(v, Y)$$
$$\frac{dX'}{dt} = -X'\frac{F(v, Y)}{v} = -X'f(v, Y)$$

the initial conditions of motion being:

$$X(0) = Y(0) = 0, \qquad v(0) = v_0$$
$$X'(0) = v_0 \cos \varphi_0, \qquad Y'(0) = v_0 \sin \varphi_0$$

Introduce instead of X and Y new variables x, and y by means of the relations

(7)
$$X = x \cos \varphi_0, \qquad Y = x \sin \varphi_0 - y$$

Then system (6) goes over into the following one:

(8)
$$x' = \frac{dx}{dt}, \qquad y' = \frac{dy}{dt}$$
$$\frac{dx'}{dt} = -x'f[v, 2k^2x - (x + y)] = x''$$
$$\frac{dy'}{dt} = g - y'f[v, 2k^2x - (x + y)] = y''$$

where

$$v^2 = (x' + y')^2 - 4k^2x'y', \qquad k = \sin\left(\frac{\pi}{4} + \frac{\varphi_0}{2}\right)$$

and the initial conditions of motion are:

$$x(0) = y(0) = 0$$
$$x'(0) = v_0, \qquad y'(0) = 0$$

Further Biggeri assumes that the acceleration g due to gravity is a function of the altitude Y, i.e.,

$$g = g(Y) = g[2k^2x - (x + y)]$$

Biggeri solves the principal problem of exterior ballistics, stated in the form (8), as follows. Assume that λ is a complex parameter and consider the system of equations:

$$x' = \frac{dx}{dt}, \qquad y' = \frac{dy}{dt} \tag{9}$$

$$x'' = \frac{dx'}{dt} = -x'f(v, \xi)$$

$$y'' = \frac{dy'}{dt} = g(\xi) - y'f(v, \xi)$$

where for every λ such that $|\lambda| \leq 1$, we have

$$v^2 = (x' + y')^2 - 4k^2\lambda x'y'$$

$$\xi = \xi(\lambda) = 2k^2\lambda x - \lambda(x + y)$$

The initial conditions of motion are:

$$x(0) = y(0) = 0$$

$$x'(0) = v_0, \qquad y'(0) = 0$$

Fundamental theorem 2. The solution

$$x = x(t, \lambda)$$

$$y = y(t, \lambda)$$

of the system (9) is analytic with respect to the parameter λ for every real and positive value of t, i.e., the following expansions in terms of the parameter λ hold:

$$x = x(t, \lambda) = x_0(t) + \lambda x_1(t) + \lambda^2 x_2(t) + \cdots = \sum_0^\infty \lambda^n x_n(t) \tag{10}$$

$$y = y(t, \lambda) = y_0(t) + \lambda y_1(t) + \lambda^2 y_2(t) + \cdots = \sum_0^\infty \lambda^n y_n(t)$$

In addition,

$$v = v(t, \lambda) = v_0(t) + \lambda v_1(t) + \lambda^2 v_2(t) + \cdots = \sum_0^\infty \lambda^n v_n(t) \tag{11}$$

where

$$x_n = x_n(t)$$
$$y_n = y_n(t)$$
$$v_n = v_n(t)$$

are analytic functions of t for every real and positive value of t. The radii of convergence of the power series (10) and (11) with respect to λ are greater than one for every real and positive t. The series themselves are uniformly convergent in any finite interval on the real and positive semiaxis, $0 \leq t < \infty$, assuming that $|\lambda| \leq 1$.

Substituting (10) and (11) into the system (9) and identifying the coefficients of equal powers of λ on both sides of the equations so obtained, an infinite number of systems for $x_0(t)$, $y_0(t)$, $v_0(t)$; $x_1(t)$, $y_1(t)$, $v_1(t)$; $\cdots$ is obtained, all being integrable by successive quadratures.

Substantially, Biggeri's work is a contribution to the analytic theory of ordinary differential equations, integral equations, and integro-differential equations. Some of his proofs, however, are based on results obtained in some of his previous papers and have not been available to the present writer.

5. Reduction to integral equations

Bucerius [1] gives a new approach to the particle problem of exterior ballistics by treating it as a boundary-value problem with given range and time of flight, instead of as an initial-value problem. This method, however, does not seem to be of great practical value because of the laborious calculations involved and the fact that the initial values of the solutions can be determined only after the calculations have been completed.

6. Numerical integration. Method of Kazakov and Palechek

At present there are several methods for numerical integration of the various systems of differential equations of the particle problem corresponding to particular choices of the independent variable. Among the more recent ones an interesting method is that of Kazakov [1]. The importance of his method lies in the fact that in carrying out the numerical integration the extrapolations are not directly performed upon one or the other variables used, but rather upon certain auxiliary quantities which have little influence on the results of extrapolation. Kazakov himself used as the independent variable the abscissa of the center of mass and the time of flight of the projectile.

Palechek [1, 2] has extended Kazakov's method to the case where one of the following is taken as the independent variable, either the tangent of the angle of inclination of the tangent to the trajectory, or the horizontal component of the velocity, or the ordinate of the center of mass, or the vertical component of the velocity of the projectile. Palechek also gives estimates of the relative error of the results obtained for each choice of the independent variable, and discusses the appropriateness of the choice in each case.

7. Hodograph of the resistance vector. Method of Bouffard

Whereas usually the ballisticians are concerned with the hodograph of the velocity vector, Bouffard [1] considers the more involved hodograph representation of the resistance vector **R**, plotted against the angle φ in a polar diagram with the origin at a fixed point O. The ratio $p = g/\overline{OP}$ of the acceleration g due to gravity to the intercept $\overline{OP}$ on the vertical axis of the tangent to this resistance hodograph at a point A is called the ballistic index of the trajectory at the point M corresponding to the point A. In terms of the equation of the trajectory $y = y(x)$, this number p is given by the expression

$$p = 4 - 2y''y^{iv}/y'''^2$$

where y'', y''' and y^{iv} denote the second, the third, and the fourth derivatives respectively of y with respect to x.

If the given trajectory is represented by an arc of a parabola with oblique axis, or by a hyperbola with a vertical asymptote, or by an arc of a cubic parabola (trajectory of Piton-Bressant), the index p is a constant, i.e., in all three cases mentioned above the polar diagram of **R** reduces to a straight line.

Next Bouffard studies all motions M_p with constant ballistic index p; or, in other words, all motions for which the polar diagram of **R** is a straight line. He then applies the results obtained for calculating trajectories, using surosculating arcs of trajectories. His idea is to replace the actual motion of a projectile by a fictive osculating motion M_p for which the straight-line **R**-diagram is tangent to the polar diagram of **R** in the actual motion. The surosculatory motion mentioned above is one which fits the actual motion at a point $M_0(\varphi_0, v_0, R_0, p_0)$ not only in the first three elements φ_0, v_0, R_0 (osculating motion) but also in the ballistic index p_0, calculated for the actual motion. This method has a high degree of accuracy since at M_0 both trajectories have a contact of order four.

The above idea is applied to the numerical calculation of trajectories, with the use of modified surosculatory trajectories and higher order approximations. The advantages of the method against other methods are illustrated by two examples. The number of arcs necessary for the calculation of a trajectory is approximately equal to one half of that necessary in other methods. Also, the calculation of a single arc proceeds rather rapidly.

THE DIFFERENTIAL EQUATIONS OF MOTION OF A PROJECTILE CONSIDERED AS A RIGID BODY

8. The work of Signorini

Signorini [1] has dealt with the problem of motion of a projectile in a form which may be called complete in the sense that no aerodynamical effects due to the translatory and rotatory motion of the projectile which are of importance in practice have been neglected.

The effects due to the translatory motion of the projectile P are equivalent to a single force $m\mathbf{R}$ applied at the center of pressure which lies on the figure axis of P. Those due to the rotational motion of P in comparison with $m\mathbf{R}$ are much smaller and usually are neglected. However, the force which cannot be neglected if we want to be more precise is the Magnus effect which is equivalent to a single force $\mathbf{F}^{(M)}$ applied at the center of mass G of the projectile. In addition to these aerodynamic forces, Signorini takes into account the weight of the projectile $\mathbf{W} = m\mathbf{g}^{(G)}$, applied at G, and the composite centrifugal force $\mathbf{F}^{(c)}$ due to the rotation of the Earth.

The theorem on the motion of the center of mass G of the projectile P then reads

$$\dot{\mathbf{v}} = \mathbf{g}^{(G)} + \mathbf{R} + \mathbf{F}^{(M)} + \mathbf{F}^{(c)}$$

Without introducing any simplifying assumptions Signorini first establishes more general differential equations of motion of a rigid body. These new differential equations actually correspond to a new form of the theorem of angular momentum

$$C\dot{r}\mathbf{k} + Cr\dot{\mathbf{k}} + A\mathbf{k} \times \ddot{\mathbf{k}} = \mathbf{M}$$

when applied to a rigid body with a gyroscopic structure. In particular, they are considered suitable for the solution by perturbation methods of the complete ballistic problem, denoting by this the simultaneous study of the motion of the center of mass G of the projectile and its rotation about G. Actually, however, such methods have not been developed, either by Signorini or by anybody else.

Second, Signorini has established the validity of certain simple approximations in the theory of motion of a gyroscope, called the principle of gyroscopic effect. This principle consists in the following:

Consider a gyroscope Q with a point O on its axis, fixed or coinciding with the center of mass G. Suppose that Q is subject to (i) a system of external forces with angular momentum $\mathbf{M} = \mathbf{M}^*$ at O, orthogonal to $\mathbf{k}$, and (ii) a rotation with a large initial angular velocity $\boldsymbol{\omega}_0 = r_0\mathbf{k}$ about its axis. The principle of gyroscopic effect then consists in replacing the equation, given by the theorem of angular momentum by the approximate equation

$$Cr_0 \frac{d\mathbf{k}}{dt} = \mathbf{M}^*$$

The theoretical justification of this principle has been given by Signorini for the case where the angular momentum $\mathbf{M}^*$ initially is zero. This case is of importance in ballistics where external forces of this type occur. Stoppelli [1-2] has investigated the extent to which similar approximations can be used in more general problems concerning a rigid body with one fixed point under the influence of external forces.

In the case of ordinary artillery projectiles we neglect the Magnus effect and the rotation of the Earth and assume that $g^{(G)} = \text{const} = g$, and we are faced with cases of regular firing of the projectile. In addition, because of a large initial spin of the projectile, the principle of gyroscopic effect applies. On the basis of these assumptions Signorini shows that the determination of motion of spinning artillery projectiles reduces to the solution of a system of four first-order differential equations and, in addition, certain quadratures. Two of these equations are analogous to the two simultaneous equations in the classical particle problem. In the case of constant density of the air, they reduce to a single differential equation of the type of the hodograph equation. The other two equations are of the type of equations arising in the second ballistic problem, namely, the study of motion of the projectile about its center of mass G, assuming that the motion of G has been deduced previously from the classical equations of motion.

The reason for the reduction of the order of the system of differential equations of motion of artillery projectiles to four lies in the fact that the determination of a space curve from its intrinsic equations requires the integration of an equation of the Riccati type the general solution of which is obtained from a particular solution by two quadratures.

BIBLIOGRAPHY

Biggeri, C. [1], Nuevos métodos generales en la balística racional, *Univ. Nac. La Plata Publ. Fac. Ci. Fisicomat* (2), **4,** no. 4, 314–351 (1951).

Bouffard, J. [1], Nouveau point de vue sur la balistique extérieure, *Mém. Artillerie Française*, **22,** 783–835 (1948).

Bucerius, H. [1], Neubegründung der äusseren Ballistik, *Astr. Nachr.*, **270,** 66–73 (1940).

Drach, J. [1], L'équation différentielle de la balistique extérieure et son intégration par quadratures, *Ann. Sci. École Norm. Sup.* (3), **37,** 1–94 (1920).

Kazakov, S. A. [1], Computation of the trajectories of the centers of gravity of artillery shells (Russian), *Prikl. Mat. Mekh.*, **9,** 129–138 (1945).

Leimanis, E. [1], Zur Integration der Differentialgleichungen der äusseren Ballistik durch Quadraturen, *Contrib. Baltic Univ.*, no. 5, 1–32, Hamburg (1946).

——— [2], The application of infinitesimal transformations to the integration of differential equations of exterior ballistics by quadratures, *Proc. Second Canad. Math. Congr.*, 206–217, Vancouver (1949), University of Toronto Press (1951).

Palechek, E. M. [1], On the computation of the trajectories of the centers of gravity of artillery shells (Russian), *Prikl. Mat. Mekh.*, **15,** 83–96 (1951).

——— [2], Approximate integration of the equations of exterior ballistics by the method of S. A. Kazakov for vertical parameters (Russian), *Prikl. Mat. Mekh.*, **16,** 505–510 (1952).

Popoff, K. [1], *Les méthodes d'intégration de Poincaré et le problème général de la balistique extérieure*, Gauthier-Villars, Paris (1925).

——— [2], *Das Hauptproblem der äusseren Ballistik im Lichte der modernen Mathematik*, Akademische Verlagsgesellschaft, Leipzig (1932).

——— [3], Le mouvement d'un projectile autour de son centre de gravité, *Mém. Sci. Math.*, no. 117, Gauthier-Villars, Paris (1951).

Signorini, A. [1], Complementi alla dinamica dei giroscopi e equazioni del problema completo della balistica esterna, *Atti Accad. Naz. Lincei Mem. Cl. Sci. Fis. Mat Nat.* (8), **1** (Sect. 1), no. 1, 1–41 (1946).

Stoppelli, F. [1], Sul principio dell'effetto giroscopico, *G. Mat. Battaglini* (4), **4** (80), 14–38 (1951).

——— [2], Sui fenomeni giroscopici in un solido qualsiasi, *Rend. Sem. Mat. Univ. Padova*, **21,** 25–43 (1952).

CELESTIAL MECHANICS

SECTION 2

UNIFORMIZATION OF THE THREE-BODY PROBLEM

PERIODIC SOLUTIONS OF THE RESTRICTED THREE-BODY PROBLEM

1. Periodic solutions in the Poincare restricted problem

In the circular restricted three-body problem in space, Batrakov [1] has recently shown the existence of periodic solutions, analogous to the Schwarzschild periodic solutions in the planar case. Whereas in the Schwarzschild case the apse line of the osculating Kepler ellipse, representing these solutions, rotates through a certain angle during the period of motion $T = T_0 + \Delta T$ (T_0 being the period of the generating solution and ΔT being of the order of a small perturbing mass μ), in the case of the newly obtained periodic solutions the line of nodes is subject to a rotation. The eccentricity e and the inclination I of the plane of the generating solution are connected by a relation which has been elaborated in detail for the commensurability 3:1 of the mean motions of the asteroid and Jupiter. Terms up to the fourth degree inclusive in e and I in the expansion of the perturbation function are taken into account. Inspection of a table not reproduced here which represents the above-mentioned relation between e and I shows that the generating solution deviates too much from the actual orbits of asteroids with commensurability 3:1. Hence, the new solutions obtained seem to have little value as intermediate orbits for the commensurability case considered.

For other commensurabilities $p:q$ of even order ($p - q$ = even number, where p and q are relatively prime integers) and for commensurabilities of order one the question remains open.

The Batrakov solutions as well as the Schwarzschild solutions contain the eccentricity of the orbit as an arbitrary parameter. The Poincaré solutions of the second type (*sorte*) are a particular case of the Schwarzschild solutions (when $\Delta T = 0$), whereas the Poincaré solutions of the third type are a particular case of the newly Batrakov solutions. Both the Batrakov and the Schwarzschild periodic solutions are sinodic; i.e., after the elapse of the period of motion only the relative positions of all three bodies are the same. On the other hand, the absolute positions of the bodies repeat in the Poincaré periodic solutions. The Batrakov periodic solutions with circular generating solutions, lying in the plane of motion of Jupiter, exist only for commensurabilities of odd orders: 3, 5, 7, $\cdots$. In practice, however, these cases are of little interest.

Since the equations of celestial mechanics do not contain the time t explicitly and they assume the time-independent energy integral, the Poincaré periodic solutions of the three-body problem contain only two arbitrary constants out of twelve constants which determine the motion. Because of the small number of arbitrary constants which are at our disposal in these solutions (and also in the Schwarzschild and the Batrakov solutions), their applicability to astronomy is very limited.

Merman [1] has constructed a new class of periodic solutions for two particular cases of the three-body problem: (i) the restricted problem, and (ii) its limit case, the Hill problem. These new periodic solutions contain four arbitrary constants out of six constants which determine the motion; namely, the initial values of the major semiaxis, the eccentricity, the inclination of the plane of the orbit, and the position of the asteroid in the orbit.

Merman's leading idea can be explained as follows. Suppose that by retaining in the equations of motion certain of the most influential terms which depend on the perturbing mass μ we have obtained a nonlinear but integrable system of equations. Take such a system of equations, instead of the linear system obtained by putting $\mu = 0$ (as in the Poincaré method), as a generating system. Whereas the period of a periodic solution of a linear system is completely determined by its coefficients, that of a periodic solution of a nonlinear system is itself an arbitrary constant. Let the general solution of the chosen generating system be a sum of a certain number of periodic functions with arbitrary independent periods. Then every set of commensurable periods gives a periodic solution of the generating system. The corresponding solution of the originally given system of equations will contain, as in the case of the Poincaré solutions, two arbitrary constants. In addition, however, there will be as many arbitrary constants at our disposal as there are independent periods in the general solution of the nonlinear generating system. Such a nonlinear system of equations, called the averaged system of equations, has been constructed by Merman for the two special cases mentioned above of the three-body problem.

2. Numerical integration. Periodic solutions in the Copenhagen restricted problem

Let two particles P_1 and P_2 of masses $m_1 = m_2 = \frac{1}{2}$ (the total mass being the unit of mass) revolve in a circular orbit about their common center of mass with constant angular velocity $n = 1$. Consider a third infinitesimal particle P which moves in the plane of motion of P_1 and P_2 in such a way that, although it is subject to the Newtonian attractions of P_1 and P_2, it does not disturb the Keplerian motion of the two finite bodies. The resulting model is called the Copenhagen (Thiele-Strömgren) form of the restricted problem of three bodies. Whereas in the "*problème restreint*," in the sense of Poincaré, one of the two finite masses is small in comparison with the other mass, in the Copenhagen form of the restricted problem both finite masses are of the same order of magnitude. In other words, the Poincaré model is adapted for investigation of perturbation problems in the solar system (where the mass of the Sun is much greater than that of any of the planets) whereas that of Strömgren is as far remote from a perturbation problem as possible.

Problems of the latter type were first attacked with the use of numerical integration by Thiele and Burrau, and were continued by Strömgren [1], using equal masses $m_1 = m_2$, and by Darwin with the mass ratio $m_1 = 10m_2$. During three decades (1895–1935) several hundred orbits of P were calculated numerically at the Copenhagen Observatory with the help of a large staff of computers. Since 1913 this work has been carried on systematically under the direction of Strömgren. Among these orbits are 63 simply-periodic and 13 periodic-asymptotic orbits. The orbits are given with respect to the rotating coordinate system ξ, η, which rotates uniformly with angular velocity $n = 1$ in the positive sense in such a way that the ξ-axis always coincides with the line through P_1 and P_2. The original aim of Strömgren was to obtain a complete survey of possible forms of simple-periodic and periodic-asymptotic orbits in the restricted problem with the mass ratio $m_1 = m_2$. This aim can now be considered as achieved.

In order to indicate Strömgren's working principles, we consider a particular orbit with one of the initial values fixed, and vary continuously the Jacobian constant K. In this way we obtain a group of related orbits. Those among them which are closed and which intersect one or both coordinate axes orthogonally are periodic orbits. If we vary K monotonically and follow the development of such periodic orbits, we arrive at the periodic orbit of ejection. At first Burrau and Strömgren both regarded the orbit of ejection as a limit of the group under consideration. Later on, however, when groups of orbits emerged which develop into simple orbits after passing through orbits of ejection and looped orbits, Strömgren discarded this interpretation. Finally Strömgren arrived at his dynamical termination principle for periodic groups of orbits in the restricted problem. This principle states that one of the following situations must occur:

1. The group of periodic orbits (solutions) is closed in itself, i.e., as the Jacobian constant K continues to vary, the same periodic orbits reappear.
2. The group has its natural origin and its natural end in one or two of the five libration points or in one or both of the two finite masses.
3. The dimensions or the period or the constant K of the orbits of the group become infinite. A mathematical proof of this principle was given by Wintner [1].

In a recent paper Vernić [1] showed that with reference to an inertial coordinate system, the orbits of the Copenhagen restricted problem are not, in general, closed curves but resemble an ellipse with a progressive motion of the apsides caused by the incommensurability of the angular velocity of the finite particles and the mean motion of the infinitesimal particle. As a consequence the dynamical model under consideration will not have periodic motions, but only Poisson stability in the sense of Poincaré. In other words, the periodicity of motion is not a kinematic property; it is implied only by a particular choice of the rotating coordinate system. Vernić also showed that the only motions which are periodic with respect to the rotating and the absolute coordinate systems are the linear oscillations of the infinitesimal particle P.

Since the transformation from the rotating ξ, η-coordinate system to an inertial coordinate system does not in general preserve the main properties of the orbits such as periodicity, cusps, and loops, Vernić concludes that the knowledge of the totality of orbits in the Copenhagen restricted problem cannot constitute a key to the understanding of the general three-body problem, as Strömgren had hoped it would. This criticism, however, is not meant to underestimate the methods used and the results obtained by Strömgren and his collaborators but only to deny their transferability to the general three-body problem.

Strömgren regarded the method of numerical integration of the differential equations of the three-body problem as the only possible way of tackling this most difficult problem of celestial mechanics. This opinion was also expressed by Sundman at the meeting of the *Astronomische Gesellschaft* in Budapest in 1930. Vernić [3, 4], however, does not share this view, and proposes a method of regularization and uniformization of collisions (in the sense of the theory of functions of a complex variable) as an appropriate way of solving this problem. Unfortunately, in view of some rather serious mistakes and miscalculations in his proofs, Vernić's attempt must be considered as a failure.

REGULARIZATION AND UNIFORMIZATION OF THE GENERAL THREE-BODY PROBLEM

3. Regularization of the two- and the three-body problem

Sundman's method. Vernić [3] claims to have extended the validity of Sundman's results on regularization to all types of collisions. In this connection he has formulated the following theorem.

Extended Sundman's theorem. All types of collisions (binary or triple, real or imaginary) in the two- and the three-body problem are regularizable and the problems themselves integrable by the transformations and the method used by Sundman. The most natural transformation of this type is that of Sundman and Levi-Civita,

$$C\,dt = du/V \qquad (C \text{ constant}) \tag{1}$$

whereas the most general transformation of this type is

$$C\,dt = S(r)\,du \tag{2}$$

Here V is the potential of the system, and $S(r) = S(r_1, r_2, r_3)$ is a symmetric homogeneous function of degree one in the three mutual distances. This function can be given explicitly in infinitely many ways for every case under consideration. The solution of the two- or three-body problem is uniformized by means of the parameter u:

$$r_i = r_i(u), \qquad t = T(u) \qquad (i = 1 \text{ or } i = 1, 2, 3)$$

i.e., to each value of u between $-\infty$ and $+\infty$ there is just one set of corresponding mutual distances.

Actually the theorem itself is false. Merman [2] showed that Vernić's proof is defective and that, although his theorem holds for all binary collisions and for real triple collisions in the case of the Lagrangian motions and in the case of other motions for certain values of the masses m_1, m_2, and m_3, it does not hold for all real triple collisions nor for any imaginary triple collision.

Vernić uses first the transformation

$$dt = R\,du \qquad \left(R^2 = \sum_{i=1}^{3} \frac{r_i^2}{m_i}\right) \tag{2'}$$

which is a particular case of a transformation of the type (2) and the limit relations obtained by Sundman for a real triple collision. He claims to have shown that R and all the coordinates and velocities of the particles can be expanded in series of positive integral powers of u or $t^{1/3}$. In other words, he maintains that the transformation (2′) regularizes all real triple collisions. This, however, is not true in general. From the limit relation of Sundman

$$\sqrt{R}\,\dot{R} = -A + \epsilon$$

it follows that

$$R = (c_2 + \epsilon)t^{2/3}$$

where

$$c_2 = \sqrt[3]{\tfrac{9}{4}A}$$

and ϵ denotes any quantity which approaches zero simultaneously with R as the particles approach a triple collision at $t = 0$. In the case of the Lagrangian equilateral triangle solutions, for instance, we have

$$A = 2M\left(\sum_{i=1}^{3}\frac{1}{m_i}\right)^{1/2} \qquad \text{and} \qquad M = \sum_{i=1}^{3} m_i$$

In order to show the validity of expanding the coordinates in powers of u, the transformed differential equations with u as the independent variable must be examined. These equations are

$$\frac{d^2\mathbf{r}_\upsilon}{du^2} = \frac{1}{R^2}(\sqrt{R}\,\dot{R})(\sqrt{R}\,\dot{\mathbf{r}}_\upsilon) - M\left(\frac{R}{r_\upsilon}\right)^2\frac{\mathbf{r}_\upsilon}{r_\upsilon} + m_\upsilon\sum_{i=1}^{3}\left(\frac{R}{r_i}\right)^2\frac{\mathbf{r}_i}{r_i} \qquad (\upsilon = 1, 2, 3)$$

In Merman's paper cited above, the factor $1/R^2$ is lacking in the first term on the right-hand side of this equation. From the limit relations of Sundman, it can be deduced that the coefficients $\sqrt{R}\,\dot{R}$ and R/r_υ ($\upsilon = 1, 2, 3$) approach definite limit values as t approaches zero, the instant of a real triple collision. R^2, on the other hand, approaches the limit value zero. Hence, even if the vectors $\sqrt{R}\,\dot{\mathbf{r}}_\upsilon$ and $\mathbf{r}_\upsilon/r_\upsilon$ had definite limit values, $\mathbf{r}_\upsilon(u)$, regular at $u = 0$, could not exist, by the Cauchy-Picard existence theorem.

In the case of the Lagrangian motions, however, it is known *a priori* that the above mentioned vectors approach definite limit values as t approaches zero. Furthermore, in this case, the three-body problem degenerates at the instant of a triple collision into two two-body problems, and the triple collision reduces to binary collisions for which $R^2 > 0$. In the general case, however, the above mentioned property of the vectors cannot be proved since, in a neighborhood of an instant of a triple collision, series expansions with irrational exponents arise (Siegel [1] and Wintner [2]).

Specifically, Siegel showed that the coordinates of the colliding particles, in general, have expansions of the form

$$x_k = t^{2/3}x_k^*(u_1, u_2, u_3), \qquad y_k = t^{2/3}y_k^*(u_1, u_2, u_3) \qquad (k = 1, 2, 3)$$

or

$$x_k = t^{2/3}x_k^*(v_1, v_2), \qquad y_k = t^{2/3}y_k^*(v_1, v_2) \qquad (k = 1, 2, 3)$$

depending on whether the limit configuration at the instant of a triple collision is an equilateral triangle or a collinear figure. The functions $x_k^*(u_1, u_2, u_3)$ and $y_k^*(u_1, u_2, u_3)$ are power series in terms of

$$u_1 = \alpha_1 t^{2/3}, \qquad u_2 = \alpha_2 t^{a_1}, \qquad u_3 = \alpha_3 t^{a_2}$$

where α_1, α_2, and α_3 are three well-defined constants; the coefficients of these series depend on the three masses of the particles. The functions $x_k^*(v_1, v_2)$ and $y_k^*(v_1, v_2)$, on the other hand, are power series in terms of

$$v_1 = \beta_1 t^{2/3}, \qquad v_2 = \beta_2 t^{b_1}$$

the constants β_1 and β_2 being determined by the trajectory of collision. In the case of the collinear limit configuration, there are two more two-parameter families of trajectories of collision which are obtained from the one mentioned above by a circular permutation of the masses m_1, m_2, and m_3. In particular cases u_1, u_2, and v_2 must be replaced by expressions which contain $\log t$. The exponents a_1, a_2, and b_1 are algebraic functions of the masses m_1, m_2, and m_3. Hence, these numbers, which, in general, are irrational, become rational only for exceptional values of the given masses. Consequently, the point $t = 0$ is in general an essential (logarithmic) singularity for the coordinates of the trajectories of collision.

Since the Lagrangian configurations of motion are the limit configurations only in the case of a real triple collision, reference to them cannot be made in the case of imaginary triple collisions.

In the case of a real collision, the condition $r = 0$ for a binary collision or the conditions $r_{ik} = 0$ for a triple collision imply the equality of the coordinates of the colliding particles at the instant of collision: $x_i = x_k$, $y_i = y_k$, $z_i = z_k$, i.e., both or all three particles have the same position in space. In the case of an imaginary collision, the mutual distances $r_{ik} = \sqrt{\Sigma(x_i - x_k)^2}$ will vanish if (i) $x_i = x_k$, $\cdots$, as in the case of a real collision; on the other hand, the coordinates and the instant of collision, in general, are complex numbers, and (ii) $\Sigma(x_i - x_k)^2$ may vanish without the individual differences being zero. In other words, at the instant of collision, the particles need not be at the same point in space. The two possible cases (i) and (ii) of imaginary collisions are said to be of the first and of the second kind respectively. Since the analytic structure of imaginary collisions of the first kind is essentially the same as that of the real collisions which Vernić mistakenly considered as regularizable, it remained for him only to deal with imaginary collisions of the second kind. The latter he claimed to have reduced to those of the first kind by a passage to an inertial coordinate system with the origin at the center of mass of the system of particles. His argument, however, consists of a chain of conclusions, some of which are erroneous.

In view of the foregoing remarks it follows that transformation (2′) and, in general, transformations of the type (2) regularize only binary collisions, and they do not accomplish anything more than the Sundman transformation.

Application to the two-body problem. As an illustration of the extended theorem consider the Sundman integration of the two-body problem. For the parabolic case the regularizing variable $u = \Theta = \tan(v/2)$, where v is the true anomaly; for the elliptic case $u = E$, where E denotes the eccentric anomaly; and for the hyperbolic case $u = F$, where

$$\tan(F/2) = [(e-1)/(e+1)]^{1/2} \tan(v/2)$$

The corresponding Sundman transformations and integrals are given below.

Parabolic case: $h = 0$, $\Theta = \tan(v/2)$. Using the Sundman transformation

$$pn\,dt = r\,d\Theta \qquad (n^2p^3 = \mu)$$

in the area integral

$$r^2\dot{v} = c = (\mu p)^{1/2} \qquad (c \neq 0)$$

we obtain the integrals

$$r = (p/2)(1 + \Theta^2), \qquad x = (p/2)(1 - \Theta^2), \qquad y = p\Theta$$

and the Barker equation which relates t and Θ:

$$2nt = \Theta + \tfrac{1}{3}\Theta^3$$

Elliptic case: $h < 0$, $0 < e < 1$. In this case the Sundman transformation is

$$an\,dt = r\,dE \qquad (n^2a^3 = \mu)$$

the integrals are

$$r = a(1 - e\cos E), \qquad x = a(\cos E - e), \qquad y = a(1 - e^2)^{1/2}\sin E$$

and the Kepler equation which relates t and E is

$$nt = E - e\sin E$$

Hyperbolic case: $h > 0$, $e > 1$. The Sundman transformation is

$$|a|n\,dt = r\,dF \qquad (n^2|a|^3 = \mu)$$

the integrals are

$$r = |a|(e\cosh F - 1), \quad x = |a|(e - \cosh F), \quad y = |a|(e^2 - 1)^{1/2}\sinh F$$

and the equation which relates t and F is

$$nt = e\sinh F - F$$

Collisions and general solutions. Consider u and t as complex variables which are eventually restricted to be real. Then the results on collision in the three-body problem can be stated as follows.

Theorem. The general solution of the three-body problem has at most a denumerable infinity of algebraic and logarithmic branch points t_ν (points of collision). The dates t_ν of continuable collisions cannot cluster at a finite limiting date t^*. Every solution has at least one real or imaginary branch point; if the number of branch points of a solution is infinite, as is the case in general, their cluster point $+\infty$ is an essential singularity.

Since Vernić mistakenly considered all the branch points t_ν to be algebraic, he had to (i) uniformize them simultaneously and (ii) study the transcendental functions which represent the general solution.

4. Theory of algebromorphic functions

For this purpose Vernić introduces the concept of an algebromorphic function, defined as an analytic (in general transcendental) multiple-valued function $w = w(z)$ which in every finite part of the t-plane has at most a finite number of algebraic branch points t_ν as critical points. Let such a function be denoted by $w = A(t; t_\nu)$. The algebromorphic functions are regular and single-valued in the Mittag-Leffler star, obtained by cutting the t-plane radially from the points t_ν to infinity.

The simplest algebromorphic functions, which are determined completely by their branch points t_ν, at which p_ν sheets of the Riemann surface unite, and which are of degree (dimension) $n_\nu (\gtrless 0)$, are called by

Vernić the fundamental algebromorphic functions. They are denoted by

$$w = A(t;\, t_v,\, n_v,\, p_v)$$

Such an algebromorphic function with branch points from the denumerable set $\{t_v\}$ without finite cluster points is represented in the Mittag-Leffler star by the fundamental formula (a generalization of Weierstrass' product formula for transcendental integral functions):

$$w = A(t;\, t_v,\, n_v,\, p_v) = \prod_{v=0}^{\infty} [(1 - t/t_v) \exp q_v(t)]^{\alpha_v} \tag{3}$$

where $\alpha_v = n_v/p_v \gtrless 0$, and $p = m\{p_v\}$, the least common multiple of all p_v, is bounded above by a constant P, i.e., $p < P$. The last condition implies that there can be at most a finite number of distinct p_v, and it requires that the corresponding Riemann surface have a finite number of sheets. One may choose the $q_v(t)$ in the convergence producing factors $\exp q_v(t)$, for example, as follows:

$$q_v(t) = \sum_{\lambda=1}^{k_v - 1} (1/\lambda)(t/t_v)^{\lambda} \tag{4}$$

This is a finite portion of the series for $-\ln(1 - t/t_v)$, in which the numbers k_v must be chosen in such a way that the series

$$\sum_{v=0}^{\infty} \alpha_v (t/t_v)^{k_v} = K(t)$$

is absolutely convergent for all t.

The above definition of a fundamental algebromorphic function is not precise. Lense [1], therefore, suggests that the fundamental formula (3) be taken as a definition for a fundamental algebromorphic function.[1] The convergence of the infinite product appearing in formula (3) can then be deduced easily from Weierstrass' product formula for integral functions.

Existence theorem of algebromorphic functions. The most general algebromorphic function with given properties is given by

$$w = E^*(t; A)$$

where A has the form (3), and the integral function E^* does not introduce any other critical points than those contained in A.

In terms of algebromorphic functions Vernić states the fundamental theorem of the three body problem as follows.

Fundamental theorem of the three-body problem. The general solution of the three-body problem with branch points t_v (points of collision) is given in terms of algebromorphic functions by

(5) $$r_i = E_i^*(t, A_i) = A_i(t; t_v)e^{E_i^*(t;A_i)} \qquad (i = 1, 2, 3)$$

This theorem, however, contradicts the previously mentioned result obtained by Siegel [1], according to which, in a neighborhood of an instant of a triple collision, series expansions with irrational exponents arise, i.e., the critical point is an isolated but essential (logarithmic) singularity.

Application to the two-body problem. As an illustration of the use of algebromorphic functions, let us consider the two-body problem. It is relatively easy to calculate the fundamental algebromorphic function A for the parabolic case. Since a collision takes place for $r = 0$, we have $1 + \Theta_0^2 = 0$, or $\Theta_0 = \pm i$. Further, the Barker equation gives the branch points

$$2nt_{1,2} = \pm\left(i - \frac{i}{3}\right) = \pm\frac{2i}{3}$$

or

$$t_{1,2} = \pm\frac{i}{3n}$$

at which $p_{1,2} = 2$ sheets of the Riemann surface unite and which are of dimensions $n_{1,2} = 1$. Hence we have

$$w = A(t, t_v) = \left(1 - \frac{t}{t_1}\right)^{1/2}\left(1 - \frac{t}{t_2}\right)^{1/2} = (1 + 9n^2t^2)^{1/2}$$

which is also the discriminant of the Barker equation. Application of the Cardano formula to this equation, rewritten in the form

$$\Theta^3 + 3\Theta - 6nt = 0$$

gives the regularizing variable Θ as an algebromorphic function of the time in the form

$$\Theta = [3nt + (1 + 9n^2t^2)^{1/2}]^{1/3} + [3nt - (1 + 9n^2t^2)^{1/2}]^{1/3}$$

The general integral is an algebromorphic function of t, given by

$$r = \frac{p}{2}(1 + \Theta^2) = E^*(t, A)$$

Furthermore, it can be shown that the integration of the differential equation of the two-body problem

$$r^2\ddot{r} + r\dot{r}^2 - 2hr = \mu$$

using algebromorphic functions

$$r = E^*(t, A), \qquad A = \prod_v (1 + c_v^2 t^2)^{1/2}$$

and the Sundman transformation

$$C\,dt = r\,du$$

leads to the classical integrals of the problem in all three cases—elliptic, parabolic, and hyperbolic.

5. Regularization and uniformization

In the general three-body problem one must distinguish between uniformization of the trajectories in the parametric form

$$(6) \qquad x_i = x_i(s), \quad y_i = y_i(s), \quad z_i = z_i(s) \qquad (i = 1, 2, 3)$$

and uniformization of the motion (r_i, t) in the parametric form

$$(7) \quad r_i = P_i(u), \quad x_i = x_i(u), \quad y_i = y_i(u), \quad z_i = z_i(u), \quad t = T(u)$$

Uniformization of the trajectories is a geometric uniformization of the problem, whereas that of the motion (r_i, t) is its kinematic uniformization.

Theorem. Every uniformization of the two- and the three-body problem is not necessarily a regularization of the problem concerned, hence it is not necessarily an integration of the problem. In general, it is necessary to uniformize the time t as well as the coordinates and the mutual distances; for instance, in the case of the two-body problem by the use of the Sundman transformation (1). Only by this last operation is a uniformization of the trajectories generalized to a uniformization of the motion. The function $1/V$ is an integrating factor which makes $C\,dt$ in (1) an exact differential.

For example, in the case of an elliptic motion in the two-body problem, the transformation

$$w = [(1 - e)/(1 + e)]^{1/2} \tan \frac{v}{2}$$

uniformizes r (and also the coordinates), since it provides the well-known rational representation of the equation of the ellipse

$$r = a\left(1 - e\,\frac{1 - w^2}{1 + w^2}\right)$$

This transformation, however, does not uniformize the motion. A simple calculation shows that

$$an\,dt = r\frac{2dw}{1 + w^2}$$

In order to uniformize the motion, one has to put

$$\frac{2dw}{1 + w^2} = dE$$

After this substitution has been performed, we obtain the correct Sundman transformation

$$an\,dt = r\,dE$$

which permits the integration and gives

$$\begin{aligned} E &= 2 \text{ arc tan } w \\ w &= [(1 - e)/(1 + e)]^{1/2} \tan\frac{v}{2} = \tan\frac{E}{2} \\ r &= a(1 - e\cos E) \\ x &= a(\cos E - e), \qquad y = a(1 - e^2)^{1/2} \sin E \\ nt &= E - e\sin E \end{aligned}$$

In this example the expression

$$\frac{dw}{dE} = \frac{1}{2}(1 + w^2)$$

is the integrating factor of the equation

$$an\frac{dt}{dw} = \frac{2r}{1 + w^2}$$

In the hyperbolic and the parabolic cases the kinematic uniformization is similarly accomplished by the variables F and Θ respectively.

6. Uniformization of algebromorphic functions

The problem of uniformization is concerned with the representation of a functional relation $f(z, w) = 0$ (in general multiple-valued) by means of parametric equations $z = z(t)$, $w = w(t)$, where $z(t)$ and $w(t)$ are single-valued functions of t in the neighborhood of the point, say, $t = 0$, corresponding to the point $z = z_0$, $w = w_0$ on the Riemann surface of the function $w = w(z)$ defined by $f(z, w) = 0$. The auxiliary variable t is called a local uniformizing variable for the functional relation $f(z, w) = 0$. The point (z_0, w_0) is an algebraic or a transcendental branch point of our Riemann surface, according to whether a finite or an infinite number of sheets of the Riemann surface unite at this point; or, in other words, according to whether a finite or an infinite number of branches of the multiple-valued function $w = w(z)$ permute around the point (z_0, w_0). The role of the local uniformizing variable t is to map onto a simple neighborhood of $t = 0$ a neighborhood of the point (z_0, w_0) on the Riemann surface of the function $w = w(z)$. The theorem of Poincaré and Koebe (1907) on total uniformization in the large of a given function states the existence of a total uniformizing variable t which at every point of the corresponding Riemann surface has the character of a local uniformizing variable and which maps the entire Riemann surface onto a simple region of the t-plane.

This means that all functions which are single-valued on the Riemann surface can be expressed as single-valued functions of t in the above region of the t-plane.

After these general remarks let us discuss an example which will be used later on. Consider the unit circle $z^2 + w^2 = 1$, of which a rational parametrization is

$$z = \frac{2t}{1 + t^2}, \qquad w = \frac{1 - t^2}{1 + t^2}$$

Further, consider the Riemann surface R_2 of the function $w = \sqrt{1 - z^2}$. This surface has two sheets over the z-plane, with branch points at $z = -1$ and $z = +1$. It can be mapped onto the simple t-plane by, say, the function

$$t = \frac{1 + \sqrt{1 - z^2}}{z} \tag{8}$$

Another rational parametrization of the unit circle can be obtained by using, for instance, the function

$$t = \sqrt{\frac{1 - z}{1 + z}} \tag{9}$$

for mapping the Riemann surface R_2 onto the t-plane. This would have yielded

$$z = \frac{1 - t^2}{1 + t^2}, \qquad w = \frac{2t}{1 + t^2}$$

which is a new uniformization of the circle. In general, all possible parameters t can be obtained from one of them by means of linear transformations.

The two-sheeted Riemann surface R_2 is by no means the only one on which $w = \sqrt{1 - z^2}$ is a single-valued function of the position. Let us consider, for instance, the Riemann surface R_4 of four sheets that has its branch points at $z = \pm 1$. On this surface too, $w = \sqrt{1 - z^2}$ is single-valued. If we map this surface in a one-to-one fashion onto the t-plane by means of

$$(10) \qquad t = \sqrt[4]{\frac{1 - z}{1 + z}}$$

we obtain a new parametric representation

$$z = \frac{1 - t^4}{1 + t^4} \qquad w = \frac{2t^2}{1 + t^4}$$

This, however, is not contained in the set we obtained earlier.

Finally, to arrive at a uniformization of the unit circle $z^2 + w^2 = 1$ by means of trigonometric functions, we can start by considering a Riemann surface R_∞ for which the branch points $z = \pm 1$ are of order infinity, i.e., a Riemann surface with an infinite number of sheets. The function

$$(11) \qquad t = \arcsin z = -i \ln (iz + \sqrt{1 - z^2}\,) = -i \ln (iz + w)$$

maps this surface onto the simple t-plane. Solving for z and w, we obtain a new parametrization

$$z = \sin t, \qquad w = \cos t$$

On the other hand, if one chooses for the mapping function

$$(12) \qquad it = \ln (z + i\sqrt{1 - z^2}\,) = \ln (z + iw),$$

we obtain a second trigonometric parametrization

$$z = \cos t, \qquad w = \sin t$$

The uniformizing variables (8) and (9) are single-valued on the above-mentioned R_2, whereas (10) is single-valued on an R_4 which has the same branch points as R_2 and of which R_2 is a part. The uniformizing variables (11) and (12) are single-valued on an R_∞ which is the recovering surface of R_2 and all functions that are single-valued functions of position on this surface. The variables (10), (11), and (12), on the other hand, uniformize more comprehensive surfaces R_m ($m = 4$ and $m = \infty$ respectively) and all functions that are single-valued on R_m; hence, *a fortiori* also the given surface R_2 and all single-valued functions on it.

Finally, we mention the following theorem of Koebe (1927): Given a sequence of local uniformizing variables $u_1, u_2, \cdots, u_v, \cdots$; let $\omega_1, \omega_2, \cdots, \omega_v, \cdots$ be the values corresponding to O in these mappings of the given Riemann surface onto a simple region of the u-plane. Then the most general global uniformizing variable is given by

$$\bar{u} = u \prod_{v=1}^{\infty} \frac{u_v}{\omega_n}$$

where the factor u corresponds to the branch point O. This variable $\bar{u}$ maps the complete recovering surface of the given Riemann surface one-to-one onto a simple region of the u-plane.

Applying the above-mentioned principles for solving the problem of uniformization to algebromorphic functions, we obtain immediately the following result. On the Riemann surface of a fundamental algebromorphic function $w = A(t; t_v)$ with branch points t_v of dimensions α_v a local uniformizing variable of the branch point t_v is given by the formula

$$u_v = (t_v - t)^{\alpha_v} \exp [\alpha_v q_v(t)]$$

Since the convergence-producing factor $\exp (\alpha_v q_v)$, being a regular function, is of no importance in a local uniformization, it can be introduced at the very beginning. Because of the Mittag-Leffler star we exclude the origin $t = 0$ from our consideration and apply the above-mentioned theorem of Koebe to the sequence of branch points $\{t_v^{\alpha_v}\}$. Then the most general global uniformizing variable is

$$u = \prod_{v=0}^{\infty} \frac{u_v}{t_v^{\alpha_v}} = \prod_{v=0}^{\infty} [(1 - t/t_v) \exp q_v(t)]^{\alpha_v} \tag{13}$$

This, however, is one of the fundamental algebromorphic functions. More generally, one may introduce the global uniformizing variable u^* as an entire function of u:

$$u^* = E_*(u)$$

For $t = t_v$ the right-hand side of formula (13) is zero; hence the inverse function $t = \varphi(u)$ will have a logarithmic branch point at $u = 0$. This difficulty, however, can be avoided by a proper choice of the function $E_*(u)$. Consider, for example, the Kepler equation

$$nt = E - e \sin E \tag{14}$$

If the term E were not present on the right-hand side of (14), we would have

$$E_* = - \operatorname{arc} \sin \frac{nt}{e}$$

and the above-mentioned difficulty would appear. The presence of the variable term E, however, avoids this difficulty, and the inverse function $E = E(t)$ is given as a regular function of t by the well-known Bessel series.

7. Uniformization of the two-body problem

Theorem. In the elliptic case of the two-body problem the eccentric anomaly E is a geometric uniformizing variable; the corresponding Riemann surface has an infinite number of sheets.

In fact, if we suppose that the Sun is at the right focus of the ellipse, the equation of the orbit of a planet is

$$\left(\frac{x}{a} + e\right)^2 + \left(\frac{y}{a\sqrt{1-e^2}}\right)^2 = 1$$

which, by means of the transformations

$$\xi = \frac{x}{a} + e, \qquad \eta = \frac{y}{a\sqrt{1-e^2}}$$

goes over into the equation of the unit circle

$$\xi^2 + \eta^2 = 1$$

Now any of the previously mentioned uniformizing variables can be applied; for example, put

$$\ln\,(\xi + i\sqrt{1-\xi^2}) = iE$$

Then

$$\xi = \cos E, \qquad \eta = \sin E$$

and

$$x = a(\cos E - e), \qquad y = a\sqrt{1-e^2}\sin E$$

Finally, the equation $x^2 + y^2 = r^2$ yields $r = a(1 - e\cos E)$

Similarly, in the case of a hyperbola, we have

$$\xi = e - \frac{x}{|a|}, \qquad \eta = \frac{y}{|a|\sqrt{e^2-1}}$$

$$\ln\,(\xi - \sqrt{\xi^2 - 1}\,) = F$$

$$x = |a|(e - \cosh F), \qquad y = |a|\sqrt{e^2-1}\sinh F$$

$$r = |a|\,(e\cosh F - 1)$$

Finally, in the parabolic case $\Theta = \tan(v/2)$ is a geometric uniformizing variable.

It was previously pointed out that in the three cases of the two-body problem the variables E, F, and Θ are at the same time the kinematic uniformizing variables. This coincidence of the geometric and the kinematic uniformizing variables provides the key to the complete integrability of the two-body problem in a closed form.

8. Kinematic uniformization of the three-body problem

Theorem. The Sundman and Levi-Civita transformation

$$dt = du/V \qquad (u = 0, t = 0) \tag{1a}$$

provides an integrating factor for the Lagrange-Jacobi equation

$$\frac{1}{2}\frac{d^2J^2}{dt^2} = V + 2h$$

and, hence, implies the existence of a new integral (the so-called Bohlin integral) with respect to the uniformizing variable u:

$$\frac{1}{2}\frac{dJ^2}{dt} = u + 2ht$$

Here J^2 denotes the moment of inertia and h the energy constant of the system.

The truth of the theorem is immediately seen by differentiating the last equation and using (1a).

In the case of the two-body problem, the Lagrange-Jacobi equation reads

$$\frac{1}{2}\frac{d^2r^2}{dt^2} = \frac{\mu}{r} + 2h$$

By application of the Sundman transformation for the elliptic case, for example,

$$an\,dt = r\,dE$$

the distance r which appears in the right-hand side of the Lagrange-Jacobi equation can be eliminated, and the equation assumes the form

$$\frac{1}{2}\frac{d^2r^2}{dt^2} = a^2n\frac{dE}{dt} + 2h$$

Integrating this equation, we obtain the Bohlin integral

$$\frac{1}{2}\frac{dr^2}{dt} = a^2nE + 2ht$$

which, because of the relation $2h = -a^2n^2$, can be rewritten in the form

$$\frac{1}{2a^2n}\frac{dr^2}{dt} + nt = E$$

Finally, substituting in this equation the Bessel expansion for r^2, namely,

$$r^2 = a^2\left[1 + \frac{3e^2}{2} - 4\sum_{m=1}^{\infty}\frac{J_m(me)}{m^2}\cos mnt\right]$$

we obtain the Bessel form for the regularizing or uniformizing variable E:

$$E = nt + 2\sum_{m=1}^{\infty}\frac{J_m(me)}{m}\sin mnt$$

The final theorem of Vernić on uniformization of the three-body problem is also false since the fundamental theorem of the three-body problem, as formulated by him and on which this theorem depends, does not hold.

NUMERICAL SOLUTION OF THE GENERAL THREE-BODY PROBLEM

9. Numerical regularization

The numerical solution of the general three-body problem consists in the integration of the Newtonian differential equations of motion

$$(15)\quad m_i\frac{d^2x_i}{dt^2} = \frac{\partial V}{\partial x_i},\quad m_i\frac{d^2y_i}{dt^2} = \frac{\partial V}{\partial y_i},\quad m_i\frac{d^2z_i}{dt^2} = \frac{\partial V}{\partial z_i}\qquad (i = 1, 2, 3)$$

where

$$V = \sum_{1}^{3} m_jm_k/r_{jk}\qquad (j \neq k)$$

and the constant of gravitation is one, by the method of successive approximations; or, to be more specific, in the successive calculation of the coefficients of the Taylor series expansions for the coordinates and the mutual distances of the particles. The general three-body problem cannot be solved by means of power series in t since, in general, such series diverge. The singularities of motion are binary and triple collisions of the particles. After the introduction of the variable u by the Sundman and Levi-Civita transformation, the differential equations of motion become

$$(16)\qquad \frac{d}{du}\left(V\frac{dx_i}{du}\right) = \frac{1}{m_i}\frac{\partial \ln V}{\partial x_i}\qquad (i = 1, 2, 3)$$

(and two similar equations with respect to y_i and z_i), or, explicitly,

$$x_i'' + \frac{V'}{V} x_i' = \frac{1}{m_i V^2} \frac{\partial V}{\partial x_i} \qquad (i = 1, 2, 3) \tag{17}$$

(and two similar equations with respect to y_i and z_i), where primes denote derivation with respect to u.

The general solution is analytic in the "pseudotime" u, and can be obtained from (16), respectively (17), by differentiation and elimination in the form of Taylor series

$$x_i = \sum_{v=0}^{\infty} \xi_{iv} u^v, \quad y_i = \sum_{v=0}^{\infty} \eta_{iv} u^v \quad z_i = \sum_{v=0}^{\infty} \zeta_{iv} u^v \quad (i = 1, 2, 3) \tag{18}$$

$$r_{ik} = \sum_{v=0}^{\infty} \rho_{ik,v} u^v \qquad (i \neq k) \tag{19}$$

where

$$\xi_{iv} = \frac{1}{v!}\left(\frac{d^v x_i}{du^v}\right)_0, \quad \cdots, \quad \rho_{ik,v} = \frac{1}{v!}\left(\frac{d^v r_{ik}}{du^v}\right)_0$$

The subscript zero indicates that the derivatives are evaluated at $u = 0$.

10. Numerical example

The case of the general three-body problem with equal masses $m_1 = m_2 = m_3 = 1$ was considered earlier by Zumkley [1], who used the method of numerical integration. Vernić [5] uses the same example to illustrate the numerical solution by power series in u. Explicit formulas for the first three successive approximations are given up to the eighth degree in u. The coefficients are calculated to three decimal places, and the results are given in three tables not reproduced here. Comparison of the second and third tables shows that higher order approximations do not affect the values of the coefficients calculated, i.e., the approximation process is to be considered as finished. The range of validity of both methods (Zumkley and Vernić) depends on the accuracy chosen.

The formulas for $x_1, x_2, x_3; y_1, y_2, y_3; t$, as given in table 3, are:

$$\begin{aligned} x_1 &= 2 - 0.066u^2 + 0.005u^4 + 0.000_1u^6 + 0.000_1u^8 + \cdots \\ x_2 &= 1 + 0.136u^2 - 0.003u^4 - 0.000_7u^6 + 0.000u^8 + \cdots \\ x_3 &= 3.5 - 0.071u^2 - 0.000_6u^4 + 0.000_5u^6 + 0.000u^8 + \cdots \end{aligned}$$

$$\begin{aligned} y_1 &= -0.023u^3 + 0.001u^5 + 0.000_5u^7 + \cdots \\ y_2 &= -0.726u + 0.031u^3 - 0.001u^5 + 0.001u^7 + \cdots \\ y_3 &= 0.484u - 0.009u^3 - 0.000_2u^5 - 0.000u^7 + \cdots \end{aligned}$$

$$t = 0.484u + 0.008u^3 - 0.001u^5 + 0.000_2u^7 + \cdots$$

These formulas give the general solution of a particular case of the general three-body problem which has not been dynamically simplified.

The coefficients in the above series for the coordinates and t decrease very rapidly and after the first three terms they are ≤ 0.001, i.e., the limiting accuracy is reached. On the basis of this example Vernić claims that on the average the convergence of the regularized power series (18) and (19) in u will be quite satisfactory, contrary to the views expressed by Belorizky [1–4].

PERIODIC SOLUTIONS OF THE GENERAL THREE-BODY PROBLEM

11. A periodic boundary value problem

We can now discuss the question of the existence of periodic solutions in the general three-body problem. By means of the Sundman and Levi-Civita transformation

$$dt = du/V$$

the Lagrange-Jacobi equality

$$\frac{1}{2}\frac{d^2J^2}{dt^2} = V + 2h$$

which is the most important formula of the three-body problem, assumes the form

$$\frac{d}{du}\left(V\frac{dJ^2}{du}\right) = 2\left(1 + \frac{2h}{V}\right) \tag{20}$$

Vernić [2] chooses equation (20) as his working equation, and proves the following theorems with respect to the existence of periodic solutions without collisions.

Theorem 1. For the periodic solutions of the general three-body problem the Lagrange function J^2 and the potential V are periodic functions of u with one and the same period ω.

Theorem 2. The periodic solutions of the general three-body problem are contained in the set of periodic solutions of equation (20).

Theorem 3. A necessary condition for the existence of periodic solutions in the general three-body problem is that the total energy constant h be less than zero.

Next, Vernić uses Green's function to solve the periodic boundary-value problem for an ordinary linear, nonhomogeneous second-order self-adjoint differential equation:

$$L(J^2) = \frac{d}{du}\left(p\frac{dJ^2}{du}\right) + qJ^2 = f(u)$$

$$J^2(0) = J^2(\omega)$$

$$\left(\frac{dJ^2}{du}\right)_0 = \left(\frac{dJ^2}{du}\right)_\omega$$

This method is then applied to equation (20) which is of parabolic type:

$$L(J^2) = \frac{d}{du}\left[V(u)\frac{dJ^2}{du}\right] = 2\left[1 + \frac{2h}{V(u)}\right] = f(u)$$

The principal theorem of Vernić states that the only periodic solutions of the general three-body problem are the Lagrange exact solutions. Merman [2], however, showed that Vernić's proof is defective. On the other hand, he has not proved that Vernić's theorem concerning the existence of the Poincaré periodic solutions in the general three-body problem is false. Vernić himself is aware that his theorem contradicts the views held by most mathematicians and astronomers. However, he tries to make the truth of his theorem intuitively plausible by arguments outlined in the next section.

12. Discussion of Vernić's principal theorem

Let us begin by recapitulating several basic facts. First, nobody since Lagrange (1772) has found concrete periodic solutions of the general three-body problem. Second, the starting point in the proof of the existence of the Poincaré periodic solutions of the first kind (*solutions périodiques du première genre*, 1892) is a twofold degenerate three-body problem: a finite mass $m_1 = 1$ and two infinitesimal masses $m_2 = \alpha_2\mu = 0$, $m_3 = \alpha_3\mu = 0$. The unperturbed motion consists of two independent Kepler motions of m_2 and m_3 around m_1. The motion of this system of three particles is periodic, and this motion is real if $h < 0$. If $\mu \neq 0$ but sufficiently small, periodic solutions of the problem still exist. According to the nature of the unperturbed motion ($\mu = 0$), (i) circular motion in a plane ($e = 0$, $i = 0$, where e denotes the eccentricity of the orbit and i the inclination of the plane of the orbit), (ii) elliptic motion in a plane ($e \neq 0$, $i = 0$), or (iii) elliptic motion in space ($e \neq 0$, $i \neq 0$), the periodic solutions for $\mu \neq 0$ are called of the first, the second, or the third type respectively (*solutions périodiques de la première, seconde, ou* *troisième sorte*). In all three cases the period T of the perturbed motion ($\mu \neq 0$) is the same as that of the unperturbed motion ($\mu = 0$). For the periodic solutions of the second kind of Poincaré (*solutions périodiques du deuxième genre*, 1899), the period of motion is a multiple of the period T of the unperturbed motion. However, no satisfactory proof of the existence of such periodic solutions has been given when $\alpha_2 \neq 0$ and $\alpha_3 \neq 0$, either by Poincaré or by anybody else. In fact, their existence has not been demonstrated even for the case $\alpha_2 \neq 0$, $\alpha_3 = 0$ (restricted problem). Third, Poincaré himself did not complete his investigations on the restricted problem; his investigations of the existence of periodic solutions for sufficiently small μ were continued by Strömgren, who showed, by numerical methods, the existence of periodic solutions for "large" values of the perturbing mass μ (with respect to a rotating coordinate system). In the opinion of Vernić, none of the above-mentioned facts contradict his theorem.

We should like to add the following remarks. Since the right-hand sides of the equations of motion in the general three-body problem do not depend explicitly on the time t, periodic solutions with arbitrary periods exist. Poincaré (1892) was the first to call attention to this fact. He was, however, concerned only with periodic solutions the periods of which are equal to those of the generating solutions (solutions of the first kind). Historically Schwarzschild (1898) was the first to obtain, in the planar restricted problem, periodic solutions the periods of which differ from those of the generating solutions ($\mu = 0$) by quantities of the order of the small perturbing mass μ. Solutions of this type were obtained in the planar general three-body problem by Heinrich [1] in 1926, and in the circular restricted problem in space by Batrakov (1955). Finally, Batrakov [2] showed the existence of periodic solutions of the above-mentioned type in the general three-body problem in space. The properties of Batrakov's solutions are analogous to those of the Poincaré solutions of the third type. However, in the general three-body problem, the Poincaré periodic solutions of the third type, the above-mentioned solutions of the second type obtained by Heinrich, as well as those obtained by Batrakov are merely sinodic, not sideric.

Leaving aside the question as to the existence of the Poincaré periodic solutions in the general three-body problem, Merman [2] showed first the existence of the Lyapunov periodic solutions in the restricted three-body problem. Such solutions remain periodic even after the passage to an inertial coordinate system. Merman was then able to deduce the existence of the Lyapunov periodic solutions in the general three-body problem.

Merman's proof runs as follows. Consider a canonical system of equations

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial y_i}, \quad \frac{dy_i}{dt} = -\frac{\partial H}{\partial x_i} \qquad (i = 1, 2, \cdots, n)$$

where

$$H = \sum_{m=2}^{\infty} H_m$$

and H_m is a form of degree m in $x_1, \cdots, x_n; y_1, \cdots, y_n$ with constant coefficients. According to a theorem of Lyapunov [1], to every pair of pure imaginary roots $\pm s_k i$ of the characteristic equation

$$\begin{vmatrix} \frac{\partial^2 H_2}{\partial x_1^2} - s & \cdots & \frac{\partial^2 H_2}{\partial x_1 \, \partial y_n} \\ \cdot & \cdots & \cdot \\ \frac{\partial^2 H_2}{\partial y_n \, \partial x_1} & \cdots & \frac{\partial^2 H_2}{\partial y_n^2} - s \end{vmatrix} = 0$$

(s_k/s_j not an integer), there corresponds a family of periodic solutions which depends upon two arbitrary constants c and t_0. The period is a continuous (in fact analytic) function of c in the neighborhood of $c = 0$. Consider now the planar restricted three-body problem with respect to a rotating coordinate system. Introduce new coordinates, equal to the differences between the old coordinates and the corresponding coordinates of the triangular libration points. Then a Lyapunov system is obtained, the characteristic equation of which has a pair of pure imaginary roots and a pair of real roots which depend on the masses. Because of the continuous dependence of the period of the solutions so obtained upon c, there exists an infinity of periodic solutions, the period of which is commensurable with the period of rotation of the coordinate axes. Hence, these solutions will remain periodic after a passage to an inertial coordinate system.

The existence of analogous periodic solutions in the neighborhood of the Lagrangian solutions can be easily deduced also in the general three-body problem.

Thus the theorem of Vernić on the nonexistence of periodic solutions other than the Lagrangian solutions in the general three-body problem has been shown to be false.

Certain points in the theory of the Poincaré periodic solutions of the third type have been clarified in a series of notes by Gremillard [1–4].

BIBLIOGRAPHY

Batrakov, Yu. V. [1], Periodic solutions of Schwarzschild's type in the restricted problem of three bodies (Russian), *Byull. Inst. Teor. Astr.*, **6,** 112–120 (1955).

——— [2], On periodic solutions of the third type in the general three-body problem (Russian), *Byull. Inst. Teor. Astr.*, **6,** 121–126 (1955).

Belorizky, D. [1], Sur la convergence des séries dans la solution du problème des trois corps donnée par M. Sundman, *C. R. Acad. Sci. Paris,* **193,** 314–317 (1931).

——— [2], Sur la solution du problème des trois corps donnée par M. Sundman, *C. R. Acad. Sci. Paris,* **193,** 766–768 (1931).

——— [3], Sur l'application des méthodes de M. Sundman aux problèmes de la mécanique céleste, *C. R. Acad. Sci. Paris,* **193,** 1321–1323 (1931).

——— [4], *Journal des Observateurs,* **16,** nos. 7, 9 and 11 (1933).

Gremillard, J. [1], Sur les solutions périodiques de la troisième sorte dans le problème des trois corps, *C.R. Acad. Sci. Paris,* **234,** 2339–2341 (1952).

——— [2], Sur certains solutions périodiques du problème des trois corps, *C.R. Acad. Sci. Paris,* **236,** 49–51 (1953).

——— [3], Sur la recherche des solutions périodiques de la troisième sorte, *C.R. Acad. Sci. Paris,* **236,** 1952–1953 (1953).

——— [4], Sur les racines d'une équation de la théorie des solutions périodiques de la troisième sorte, *C.R. Acad. Sci. Paris,* **239,** 153–155 (1954).

Heinrich, W. W. [1], Sur une éxtension du domain des solutions périodiques de la seconde sorte dans le problème général des trois corps, *Publ. Inst. Astr. Univ. Charles de Prague* (2), no. **3,** 1–8 (1926).

Lense, J. [1], Über algebroide und algebromorphe Funktionen, *Bayer. Akad. Wiss. Math.-Nat. Kl. S.-B. 1956,* 237–242 (1957).

Lyapunov, A. M. [1], Problème général de la stabilité du mouvement, *Ann. Fac. Sci. Toulouse* (2), **9,** 293–474 (1907). Reprinted in *Annals of Math. Studies,* no. 17, Princeton Univ. Press, Princeton (1949).

Merman, G. A. [1], A new class of periodic solutions in the restricted problem of three bodies and in Hill's problem (Russian), *Trudy Inst. Teor. Astr.*, **1,** 5–86 (1952).

——— [2], Zu Arbeiten von R. Vernić über die Regularisierung und die periodischen Lösungen des Dreikörperproblems (Russian), *Byull. Inst. Teoret. Astr.*, **6,** 408–415 (1956).

Siegel, C. L. [1], Der Dreierstoss, *Ann. of Math.* (2), **42,** 127–168 (1941).

Strömgren, E. [1], Connaissance actuelle des orbites dans le problème des trois corps, *Bull. Astr.* (2), **9,** 87–130 (1935) and *Publ. Kbh. Obs.*, no. **100,** 1–44 (1935).

Vernić, R. [1], Bahnen des restringierten Dreikörperproblems dargestellt im Inertialsystem (Serbian, German Summary), *Jugoslav. Akad. Znan. Umjetn. Rasprave Odjel. Mat. Fiz. Techn. Nauke,* **1,** 81–123 (1952).

Vernić, R. [2], Periodische Lösungen im Dreikörperproblem, *Hrvatsko Prirod. Društvo Glas. Mat. Fiz. Astr.*, ser. II, **8,** 247–266 (1953).

——— [3], Discussion der Sundmanschen Lösung des Dreikörperproblems, *Jugoslav. Akad. Znan. Umjetn., Zagreb,* 146 pp. (1954).

——— [4], Die Stossbedingungen im Dreikörperproblem, *Hrvatsko Prirod. Društvo Glas. Mat. Fiz. Astr.*, ser. II, **9,** 3–13 (1954).

——— [5], Numerische Auflosung des allgemeinen Dreikörperproblems (Serbian, German Summary), *Rad Jugoslav. Akad. Znan. Umjetn. Odjel. Mat. Fiz. Techn. Nauke,* 6, **302,** 47–76 (1955).

Wintner, A. [1], Beweis des E. Strömgrenschen dynamischen Abschlussprinzips der periodischen Bahngruppen im restringierten Dreikörperproblem, *Math. Z.*, **34,** 321–349 (1931) and *Publ. Kbh. Obs.*, no. 75, 1931.

——— [2], *The Analytical Foundations of Celestial Mechanics,* Princeton Univ. Press, Princeton, 335–338 (1947).

Zumkley, J. [1], Ein numerisch gerechneter Spezialfall des allgemeinen Dreikorperproblems in vereinfachter Behandlung, *Astr. Nachr.*, **272,** 66–76 (1941).

CHAPTER 2

CAPTURE IN THE THREE-BODY PROBLEM

CHAZY'S AND SHMIDT'S INVESTIGATIONS

1. Introduction

The problem of capture of one body by another (or mutual capture between two bodies) arose in celestial mechanics in connection with various cosmogonical hypotheses on the origin of visual binary stars. Such a theory was proposed first by Stoney (1867), and later supported by Lord Kelvin, Biccerton, Arrhenius, See (1910), F. R. Moulton, Bosler, and others.

In a series of papers beginning in 1944, Shmidt presented a new cosmogonical hypothesis on the origin of the planetary system. He pointed out that none of the existing theories are able to explain the distribution of the moment of momentum of the solar system between the Sun and its planets. Whereas the Sun contains almost the whole mass of the system but only 2 per cent of the moment of momentum, the planets have only 1/700 of the total mass and 98 per cent of the moment of momentum. This circumstance led him to the idea that the masses of the planets had been captured by the Sun from the Galaxy. The actual distribution of the moment of momentum is a result of a redistribution of the original momenta of the Sun and the captured particles in their motion about the center of the Galaxy.

In the case of two isolated bodies subject to their mutual Newtonian attraction the motion of one with respect to the other takes place in a fixed conic and capture is impossible. The case of three bodies presents a greater variety of possibilities.

Definition 1. We shall say that capture takes place between two bodies if at a time t_1 the relative motions of the three pairs of bodies are hyperbolic and all their mutual distances from one another exceed a constant R_1, whereas at a later time t_2 two of the bodies are within the distance $R_2 < R_1$ and have elliptic motions, and the third body is at a distance greater than R_1 from the common center of mass of all three bodies and has a hyperbolic motion with respect to the center of mass of the other two bodies.

This is Shmidt's definition of capture for the time interval (t_1, t_2), given in a weakened form. The constants R_1 and R_2 are interpreted respectively as the mean stellar distance in the Galaxy and the mean diameter of a binary star.

Is capture possible in the three-body problem? During the 19th century and the first half of the 20th century the majority of the astronomers and mathematicians shared the opinion that capture in the three-body problem is impossible. The astronomers discussed the problem of transformation of an asteroid into a satellite of a planet. We have then the Sun (S), Jupiter (J) and a third infinitesimal body P which moves in such a way that, although it is subject to the Newtonian attractions of S and J, it does not disturb the circular Keplerian motion of the two finite bodies about their common center of mass. This model is known as the circular case of the restricted problem of three bodies. The impossibility of capture in this case was proved by von Zeipel [1], Hopf [1], and Fesenkov [1].

2. Chazy's asymptotic investigations

For the general three-body problem the most complete results relating to the asymptotic behavior of motion (when $t \to \infty$) are those of Chazy [1–4]. Let h denote the energy constant of the system when the motion is referred to the center of mass of the three bodies. In addition, assume that the angular momentum of the bodies is different from zero about every axis through the center of mass. Chazy obtained the following types of motion of positive h, $h = 0$, and of negative h in terms of the order of magnitude (for large t) of the three mutual distances r_{ik} of the three bodies.

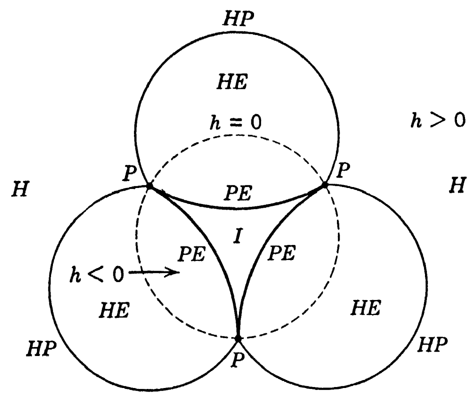
Case 1. $h > 0$: hyperbolic ($r_{ik} \sim t$); hyperbolic-parabolic (r_{13}, $r_{23} \sim t$; $r_{12} \sim t^{2/3}$); hyperbolic-elliptic (r_{13}, $r_{23} \sim t$; $r_{12} < a$, a = finite).
Case 2. $h = 0$: hyperbolic-elliptic; parabolic ($r_{ik} \sim t^{2/3}$).
Case 3. $h < 0$: hyperbolic-elliptic; parabolic-elliptic (r_{13}, $r_{23} \sim t^{2/3}$, $r_{12} < a$); bounded ($r_{ik} < a$); oscillating (at least one of the r_{ik} is neither infinite nor bounded).

Except for the last two types of motion there are three classes of each type, depending on which of the three bodies recedes to infinity.

In order to define analytically the motion of the three bodies P_0, P_1, and P_2, introduce Jacobian coordinates. Let x, y, z be the coordinates of P_1 with respect to a coordinate system with the origin at P_0, and let ξ, η, ζ be the coordinates of P_2 with respect to a coordinate system with the origin at the center of mass G of P_0 and P_1, the corresponding coordinate axes being parallel in both systems. Consider the following two relative motions: (i) the motion of P_1 with respect to P_0, and (ii) the motion of P_2 with respect to G, and denote them by P_1/P_0 and P_2/G respectively.

Analytically the motion of the three bodies is determined by six second-order differential equations corresponding to the motions P_1/P_0 and P_2/G. The corresponding trajectories are defined by the twelve initial values of the coordinates and their derivatives, x_0, y_0, z_0, x_0', y_0', z_0', ξ_0, η_0, ζ_0, ξ_0', η_0', ζ_0'. It is convenient to associate with these twelve initial values a point of a twelve-dimensional Euclidean space E_{12} which has these values for its coordinates. To the motion of the three bodies with the above-mentioned initial values then corresponds a well-defined trajectory in the twelve-dimensional manifold of states of motion E_{12} of the three bodies.

The previously mentioned types of motion divide this manifold E_{12} into five regions, represented symbolically by the following scheme:



In the twelve-dimensional manifold E_{12} the bounded and oscillating motions correspond to the interior region I. The hyperbolic-elliptic motions (HE) correspond to three regions bounded by the eleven-dimensional submanifolds (PE) and (HP), representing parabolic-elliptic and hyperbolic-parabolic motions respectively. The motions are parabolic on the ten-dimensional submanifold (P) along which the surfaces (PE)

and (HP) meet again. This manifold (P) is situated on the surface corresponding to motions for which the energy constant h is zero. The hyperbolic motions (H) correspond to the exterior of the bounding regions (HP).

Furthermore, using integral invariants, Chazy [3] arrived at the conclusion that capture in the domain $h < 0$ is impossible. Recently Gazaryan [1] has expressed some doubts as to the correctness of certain of Chazy's deductions. What Chazy really proved in the last paper mentioned above is the following. For $h < 0$ there do not exist pencils of trajectories which are bounded as $t \to \infty$ and hyperbolic-elliptic as $t \to -\infty$, or conversely. Later on this theorem was misinterpreted as a proof for the impossibility of capture in the general three-body problem, the fact being forgotten that it was proved for the case $h < 0$ only.

Merman [5] has recently re-examined the above-mentioned paper of Chazy, and he proves by simplifying and carrying out in detail some of the proofs only sketched by Chazy that the probability of capture in the domain $h < 0$ is zero. In particular, he shows that almost all motions which are hyperbolic-elliptic for $t \to \infty$ are also hyperbolic-elliptic for $t \to -\infty$. Chazy claims that these two hyperbolic-elliptic motions belong to the same class. His proof, however, is not convincing. Nor does Merman give a proof of this, but, by reference to a theorem of Birkhoff [1], shows that the probability of these two motions belonging to the same class is different from zero.

An approach of three bodies coming from infinity along independent trajectories implies $h > 0$. Therefore, in this case, the problem of the occurrence of capture is equivalent to the following one: do there exist motions which are hyperbolic for $t \to -\infty$ and hyperbolic-elliptic for $t \to +\infty$? In 1932 Chazy [4] discussed also this case and gave a negative answer to the question. However, instead of giving a rigorous proof, he made use of certain conjectures, the truth of which has neither been proved nor disproved, and of conclusions by analogy. Therefore Chazy's investigations for $h > 0$ left open the question as to whether a capture takes place or not in the general three-body problem.

3. Shmidt's numerical example

In 1947 Shmidt [1] produced a first numerical example to illustrate the possibility of capture for $h > 0$ in the general three-body problem. He assumed that the three masses are all equal to unity, the mass of the Sun. The unit of distance was chosen equal to the astronomical unit, and the unit of time equal to the mean solar year divided by 2π. The gravitational constant is then one, a fact which simplifies calculations. In the example of Shmidt the osculating orbit of P_1 with respect to P_0 changed from hyperbolic to elliptic whereas the orbit of the third body P_2 remained hyperbolic during the whole motion. The original calculations were carried out at the Geophysical Institute of the Academy of Science of the U.S.S.R. under the supervision of Pariiskiĭ [1] for the time interval from $t = -129764$ to $t = 8000$. In a series of papers since 1947, Sizova [1], Sitnikov [1], Merman [1], Proskurin [1], Hrapovickaya [1], and Kochina [1, 2] have constructed numerical examples of capture in the restricted and in the general three-body problem, using numerical integration and criteria established by Hil′mi [1] and Merman [1–8].

HIL′MI'S AND MERMAN'S INVESTIGATIONS

4. Hil′mi's three theorems

From the mathematical point of view the problem of capture was first considered by Hil′mi [1]. Assume that the state of motion of a system of three particles P_i $(i = 0, 1, 2)$, subject to the mutual Newtonian attractions, is given at the initial instant $t = 0$. Denote by m_i $(i = 0, 1, 2)$ the masses of the three particles respectively and by $r_{jk}(t)$ $(j \neq k; j = k = 0, 1, 2)$ their mutual distances at the instant t.

Definition 2. We shall say that a capture takes place between P_0 and P_1 if (i) all three mutual distances $r_{jk}(t) \to \infty$ $(j \neq k)$ as $t \to -\infty$, and (ii) there exists an instant T and a positive constant R such that $0 < r_{10} < R$ for all $t > T$.

Notations. Let

$$r(t) = \min\,[r_{jk}(t)], \qquad r'(t) = \min\left[\frac{dr_{jk}(t)}{dt}\right] \qquad (j \neq k)$$

$$M' = \min\,(m_0m_1 + m_1m_2,\quad m_1m_2 + m_0m_2,\quad m_0m_1 + m_0m_2)$$

$$M'' = \max\left(\frac{m_0m_1}{m_0 + m_1},\ \frac{m_1m_2}{m_1 + m_2},\ \frac{m_0m_2}{m_0 + m_2}\right), \qquad M^* = \frac{M'}{M''}$$

Hil′mi [1–2] stated the following two theorems.

Theorem 1. If there exists a time t_1 such that

$$(A) \qquad r'(t_1) < 0, \qquad r'^2(t_1) > 8M^*/r(t_1)$$

then all three mutual distances increase indefinitely as $t \to -\infty$ (hyperbolic motion).

This theorem states that up to the time t_1 capture has occurred for no pair of the three bodies and, consequently, their motions have been independent.

Theorem 2. Let $\rho(t)$ denote the distance between the third particle P_2 and the center of mass of P_0 and P_1, and let

$$M = m_0 + m_1 + m_2, \qquad \mu = \frac{m_2(m_0 + m_1)}{M}, \qquad m = \frac{m_0 m_1}{m_0 + m_1}$$

If the energy constant h of the system is greater than zero and there exists a time t_2 and two positive constants R and $\epsilon < R$ such that

$$(B) \qquad r_{01}(t_2) < R, \qquad \rho(t_2) > 2R, \qquad \rho'(t_2) > 0$$

$$\rho'^2(t_2) - \frac{16M}{\rho(t_2)} > \frac{2}{\mu} h + \frac{2m(m_0 + m_1)}{\mu(R - \epsilon)}$$

then the two distances r_{02} and r_{12} increase indefinitely as $t \to \infty$, whereas the third distance r_{01} remains less than R (bounded) for $t > t_2$ (hyperbolic-elliptic motion).

This theorem states that, if capture between the two particles P_0 and P_1 has occurred prior to t_2, then in the future it cannot be dissolved.

Definition 3. We shall say that capture between the two particles P_0 and P_1 takes place in the interval (t_1, t_2) if at the instants t_1, t_2 the conditions of theorems 1 and 2 are satisfied respectively.

It is easy to see that capture in the sense of this definition is a particular case of that given by definition 2.

Shmidt's numerical example satisfies the assumptions of theorem 1 for $t = -129764$ and those of theorem 2 for $t = 8000$. Hence this example illustrates capture not only in a weakened form (definition 1) or in a special form (definition 3) but for the whole time interval between $-\infty$ and $+\infty$ (definition 2).

Corollary. Capture in the general three-body problem is possible.

Theorem 3. In the twelve-dimensional manifold of states of motion E_{12} of the three bodies the measure of the set Φ of points which represent initial conditions leading to capture is positive.

The proof of this fundamental theorem runs as follows. Introduce Jacobian coordinates and consider the twelve-dimensional Euclidean space E_{12} with coordinates x_0, y_0, z_0, x_0', y_0', z_0', ξ_0, η_0, ζ_0, ξ_0', η_0', ζ_0', the mutual distances being well-defined functions of these variables. Let E and E' denote two sets of points in E_{12} the coordinates of which satisfy the conditions (A) and (B) respectively. The Shmidt numerical example of motion is represented in the phase space E_{12} by an arc of a curve, beginning at a point in E and ending at a point in E'. Because the conditions (A) and (B) are given in the form of inequalities, both sets E and E' are open. Further, from the existence of a single arc of the curve joining corresponding points of E and E', follows (by a well-known property of differential equations) the existence of a tube of such arcs the initial points of which in E form an open subset Φ. Since the measure of every open set is positive, it follows that the measure of Φ is greater than zero. This completes the proof of the theorem.

5. Merman's criteria

Hil'mi's sufficient conditions for a hyperbolic-elliptic motion require (i) a sufficiently large initial radial velocity of the body P_2, receding to infinity along a hyperbola, and (ii) an initial configuration of the three bodies, stretched in the direction of P_2. In the case of a hyperbolic motion, the sufficient conditions of Hil'mi require large values for all three radial velocities.

Merman [2] has given new criteria for hyperbolic and hyperbolic-elliptic motion in which the above-mentioned restrictions on the radial velocities and on the initial configuration of the three bodies are removed. Instead it is required that the corresponding full velocities be sufficiently large, a condition which is to be expected for the occurrence of these two types of motion.

Let m_i denote the mass of P_i ($i = 0, 1, 2$), $\mathbf{r}_\nu$ the vector from $P_{\nu+1}$ to $P_{\nu+2}$ (indices, differing modulo three are considered as equivalent), $\mathbf{r} = \mathbf{r}_2$, $\boldsymbol{\rho}$ the vector from the center of mass G of P_0 and P_1 to P_2, and let

$$M = \Sigma m_i, \qquad \lambda = \frac{m_1}{m_0 + m_1}, \qquad \mu = \frac{m_0}{m_0 + m_1}$$

The differential equations of motion are then

$$\ddot{\mathbf{r}}_\nu + \frac{m_{\nu+1} + m_{\nu+2}}{r_\nu^{\,3}}\mathbf{r}_\nu = m_\nu\left(\frac{\mathbf{r}_{\nu+1}}{r_{\nu+1}^{\,3}} + \frac{\mathbf{r}_{\nu+2}}{r_{\nu+2}^{\,3}}\right) \qquad (\nu = 0, 1, 2)$$

$$\ddot{\boldsymbol{\rho}} = -M\left(\lambda\frac{\mathbf{r}_0}{r_0^{\,3}} - \mu\frac{\mathbf{r}_1}{r_1^{\,3}}\right)$$

Under the assumption that a collision does not occur during the motion, Merman proved the following two criteria.

A criterion for hyperbolic-elliptic motion. If at the initial instant $t = 0$ there exist three constants R, $a > \max(\lambda, \mu)$, and ϵ such that

$$\rho(0) < a_0 r_0(0), \qquad \rho(0) < a_1 r_1(0), \qquad \frac{1}{2}[|\dot{\boldsymbol{\rho}}(0)| + \dot{\rho}(0)] > \epsilon$$

$$\frac{A \arctan (k'/k)}{k'[|\dot{\boldsymbol{\rho}}(0)| - \epsilon]\rho(0)} \leq \epsilon$$

$$\frac{m_0 + m_1}{r(0)} - \frac{\dot{r}^2(0)}{2} \geq \frac{m_0 + m_1}{R} + \frac{B}{L\rho^2(0)}$$

$$\rho(0) \geq ar(0) \qquad \text{if} \qquad \dot{\rho}(0) - \epsilon \geq aD$$

$$a \sqrt{\left\{\frac{r(0)}{\rho(0)}[|\dot{\boldsymbol{\rho}}(0)| - \epsilon] - kD\right\}^2 + k'^2D^2} < k'[|\dot{\boldsymbol{\rho}}(0)| - \epsilon]$$

if

$$\dot{\rho}(0) - \epsilon < aD$$

where

$$L = \tfrac{1}{2}[|\dot{\boldsymbol{\rho}}(0)| + \dot{\rho}(0)] - \epsilon, \qquad k = \frac{\dot{\rho}(0) - \epsilon}{|\dot{\boldsymbol{\rho}}(0)| - \epsilon}, \qquad k^2 + k'^2 = 1$$

$$a_0 \geq \frac{1}{1 - (\mu/a)}, \qquad a_1 \geq \frac{1}{1 - (\lambda/a)}$$

$$A = M \max (\lambda a_0^2 + \mu, \lambda + \mu a_1^2), \qquad D = \sqrt{\dot{r}^2(0) + \frac{2B}{L\rho^2(0)}}$$

$$2B = m_2\sqrt{2(m_0 + m_1)R} \max [(1 + \mu)a_0^3 + a_0^2 + a_0 + \lambda, \\ (1 + \lambda)a_1^3 + a_1^2 + a_1 + \mu]$$

then the motion is hyperbolic-elliptic for $t > 0$, i.e., $\rho \to \infty$ as $t \to \infty$, whereas $r < R$ for all $t \geq 0$.

First criterion for hyperbolic motion. If at the initial instant $t = 0$ there exist three constants $\epsilon_v > 0$ such that the following conditions hold,

$$\tfrac{1}{2}[|\dot{\mathbf{r}}_v(0)| + \dot{r}_v(0)] \geq \epsilon_v \qquad (v = 0, 1, 2)$$

$$\frac{(m_{v+1} + m_{v+2}) \arctan (k_v'/k_v)}{k_v'[|\dot{\mathbf{r}}_v(0)| - \epsilon_v]r_v(0)} + \frac{m_v \arctan (k_{v+1}'/k_{v+1})}{k_{v+1}'[|\dot{\mathbf{r}}_{v+1}(0)| - \epsilon_{v+1}]r_{v+1}(0)}$$
$$+ \frac{m_v \arctan (k_{v+2}'/k_{v+2})}{k_{v+2}'[|\dot{\mathbf{r}}_{v+2}(0)| - \epsilon_{v+2}]r_{v+2}(0)} \leq \epsilon_v \qquad (v = 0, 1, 2)$$

where

$$k_\nu = \frac{\dot{r}_\nu(0) - \epsilon_\nu}{|\dot{\mathbf{r}}_\nu(0)| - \epsilon_\nu}, \qquad k_\nu^2 + k_\nu'^2 = 1 \qquad (\nu = 0, 1, 2)$$

then the motion is hyperbolic for $t \to \infty$.

Recently Merman [6] has given a new criterion for hyperbolic motion in the three-body problem. This criterion is applicable if the relative velocity of some pair of the bodies is not large and the third body is at a considerable distance from that pair or recedes from it sufficiently rapidly.

Second criterion for hyperbolic motion. If at the initial instant $t = 0$ there exist three positive constants $a > \max(\lambda, \mu)$, v, and ϵ such that

$$\rho(0) > ar(0), \qquad \dot{\rho}(0) - \epsilon > av, \qquad \tfrac{1}{2}[|\dot{\boldsymbol{\rho}}(0)| + \dot{\rho}(0)] > \epsilon$$

$$\frac{A \arctan k'/k}{k'[|\dot{\boldsymbol{\rho}}(0)| - \epsilon]\rho(0)} \leq \epsilon$$

$$\frac{\dot{r}^2(0)}{2} - \frac{m_0 + m_1}{r(0)} > \frac{C}{L\rho(0)}, \qquad \frac{v^2}{2} \geq \frac{\dot{r}^2(0)}{2} + \frac{C}{L\rho(0)}$$

where

$$A = M \max(\lambda a_0^2 + \mu, \lambda + \mu a_1^2), \qquad k = \frac{\dot{\rho}(0) - \epsilon}{|\dot{\boldsymbol{\rho}}(0)| - \epsilon}, \qquad k^2 + k'^2 = 1$$

$$a_0 = \frac{1}{1 - (\mu/a)}, \qquad a_1 = \frac{1}{1 - (\lambda/a)}, \qquad L = \frac{1}{2}[|\dot{\boldsymbol{\rho}}(0)| + \dot{\rho}(0)] - \epsilon$$

$$2C = m_2 v \max[(1 + \mu)a_0^3 + a_0^2 + a_0 + \lambda, (1 + \lambda)a_1^3 + a_1^2 + a_1 + \mu] \times \left[\frac{r(0)}{\rho(0)} + \frac{v}{|\dot{\boldsymbol{\rho}}(0)| - \epsilon}\right]$$

then the motion is hyperbolic for $t \to \infty$.

The first two criteria applied to Shmidt's example have considerably shortened the interval of numerical integration (the first criterion is satisfied at $t = -1364$ and the second at $t = -896$, capture taking place around $t = -1250$) as compared with the interval between $t = -129764$ and $t = 8000$ in Hil'mi's theorems 1 and 2.

6. Capture in the three-body problem

The Shmidt and Hil'mi proof of the occurrence of capture in the three-body problem is a combination of a mathematically rigorous method and a method of numerical integration. In applying the Hil'mi criteria to the ends of a trajectory calculated by numerical integration, one needs (i) to estimate the error committed over the interval of integration, and (ii) to show that, to within this error, the criteria of Hil'mi still hold at the end points of the trajectory. Finally, the domain of the initial data leading to a capture is to be determined, or, in other words, sufficient conditions for capture have to be found.

These problems have been solved by Merman [3], first in the case of the restricted hyperbolic problem of the three bodies (when the mass m_2 of P_2 is supposed to be zero). Since $m_2 = 0$, the particle P_2 does not influence the relative motion of P_0 and P_1 which is hyperbolic, and the error committed over the interval of numerical integration has to be estimated only for P_2. Merman replaces the actual problem by a simplified scheme in which the actual motion of P_2 is replaced by a rectilinear and uniform motion of this body with respect to the center of mass G of P_0 and P_1. His method is essentially a somewhat modified Hil'mi's method. Using the same method, Merman has derived a great number of criteria, more or less convenient in various cases, sufficient for capture in the restricted hyperbolic problem of three bodies. A common feature of these sufficient criteria is that the interval of numerical integration needed for their application is several times shorter than that given by Shmidt and Hil'mi.

In addition to these sufficient criteria, Merman has also given several necessary criteria for capture in the above-mentioned particular case of the three-body problem. These last criteria exhibit one common feature, namely, that a capture can take place only in the case of a close approach of P_0 and P_1.

Sufficient conditions of capture for the general three-body problem have also been given by Merman [4, 7]. In this case, as compared with the previously mentioned restricted hyperbolic case, the deviation of the actual relative orbit of the approaching bodies P_0 and P_1 from the hyperbola in the simplified scheme has to be estimated. Merman [7] proceeds as follows. Consider the motion of the three particles with the following initial conditions: (i) the mutual distance r_{01} between P_0 and P_1 is small in comparison with the other two distances r_{02} and r_{12}; (ii) the relative velocities of P_0 and P_1, P_2 and P_1 are hyperbolic whereas the relative velocity of P_2 and P_0 is close to a parabolic one. Any motion with such initial conditions is called a close approach between P_0 and P_1. In any domain of such a close approach the receding body P_2 has little influence on the relative motion of P_0 and P_1, and the influence of the latter, in turn, is negligible on the motion of P_2 with respect to the center of mass G of P_0 and P_1. The motion of the three particles in the early stages will thus differ little from the motion of the same three particles in a simplified scheme (I), in which P_1 moves relative to P_0 in a hyperbola, but the motion of P_2 relative to G is uniform and rectilinear.

The relative speed $|\dot{\mathbf{r}}_{10}|$ being great, the branch of the hyperbola described by the radius vector $\mathbf{r}_{10}$ rapidly approaches its asymptote, and the approach between P_0 and P_1 ceases to be close. As the motion goes on, the orbit described by the radius vector $\mathbf{r}_{20}$ will change its form into an elliptic or a hyperbolic orbit. It is therefore assumed that in the intermediate stages the motion proceeds either according to a simplified scheme (II), in which the motion of P_2 with respect to G continues to be uniform and rectilinear, but the relative orbit of $\mathbf{r}_{20}$ is a circle $r_{20} =$ const; or according to the scheme (III), in which the relative motions of all three particles are uniform and rectilinear.

In the later stages of motion criteria are applied which guarantee the preservation of the one or the other type of motion at infinity.

The principal problem is to estimate the deviations of the true motions of the three particles from those in the three above-mentioned simplified schemes. Merman solves this problem by an immediate generalization to the general three-body problem of the methods developed by him for the restricted hyperbolic three-body problem. This generalized method for obtaining bounds for the coordinates and the velocities of the three particles for a given time interval $(0, \tau)$, their values being known at $t = 0$, provides at the same time sufficient conditions for capture in the general three-body problem. The method outlined is illustrated by the numerical example of Shmidt cited before. Since the conditions of capture are in the form of inequalities, the corresponding domain of the initial conditions of motion constitute an open set. That this set is not vacuous is shown in the paper by Merman and Kochina [8]. The previously mentioned paper by Merman [7] together with this joint paper of the two authors gives the first rigorous proof that the probability of capture in the general three-body problem is positive.

It is relatively easy to establish necessary and sufficient conditions for capture in the simplified scheme of the restricted hyperbolic three-body problem considered before. In the exact hyperbolic case of the restricted three-body problem the necessary and sufficient conditions differ little from those obtained for the approximate scheme, since all the coordinates and velocities deviate little from the corresponding quantities in the approximate scheme for the range of close approaches between P_0 and P_1.

The problem of finding necessary and sufficient conditions for capture in the general three-body problem is difficult and, as yet, has not been solved.

BIBLIOGRAPHY

Birkhoff, G. D. [1], *Dynamical Systems*, American Mathematical Society Colloquium Publications, **9**, chapter 9, New York (1927).

Chazy, J. [1], Sur certaines trajectoires du problème des n corps, *Bull. Astr.*, **35**, 321–389 (1918).

——— [2], Sur l'allure du mouvement dans le problème des trois corps quand le temps croit indéfiniment, *Ann. École Norm. Sup.* (3), **39**, 29–130 (1922).

——— [3], Sur l'allure finale du mouvement dans le problème des trois corps, *J. Math. Pures Appl.* (9), **8**, 353–380 (1929).

——— [4], Sur l'allure finale du mouvement dans le problème des trois corps, *Bull. Astr.* (2), **8**, 403–436 (1932).

Fesenkov, V. G. [1], On the possibility of capture at close passages of attracting bodies (Russian), *Astr. Zhurn.*, **23**, 45–48 (1946).

Gazaryan, Yu. L. [1], On Chazy's proof of the impossibility of capture in the three-body problem (Russian), *Moskov. Gos. Univ. Soobshch. Astr. Inst.*, no. 92, 23–45 (1953).

Hil'mi, G. F. [1], On the possibility of capture in the problem of three bodies (Russian), *Dokl. Akad. Nauk S.S.S.R.* (N.S.), **62**, 39–42 (1948).

——— [2], *The problem of* n *bodies in celestial mechanics* (Russian), chapter 3, Izdat. Akad. Nauk S.S.S.R., Moscow (1951), 155 pp.

Hopf, E. [1], Zwei Sätze über den wahrscheinlichen Verlauf der Bewegungen dynamischer Systeme, *Math. Ann.*, **103**, 710–719 (1930).

Hrapovickaya, G. E. [1], On a case of capture in the problem of three bodies (Russian), *Akad. Nauk S.S.S.R. Byull. Inst. Teor. Astr.*, **5**, 435–444 (1953).

Kochina, N. G. [1], Examples of hyperbolic and hyperbolic-elliptic motion in the restricted hyperbolic problem of three bodies (Russian), *Akad. Nauk S.S.S.R. Byull. Inst. Teor. Astr.*, **5**, 445–454 (1953).

——— [2], Examples of motion in the restricted parabolic problem of three bodies (Russian), *Akad. Nauk S.S.S.R. Byull. Inst. Teor. Astr.*, **5**, 617–622 (1954).

Merman, G. A. [1], An example of capture in the plane restricted hyperbolic problem of three bodies (Russian), *Akad. Nauk S.S.S.R. Byull. Inst. Teor. Astr.*, **5**, 373–391 (1953).

Merman, G. A. [2], New Criteria of hyperbolic and hyperbolic-elliptic motion in the problem of three bodies (Russian), *Astr. Zhurn.*, **30**, 332–339 (1953).

——— [3*a*], On sufficient conditions of capture in the restricted hyperbolic problem of three bodies with close double approaches (Russian), *Akad. Nauk S.S.S.R. Byull. Inst. Teor. Astr.*, **5**, 325–372 (1953).

——— [3*b*], Restricted parabolic problem of three bodies (Russian), *Akad. Nauk S.S.S.R. Byull. Inst. Teor. Astr.*, **5**, 606–616 (1954).

——— [4], On sufficient conditions for capture in the three-body problem (Russian), *Dokl. Akad. Nauk S.S.S.R.* (N.S.), **99**, 925–927 (1954).

——— [5], On Chazy's investigations in the problem of three bodies, *Akad. Nauk S.S.S.R. Byull. Inst. Teor. Astr.*, **5**, 594–605 (1954).

——— [6], A new criterion for hyperbolic motion in the problem of three bodies (Russian), *Akad. Nauk S.S.S.R. Byull. Inst. Teor. Astr.*, **6**, 69–72 (1955).

——— [7], Hyperbolic approaches in the problem of three bodies (Russian), *Akad. Nauk S S.S.R. Byull. Inst. Teor. Astr.*, **6**, 73–84 (1955).

Merman, G. A. and Kochina, N. G. [8], Application of a method of estimates to O. Yu. Shmidt's example (Russian), *Akad. Nauk S.S.S.R. Byull. Inst. Teor. Astr.*, **6**, 85–111 (1955).

Pariiskiĭ, N. N. [1], Computation of a trajectory of capture in the plane problem of three bodies of equal mass (Russian), *Vopr. Kosmog.*, **3**, 33–62 (1954).

Proskurin, V. F. [1], On capture in the example of O. Yu. Shmidt (Russian), *Akad. Nauk S.S.S.R. Byull. Inst. Teor. Astr.*, **5**, 429–434 (1953).

Shmidt, O. Yu. [1], On a possibility of capture in celestial mechanics (Russian), *Dokl. Akad. Nauk S.S.S.R.* (N.S.), **58**, 213–216 (1947).

Sitnikov, K. A. [1], On the possibility of capture in the problem of three bodies (Russian), *Mat. Sborn.* (N.S.), **32** (74), 693–705 (1953).

Sizova, O. A. [1], On the possibility of capture in the restricted problem of three bodies (Russian), *Dokl. Akad. Nauk S.S.S.R.* (N.S.), **86**, 485–488 (1952).

Zeipel, H. von [1], *Bull. Astr.*, **22**, 449–490 (1905).

GENERALIZED n-BODY AND THREE-BODY PROBLEMS

A GENERALIZED n-BODY PROBLEM

1. Statement of the problem

A generalized n-body problem, obtained by replacing the Newtonian law of gravitation by a more general law, has been the subject of study by Sokolov [3, 4] since 1934. He considers a system of n (≥ 3) particles P_i of masses m_i ($i = 1, 2, \cdots, n$), which attract or repel each other, the interaction between P_i and P_j ($i \neq j$) having magnitude

$$m_i m_j |f(r_{ij})| \tag{1}$$

and representing an attraction or repulsion according to whether $f(r)$ is negative or positive. It is assumed that $f(r) = dF(r)/dr$ is real and analytic for every real and positive r and may have singularities at the points $r = 0$ and $r = \infty$ on the real r-axis.

If, in particular, $f(r) = Cr^{-(2\alpha+1)}$, C being a constant, then, for $\alpha = \frac{1}{2}$, we have Newton's inverse square law, and for $\alpha = 1$ Jacobi's inverse cube law.

Consider an inertial coordinate system and denote by $\mathbf{x}_i$ the coordinate vector of the point P_i with respect to this system. Then the differential equations of motion are

$$m_i \mathbf{x}_i'' = U_{\mathbf{x}_i} \qquad (i = 1, 2, \cdots, n) \tag{2}$$

where

$$U = \sum_1^n m_j m_k \, F(r_{jk}), \tag{3}$$

$$r_{jk} = |\mathbf{x}_j - \mathbf{x}_k| \qquad (j \neq k), \qquad ' = d/dt \tag{4}$$

and $U_{\mathbf{x}_i}$ denotes the $\mathbf{x}_i$ gradient of the scalar U.

The system (2) admits the energy integral

$$T - U = h \tag{5}$$

where

$$T = \frac{1}{2} \sum_{1}^{n} m_i x_i'^2$$

and h is a constant of integration.

If at the initial instant $t = 0$ the coordinates $x_{i\nu}(0)$ and the velocity components $x_{i\nu}'(0)$ ($i = 1, 2, \cdots, n$; $\nu = 1, 2, 3$) of the n particles are real and finite and none of the mutual distances r_{jk} are zero, the right-hand sides of the system (2) are analytic functions in certain neighborhoods of these initial values. According to the classical existence theorem, there exists a unique solution $x_{i\nu}(t)$, $x_{i\nu}'(t)$, analytic in a neighborhood of $t = 0$, satisfying the system (2), and assuming for $t = 0$ the given initial values $x_{i\nu}(0)$, $x_{i\nu}'(0)$.

2. Regular motion

A position of the system of n particles is said to be regular if in a neighborhood of the values $\mathbf{x}_i^0$ of the n vectors $\mathbf{x}_i$ all the functions $U_{\mathbf{x}_i}$ are analytic. Because of the assumption made relative to $f(r)$ this means that every position of the system corresponding to real, finite, and positive values of r_{jk} will be regular. A motion of the system of n particles will be said to be regular at an instant if its position at that instant is regular, and the velocities of the n particles are finite.

A lower bound for the radii of convergence of the expansions for the coordinates of the n particles about an instant t in the regular motion has been determined by Sokolov. In the particular case of three bodies under Newtonian attraction the value of this lower bound obtained is approximately one and a half times greater than the value given by the corresponding formulas of Sundman [1] and Mendes [1].

Only real motions of the system of n particles are considered, i.e., motions for which the coordinates and the components of the velocities of the particles are real for real values of t. If a motion, being regular at $t = 0$, is continued analytically along the real t-axis, then two cases are possible: the motion is regular for all values of t between $-\infty$ and $+\infty$, or it ceases to be regular at a certain point $t = t_1$.

In the first case the coordinates of the particles, being analytic functions for all real t, are expansible in series of polynomials, uniformly convergent in every interval of the real t-axis. If, however, these coordinates are analytic functions in a strip of the complex t-plane:

$$-T_0 < \operatorname{Im} t < +T_0$$

they can be expanded, as is well known, in series of integral and positive powers of

$$\tau = (e^{\pi t/2T_o} - 1)/(e^{\pi t/2T_o} + 1)$$

3. Singular points

In this section certain theorems on the behavior of a system of n particles, interacting according to the law (1), in a neighborhood of a singular point $t = t_1$ will be given. These theorems are partly generalizations and extensions of the corresponding theorems of Painlevé, proved for the classical n-body problem, when $F(r) = 1/r$.

Theorem 1. If the motion of the n particles is regular up to the instant $t = t_1$ but ceases to be regular at this instant, then min $(\underline{r}, 1/\bar{r})$ approaches zero as $t \to t_1$.

The quantities $\bar{r}(t)$ and $\underline{r}(t)$ denote respectively the largest and the smallest of the mutual distances of the particles at the instant t. In particular, we have the following theorem.

Theorem 2. Let $f(r)$, analytic for $-\delta < \arg r < +\delta$, be of bounded modulus for $|r| > d > 0$, and let $F(r)$ be bounded above as r varies from $r = d > 0$ to $+\infty$. Then, if a motion of the n particles is regular up to the instant t_1 but ceases to be regular at this instant, the smallest of the mutual distances of the particles approaches zero as $t \to t_1$.

The assumptions of this theorem are satisfied, for example, for $F(r) = 1/r^{2\alpha}$, $\alpha > 0$. A particular case of this theorem is the corresponding theorem of Painlevé for the classical n-body problem.

The assumptions of theorem 2 are not necessary, and $\lim \underline{r}(t) = 0$, as $t \to t_1$, holds also under the following assumptions given in the next theorem.

Theorem 3. Let

$$J^2 = \sum_1^n m_i x_i^2 = \frac{1}{M} \sum_1^n m_j m_k \, r_{jk}^2 \qquad (j \neq k)$$

be the moment of inertia of the system about its center of mass, assumed to be fixed and at the origin, and $M = \sum_1^n m_i$ be the total mass of the system. If for sufficiently large values of J^2 the ratio U/J^2 remains bounded above, i.e.,

$$U/J^2 < A \qquad \text{for} \qquad J \geq J_0 > 0$$

then

$$\lim_{t \to t_1} \underline{r}(t) = 0$$

The assumptions of this theorem are satisfied, for example, in the case $F(r) = r^{2\beta}(0 < \beta < 1)$, since for $\bar{r} > 1$ $(J > m/M^{1/2})$ we have

$$U/J^2 < \frac{M}{m^2} \Sigma m_j m_k$$

Theorem 4. Let the coordinates of the n particles be analytic functions of t for all t from $t = 0$ up to $t = t_1$, the last value being excluded, and let $f(r)/r$ be an analytic function of r^2 at the point $r^2 = 0$. Then, as $t \to t_1$, the largest of the mutual distances of the particles increases indefinitely, i.e.,

$$\lim_{t \to t_1} \bar{r}(t) = +\infty$$

4. Nature of singularities. Collisions and "approches"

Let $\lim \underline{r}(t) = 0$ as $t \to t_1$. Then it is possible that at least one of the mutual distances actually approaches zero. In such a case we say that there is at least a collision of two particles. This, however, does not mean that a simultaneous collision of two or more particles at $t = t_1$ is the only singularity possible. It is conceivable that $\lim \underline{r}(t) = 0$ as $t \to t_1$, but the pairs of closest particles change all the time so that the system of n particles does not approach any definite limit position.

If, for example, $n = 3$ and $F(r) = 1/r^{2\alpha}$, $\alpha > 0$, singularities of the last type do not exist on the real t-axis. In particular, in the case of the classical three-body problem $(\alpha = \frac{1}{2})$, Painlevé showed that in a neighborhood of a binary collision one and the same of the three mutual distances remains the smallest and, ultimately, approaches zero. Consequently, in the case of the classical three-body problem, binary collisions and triple collisions are the only singularities of motion which are possible.

For the classical n-body problem $(n > 3)$ Painlevé proved the following theorem.

Theorem 5. As $t \to t_1$, all particles of the system approach limit positions at finite distances, or there exist at least four particles of the system which do not approach any such limit positions. These exceptional particles are such that the minimum of their mutual distances approaches zero, with $t - t_1$, without any of these distances steadily approaching zero.

It has never been proved, however, that singularities of this last type actually occur.

In connection with binary collisions, there arises the phenomenon of an "approche." Lévy [1] defines an "approche" in the classical three-body problem as a motion in which one of the mutual distances $r(t)$ satisfies the conditions

$$\liminf_{t \to \infty} r(t) = 0, \qquad \limsup_{t \to \infty} r(t) > 0$$

Motions with binary collision and motions with "approche" are closely related. They are, so to say, if not adjacent, then at least complementary. For instance, two bodies which do not approach each other as t increases indefinitely, a collision at $t = \infty$ being impossible, would exhibit the phenomenon of an "approche."

According to Chazy the types of motions of the three bodies for large t can be subdivided into two large categories (see Chapter 2, Art. 2 of this section): those in which at least two mutual distances become infinite with t (hyperbolic-elliptic, hyperbolic-parabolic, and hyperbolic motions, *mouvements à allure binaire*), and those for which the mutual distances remain finite or oscillate between finite values and infinity (bounded and oscillating motions, *mouvements bornés et oscillants*). Lévy proves the existence of motions with "approche" in both of these categories of motions. In the first category, planar as well as spatial motions with "approche" exist in the hyperbolic-elliptic region only. The spatial motions with "approche" have to satisfy three initial conditions, whereas the planar motions only one condition. The motions leading to an "approche" depend upon seven parameters in the plane and upon nine parameters in space. It should be pointed out, however, that in space there are two categories of motions with "approche," each depending upon nine parameters, one of these categories consisting of planar motions only.

In comparison with our knowledge about the asymptotic behavior of motions "*à allure binaire*," very little is known about bounded and oscillating motions, although the latter are of the utmost interest in the planetary theories. According to the results accumulated for bounded and oscillating motions, these motions possess "*stabilité à la Poisson*," except certain motions which are called exceptional. These last motions, when represented in the twelve-dimensional manifold of states of motion E_{12} (see Chapter 2, Art. 2 of this section), do not fill up a volume of nonzero measure.

In order to investigate motions with "approche" in the category of bounded and oscillating motions, Lévy distinguishes two cases: (i) all motions with binary collision that are nonexceptional, and (ii) all such motions that are exceptional. It is extremely probable that this second assumption cannot be realized. In the first case, however, all motions with binary collision exhibit the phenomenon of an "approche." It is also conceivable that there are other motions with "approche," and that, under certain conditions, even all motions exhibit this phenomenon. These questions, however, remain unanswered at present.

5. The generalized Lagrange-Jacobi equality

Sokolov, following Sundman, uses the generalized Lagrange-Jacobi equality which is obtained as follows. Differentiate J^2 twice with respect to t and make use of the differential equations of motion (2) and the energy integral (5). Then there results the equality

$$\frac{1}{2}\frac{d^2J^2}{dt^2} = \Sigma\, m_j m_k\, F_1(r_{jk}) + 2h \tag{6}$$

where

$$F_1(r) = 2F(r) + rf(r)$$

If $F(r)$ is bounded above for $0 \leq r < +\infty$, then it follows from the energy integral (5) that also all the x_i' $(i = 1, 2, \cdots, n)$ are bounded. Consequently, as $t \to t_1$, all the vectors $\mathbf{x}_i$ and also the mutual distances between the particles will approach definite limits. We arrive at the same conclusion, using the differential equations of motion (2) under the assumption that $|f(r)|$ remains bounded for $0 \leq r < +\infty$. Hence, under the above-mentioned assumptions with respect to $F(r)$ or $f(r)$, we have J^2 approaching a definite limit $J_1{}^2$ as $t \to t_1$.

If $rf(r)$ or $F_1(r)$ are bounded below for $0 \leq r < +\infty$, then the right-hand side of (6), being a continuous function of t in the interval $0 \leq t < t_1$, will also be bounded below on the positive semiaxis of r. Consequently, as $t \to t_1$, dJ^2/dt will approach a definite finite limit or it will increase indefinitely.* Then also J^2 will approach a definite limit $J_1{}^2 \geq 0$, or it will increase indefinitely.

If $F_1(r)$ is bounded above for $0 \leq r < +\infty$, then in a similar way it follows from (6) that dJ^2/dt approaches a finite limit as $t \to t_1$ or $dJ^2/dt \to -\infty$, and, consequently, $J^2 \to J_1{}^2 \geq 0$.

If the potential U is a homogeneous function of degree two in the coordinates of the particles, then instead of (6) we have the Jacobi equation

$$\frac{1}{2}\frac{d^2J^2}{dt^2} = 2(\beta + 1)U + 2h \tag{7}$$

Hence, for $\beta = -1$, there exists an integral of motion

$$J^2 = 2ht^2 + 2J_0J_0' + J_0{}^2$$

where J_0, and J_0' are constants of integration.

* If the derivative $f'(x)$ of a function $f(x)$ remains bounded as $x \to x_1(\neq \infty)$, then, as $x \to x_1$, $f(x)$ tends to a finite limit.

Sokolov draws attention to an erroneous conjecture made by Jacobi [1], which has been repeated by Birkhoff [1] and Chandrasekhar [1]. Namely, if the relation

$$\lim_{t \to +\infty} J^2 = +\infty$$

holds, the above-mentioned authors draw the conclusion that at least one of the particles would then recede infinitely far from their common center of mass, whereas, in fact, one may only infer that $\lim_{t \to +\infty} \bar{r}(t) = +\infty$. Further Jacobi concludes that, if $\lim_{t \to t_1} U = +\infty$, then, at the instant t_1, there is at least a collision between two particles, whereas one can only infer that $\lim \underline{r}(t) = 0$ as $t \to t_1$. In the case of collision between two particles at $t = t_1$, Jacobi assumes that

$$\int_0^{t_1} dt \int_0^t U\, dt$$

is infinite, whereas U is of the order $(t_1 - t)^{-\frac{2}{3}}$.

If the motion remains regular in the whole interval from $t = 0$ to $t = +\infty$, then from (6) some simple conclusions as to the behavior of J^2 as $t \to +\infty$, and, consequently, on the stability of the system under consideration can be made.

Consider, for instance, the three above-mentioned cases:

$$\text{(8)} \qquad \begin{aligned} &\text{(i)} \quad rf(r) \geq A \\ &\text{(ii)} \quad F_1(r) \geq A_1 \\ &\text{(iii)} \quad F_1(r) \leq A_2 \end{aligned}$$

where A, A_1, and A_2 are constants.

Under the assumption (i) for $A > 0$ and also under the assumption (ii) for $A_1 > -2h/\Sigma m_j m_k$ the derivative dJ^2/dt steadily increases as $t \to +\infty$. Hence, beginning with some instant, J^2 will also steadily increase.

If, however, $A = 0$ or $A_1 = -2h/\Sigma m_j m_k$, then dJ^2/dt monotonically increases and $J^2 \to {J_1}^2 \geq 0$ (monotonically decreasing beginning with some instant) or $J^2 \to +\infty$ (monotonically increasing beginning with some instant).

If (iii) holds for $A_2 = -2h/\Sigma m_j m_k$, then dJ^2/dt monotonically decreases and approaches a nonnegative limit. Then also $J^2 \to {J_1}^2 > 0$ or $J^2 \to +\infty$, monotonically increasing beginning with some instant.

If (iii) holds for $A_2 < -2h/\Sigma m_j m_k$, then it is impossible that the motion is regular up to $t = +\infty$.

Hence, if we assume that each of the assumptions (8) holds only in a neighborhood of $r = 0$, we arrive at the conclusion that an unlimited approach of all the particles, as $t \to +\infty$, is possible in the cases (i) and (ii) for $A \leq 0$ and $A_1 < -2h/\Sigma m_j m_k$ respectively, and in the case (iii) for $A_2 > -2h/\Sigma m_j m_k$.

If $\lim F_1(r) = +\infty$ as $r \to 0$ and $F_1(r)$ remains bounded below as $r \to +\infty$ (or $F_1(r) = 0$ and $h > 0$), the assumption $\lim_{t \to +\infty} \underline{r}(t) = 0$ leads to the conclusion that

$$\lim_{t \to +\infty} J^2 = +\infty$$

From this fact it is easy to make sure that in the case $n = 3$, as $t \to +\infty$, one and the same of the three mutual distances approaches zero while the other two distances increase indefinitely.

Generalizing certain results of Chazy, proved for the classical three-body problem, Sokolov has proved a theorem to the effect that if under the assumption

$$\lim_{r \to +0} r^{2\alpha+1} f(r) = -2\alpha < 0 \left[\lim_{r \to +0} r^{2\alpha} F(r) = 1 \right]$$

we have $r^\gamma \, |f(r)| \leq A \;\; (0 < d \leq r < +\infty, \; \gamma > 1)$, and, for $\alpha = +1$, $\lim r^\gamma F_1(r) = g \neq 0$ (or $F_1(r) = 0$ and $h \neq 0$), then in the generalized three-body problem the smallest of the mutual distances cannot approach zero as $t \to +\infty$.

This theorem implies that the hypotheses made by Painlevé [1, pp. 585–6] concerning the smallest of the three mutual distances in the classical three-body problem must be rejected.

A GENERALIZED THREE-BODY PROBLEM

In this section we shall discuss the singular trajectories of the generalized three-body problem ($n = 3$) in the following three cases:

$$\lim_{t \to t_1} J^2 = J_1{}^2 > 0, \qquad \lim_{t \to t_1} J^2 = 0, \qquad \lim_{t \to t_1} J^2 = +\infty$$

6. Trajectories of binary collision: $\lim_{t \to t_1} J^2 = J_1{}^2 > 0$

Sokolov discusses the singular trajectories of binary collision under the assumption

(9) $$\lim_{r \to +0} r^{2\alpha+1} f(r) = -2\alpha < 0 \left[\lim_{r \to +0} r^{2\alpha} F(r) = 1 \right]$$

Sundman [1] and Levi-Civita [1] developed analytic weapons sufficiently powerful to deal with the singularity of binary collision in the classical three-body problem. The method of Sokolov is essentially that of Levi-Civita, suitably generalized. He considers the relative motion of P_1 with respect to P_0 and that of P_2 with respect to the common center of mass of P_0 and P_1. Spherical coordinates of P_1 and P_2, and six suitably chosen expressions of them and their derivatives with respect to t, are taken for the new variables. The corresponding differential system of order twelve is given, and the behavior of the new variables in a neighborhood of a binary collision is investigated. The special character of the case $\alpha = 1$ is pointed out. Finally, the distance r between the two colliding particles is introduced as the independent variable and several transformed systems of equations of motion are obtained. The existence of trajectories of binary collision for $\alpha \neq 1$ is proved and methods for the construction of such solutions, valid in a neighborhood of the instant t_1 of collision, are given.

The trajectories of binary collision are characterized by two invariant relations. Two first-order partial-differential equations, satisfied by the two functions G_1, G_2 which occur in the above-mentioned conditions for a binary collision, are established. The case $F(r) = 1/r^{2\alpha}$ $(\alpha < 1)$ is discussed in detail, and the first terms of the expansions for G_1 and G_2 are given. For $\alpha = \frac{1}{2}$ the corresponding formulas of Levi-Civita and Bisconcini [1] for the classical three-body problem are obtained, and certain errors in the coefficients calculated by Bisconcini are noted and corrected.

A second method, based on the reduction of the equations of motion to two special forms, permit a more uniform treatment of the cases $\alpha = 1$ and $\alpha \neq 1$ (the first method ceases to be applicable for $\alpha = 1$).

Finally, Sokolov discusses trajectories of binary collisions for certain particular and limit cases of motion; for instance, motions having a plane or an axis of symmetry, planar motions, and the limit case of the restricted problem considered by Hill.

7. Trajectories of general collision: $\lim_{t \to t_1} J^2 = 0$

Sludskiĭ [1] and Weierstrass [1] conjectured that for a general collision of n particles, subject to the Newtonian attraction, it is necessary that the angular momentum of the particles be zero about every axis through their common center of mass. The truth of this conjecture was first proved by Sundman [1] and Chazy [1]. Generalizing the above-mentioned theorem, as we may now call this conjecture, Sokolov proved the following two theorems.

Theorem 6. Suppose that a system of n particles moving in a p-dimentional Euclidean space be given. Let $\lim_{t \to t_1} J^2 = 0$, and for $\lim_{t \to t_1} U = +\infty$ let the inequality

$$(2 - B)U + \Sigma r_{ij} \frac{\partial U}{\partial r_{ij}} \geq 0$$

be satisfied in a neighborhood of $r_{ij} = 0$, where $-U$, the potential energy of the system, depends only on the mutual distances of the particles, and B is a positive constant. Then, in the motion of the system referred to its center of mass, all the constants of the angular momentum of the particles are zero.

The vanishing of all the constants of the angular momentum of the system turns out to be also a necessary condition for a general collision of n particles, mutually interacting according to the law (1), if the assumption (9) holds and $\alpha < 1$. The truth of the conjecture of Sludskiĭ and Weierstrass follows from theorem 6 for $p = 3$ and $U = \Sigma m_i m_j / r_{ij}$.

Theorem 7. If all the constants of the angular momentum of a system of n particles in their motion in a p-dimensional space ($p \geq n$) referred to their common center of mass are zero, the motion takes place in a fixed hyperplane of $(n - 1)$ dimensions, passing through their center of mass.

A special case of this theorem for $n = 3$, $p = 3$, $F(r) = 1/r$ is a well-known theorem of Dziobek [1].

If $\alpha > 1$ the vanishing of all constants of the angular momentum of the system ceases to be a necessary condition for the general collision of the n particles. In this case there exist spatial trajectories of general collision.

For the case $n = 3$, Sokolov uses a generalized method of Sundman to investigate the behavior of certain quantities characterizing the motion in a neighborhood of the instant of triple collision. He assumes that for small r and $\alpha \neq 1$ the inequality

$$(1 - \alpha) \frac{d}{dr} [r^{2\alpha} F(r)] \geq 0$$

holds. Further he proves that in the planar motion as $t \to t_1$, the instant of triple collision, and $\alpha \neq 1$, the three particles tend to form one of the two limit configurations obtained by Sundman in the case of Newtonian attraction, namely an equilateral triangle or a rectilinear configuration with ratios of the distances determined by the masses.

For $\alpha = 1$, there exist trajectories of triple collision such that the ratios of the mutual distances do not tend to definite limits as $t \to t_1$.

Next, by introduction of new variables, Sokolov reduces the equations of motion to a special form, used later on. In the planar motion, after the equations have been reduced to four first-order differential equations with J as the independent variable and two quadratures, certain generalized results of Bohl [1], Cotton [1], and others on asymptotic solutions of differential equations are applied to establish the existence of trajectories of triple collisions. For the construction of the corresponding solutions, in general, the method of successive approximations can be used.

Assuming that $r^{2\alpha}F(r)$ or $r^{2\alpha+1}f(r)$ can be represented in a neighborhood of the point $r = 0$ by series in terms of positive integral powers of $r^{\alpha_4}, r^{\alpha_5}, \cdots, r^{\alpha_n}$, where $\alpha_4, \alpha_5, \cdots, \alpha_n$ are arbitrary positive numbers, none of which is a linear combination of the remaining ones, Sokolov gives analytic representations of the solutions around $J = 0$ for the two previously mentioned limit configurations.

He also considers spatial trajectories of triple collision with the above-mentioned two limit configurations, and shows that the corresponding solutions depend upon ten arbitrary parameters in the case of the equilateral triangle limit configuration, and upon eleven parameters in the case of the collinear limit configuration.

Finally, motions in space, having a plane of symmetry or an axis of symmetry, are considered. In particular, trajectories of triple collisions with a collinear limit configuration are investigated for motions having a plane of symmetry.

8. Trajectories of unlimited divergence of particles: $\lim_{t \to t_1} J^2 = +\infty$

Sokolov assumes that $f(r)$ is analytic for positive r, continuous for $r = 0$, and increases indefinitely, as $r \to +\infty$, in such a way that

$$\lim_{r \to +\infty} r^{1-2\beta}f(r) = 2\beta > 0 \qquad \left[\lim_{r \to +\infty} r^{-2\beta}F(r) = 1\right]$$

In order that

$$\lim_{t \to t_1} J^2 = +\infty \tag{10}$$

holds for a finite t_1, it is shown that, of necessity, $\beta > 1$.

Let $\underline{r}(t)$ denote the smallest of the three mutual distances r_0, r_1, r_2 at the instant t, and let $\underline{p} = \underline{r}/J$. Then *a priori* under the assumption (10), the following three cases are possible:

(i) $\lim\limits_{t \to t_1} \underline{p} = 0;$

(ii) the lower bound of $\underline{p}$ is positive in a neighborhood of $t = t_1$,

(iii) $\underline{p}$ oscillates between zero and a certain positive upper bound.

It can be shown that in case (i) the ratios

$$p_i = r_i/J \qquad (i = 0, 1, 2)$$

as well as

$$V = J^{-2\beta}\, U = J^{-2\beta}\Sigma m_i m_j F(r_{ij}) \qquad (i \neq j)$$

tend to finite limits.

Assume that $rf(r) - 2\beta F(r) = f_1(r) f_2(r)$, where for $r > r^* > 0$ the function $f_1(r)$ remains bounded, whereas the function $f_2(r)$, being positive and nondecreasing, is such that

$$\int_r^{\infty} f_2(r)\, dr/r^{2\beta+1}$$

exists. Then it can be shown that in all three cases mentioned above the function

$$R^2 = J^{-2\beta}(dJ/dt)^2$$

tends toward a finite positive limit R_1^2 as $t \to t_1$. Under the assumption $\lim\limits_{t \to t_1} r_i = +\infty$, which is always satisfied in case (ii), we have also $\lim\limits_{t \to t_1} 2V = R_1^2$.

Considering planar motions Sokolov shows that in the cases (i) and (ii) only the following three cases are possible:

(*a*) $\lim p_0 = 0, \qquad \lim p_1 = \lim p_2 = \{M/[m_0(m_1 + m_2)]\}^{1/2}$
where $M = m_0 + m_1 + m_2$;

(*b*) $\lim \dfrac{p_0}{p_2} = \lim \dfrac{r_0}{r_2} = 1, \qquad \lim \dfrac{p_1}{p_2} = \lim \dfrac{r_1}{r_2} = 1$
i.e., the three particles approach the vertices of an equilateral triangle as their limit positions;

(*c*) $\lim p_2 = \{M/[m_0(m_1 + m_2) + 2m_0 m_2 q + m_2(m_0 + m_1)q^2]\}^{1/2}$

$$\lim \frac{p_0}{p_2} = \lim \frac{r_0}{r_2} = q, \qquad \lim \frac{p_1}{p_2} = \lim \frac{r_1}{r_2} = 1 + q$$

where q is a positive root of the equation

$$m_0(1 + q)[(1 + q)^{2\beta-2} - 1] + m_1 q(q^{2\beta-2} - 1) + m_2 q(1 + q)[(1 + q)^{2\beta-2} - q^{2\beta-2}] = 0$$

Further, the existence of planar and spatial trajectories of unlimited divergence of particles is proved, corresponding to the previously mentioned three limit configurations. For this to be possible, it is necessary to assume that $r^{2-2\beta}\, df(r)/dr$ tends toward a definite limit $2\beta(2\beta - 1)$ and that in case (a) the function $F(r_0)/J^{2\beta}$ satisfies certain supplementary conditions.

The family of solutions of the equations of motion, corresponding to spatial trajectories with equilateral triangle as the limit configuration [case (b)] depends upon ten arbitrary parameters, whereas, those corresponding to the limit cases (a) and (c), depend upon eleven or twelve arbitrary parameters.

Further, Sokolov studies trajectories of spatial motion having a plane or an axis of symmetry, the latter case being investigated in detail, assuming that $\lim_{t \to t_1} r_2/r_0 = 0$.

Also the trajectories of an asteroid have been studied, in the spatial restricted problem, when the asteroid recedes to infinity. This case has some peculiarities, and the results obtained cannot be derived from the general problem by passage to the limit, in contrast to the case of the trajectories of binary collision.

9. Two other generalized three-body problems

Sokolov [1, 2] and Sklyanskiĭ [1–3] have also discussed binary and triple collisions in two other generalized three-body problems, obtained by specifying the function $f(r_k)$ in (1) to

$$g^2 \ln r_k/a \qquad \text{or} \qquad g^2 e^{a/r_k} \qquad (i, j, k = 0, 1, 2;\, i \neq j \neq k)$$

where r_k denotes the distance between the particles P_i and P_j, g^2, and a being two positive constants.

BIBLIOGRAPHY

Birkhoff, G. D. [1], *Dynamical Systems*, American Mathematical Society Colloquium Publications, **9**, chapter 9, New York (1927).

Bisconcini, G. [1], Sur le problème des trois corps, *Acta Math.*, **30**, 49–92 (1906).

Bohl, P. [1], Uber einige Differentialgleichungen allgemeinen Characters, welche in der Mechanik anwendbar sind (Thesis), 113 pages, Dorpat (1900), or *Bull. Soc. Math. France*, **38**, 5–138 (1910).

Chandrasekhar, S. [1], *Principles of Stellar Dynamics*, University of Chicago Press, chapter 5, Chicago (1942).

Chazy, J. [1], Sur certaines trajectoires du problème des *n* corps, *Bull. Astr.*, **35**, 321–389 (1918).

Cotton, E. [1], Sur les solutions asymptotiques des équations différentielles, *Ann. École Norm. Sup.* (3), **28**, 473–521 (1911).

Dziobek, O. [1], *Die mathematischen Theorien der Planetenbewegungen*, p. 68, J. A. Barth, Leipzig (1888).

Jacobi, C. G. J. [1], *Vorlesungen über Dynamik* (Lecture 4), Gesamm. Werke, Supplementband, G. Reimer, Berlin (1884).

Levi-Civita, T. [1], Sur la regularization du problème des trois corps, *Acta Math.*, **42**, 99–144 (1920).

Lévy, J. [1], Les approches dans le problème des trois corps (Thesis), 58 pages, Gauthier-Villars, Paris (1943).

Mendes, M. [1], Recherches sur le problème des *n* corps à masses variables, *Ann. Fac. Sci. Toulouse* (3), **27**, 1–169 (1935).

Painlevé, P. [1], *Leçons sur la théorie analytique des équations différentielles*, A. Hermann, Paris (1897).

Sklyanskiĭ, A. L. [1], On trajectories of a double collision in the problem of three bodies interacting with forces proportional to the logarithms of the mutual distances (Russian), *Ukrain. Mat. Zhurn.*, **6**, 349–362 (1954).

——— [2], On the theory of pairwise collision in the generalized three-body problem (Russian), *Ukrain. Mat. Zhurn.*, **7**, 160–166 (1955).

——— [3], Sur la question de la classification des cointonations dans le problème généralisé des trois corps (Russian, French summary), *Ukrain. Mat. Zhurn.*, **9**, 66–81 (1957).

Sludskiĭ, F. [1], On the problem of several bodies (Russian), *Mat. Sborn.*, **9**, 536–545 (1879).

Sokolov, Yu. D. [1], On rectilinear motion with a common collision of a system of three material points mutually attracting according to an exponential law (Russian), *Ukrain. Mat. Zhurn.*, **2**, no 4, 18–24 (1950).

——— [2], On the motion on a straight line of a system of three material points, each acting with forces proportional to the logarithms of their mutual distances (Russian), *Ukrain. Mat. Zhurn.*, **2**, no. 4, 25–36 (1950).

——— [3], Singular trajectories of a system of three material points (Russian), *Monografiĭ Instituta Matematiki*, **1**, Akad. Nauk Ukrain. S.S.R., 126 pages, Kiev, 1951.

——— [4], On some general characteristics of the behavior of a material system in the neighborhood of a singular instant (Ukrainian, Russian Summary), *Dopovidi Akad. Nauk Ukrain. R.S.R. 1951*, 227–233 (1951).

Sundman, K. F. [1], Mémoire sur le problème des trois corps, *Acta Math.*, **36**, 105–179 (1912).

Weierstrass, K. [1], see Mittag-Leffler, G., Zur Biographie von Weierstrass, *Acta Math.*, **35**, 29–65 (1912).

THE THEORY OF OSCILLATIONS

BY N. MINORSKY
Stanford University

INTRODUCTION

In this review we shall attempt to outline briefly the existing trends in the new theory of oscillations, as contrasted to the old theory based on the use of the linear differential equations (d.e. for short).

The necessity for the enlargement of the old theory of oscillations was recognized long ago in view of a gradually increasing number of experimental facts which could not be well explained on that basis. Thus, for instance, difficulties appeared when one tried to explain the stationary state of oscillations produced by electric arcs. It was obvious that the method of small motions of the classical theory could not be applied here, as these oscillations cannot be considered as *small* in the usual sense of this word. Moreover, the old (linear) theory indicated in this case an exponential increase of amplitudes without any bound which was obviously absurd on physical grounds. At the beginning of this century a similar situation arose with the theory of the electron tube circuits producing self-sustained oscillations.

Although the reason for these difficulties was seen in the nonlinearity of the experimental characteristics of these oscillators, the lack of any analytical methods directed the efforts of physicists to the elaboration of all kinds of graphical methods with a view to determining the stationary state.

The first step in the direction of an analytic approach to the study of these phenomena was made by van der Pol [1] who formulated a nonlinear d.e.

$$\ddot{x} + \mu(x^2 - 1)\dot{x} + x = 0 \tag{1}$$

which acquired at once the role of a nucleus around which the new studies began to form. Van der Pol investigated the solution of his d.e.

graphically by the isocline method and found that, in fact, it possessed a unique closed integral curve (i.e., a periodic solution) in the plane of the variables (x, $\dot{x}$)—the phase plane—toward which approach all other (nonperiodic) solutions in the manner of spirals winding themselves onto the closed curve.

This discovery was fundamental indeed, because, for the first time, it became possible to understand the true mechanism of the self-sustained oscillations which, for many years, could not be explained by the linear theory.

Although later contributions of E. and H. Cartan [2] and of Liénard [3] justified analytically the graphical result obtained by van der Pol, the matter remained limited to this isolated conclusion until Andronov [4] pointed out that the closed curve of van der Pol is nothing but the *limit cycle* of the theory of Poincaré [5].

Once this contact with the theory of Poincaré had been established, the subsequent developments proceeded systematically and everything happened in the manner as if the mathematical theories of the great analyst were merely awaiting the experimental facts to give rise to the new theory of oscillations. Thus, for instance, the singular points were identified with the positions of equilibria, the limit cycles with the stationary motions, the self-excitation from rest with the theory of bifurcations, and so on. This early period (1930–1940) of the codification of the new science took place almost exclusively in the U.S.S.R. [6].

The parameter μ in (1) turned out to be a real "yardstick" of possibilities. If μ is small, the above-mentioned connection with the theory of Poincaré becomes possible and all known nonlinear phenomena become explained on this basis, as we shall see in Chapter 3.

Certain difficulties were experienced when, from the early studies of self-sustained oscillatory phenomena, the investigation of other phenomena like subharmonic resonance [7], asynchronous action [8], etc., was undertaken. In the first case, the d.e. do not contain the independent variable t (time) explicitly (the so-called autonomous systems), whereas in the second they do contain it (the nonautonomous systems). Although from the standpoint of the theory of d.e. this difference is not essential, it turned out to be of importance in the new theory of oscillations merely because the latter started developing on the basis of the topological concepts (singular points, limit cycles, etc.) introduced by Poincaré, and it became difficult to give up these concepts in the case of the nonautonomous systems for which they do not apply. This, however, did not constitute any essential difficulty but merely complicated the calculations, particularly in the question of stability. We shall see this complication in the case of the theory of the subharmonic resonance developed by Mandelstam and Papalexi [7]. Another possibility is to reduce a nonautonomous system to an autonomous form by a formal transformation which simplifies the analysis once this reduction is effected; this method is outlined in Chapter 2, Art. 5. The question of the nonautonomous systems thus does not present any special difficulties as long as the parameter μ is small.

An entirely different order of difficulties appeared in the case of large parameter values. One can see this difficulty on purely physical grounds if one considers (1) as representing an oscillator. Already the early graphical results of van der Pol [9] showed that, if μ is small, the oscillator behaves *practically* like a harmonic oscillator. If, however, μ is not small, the oscillation exhibits a very complicated form consisting of portions of a cycle traversed with a finite velocity with interposed stretches traversed quasi-discontinuously. It was obvious that it would be extremely difficult to determine an analytical solution (e.g., in a form of a series) which would be adequate to represent these two widely different behaviors. In fact, a long and persistent effort of mathematicians to find an analytic solution was not crowned with a success [10], and this resulted in a certain "parting of the ways" between the two lines of endeavor to find an issue.

One of these attempts [11] consisted in treating the quasi-discontinuous stretches as real discontinuities and applying the procedure of the theoretical mechanics used in the theory of shocks. It is recalled that this procedure disregards what happens *during* a shock and merely connects the conditions *before* and *after* the shock by a certain *additional information* not related directly to the d.e. (theorems of momentum and kinetic energy). Following this approach, a discontinuous theory was evolved in U.S.S.R. which resembles closely the theory of shocks with a certain "additional information" appearing here in the form of the so-called condition of Mandelstam. This discontinuous theory turned out to be very successful inasmuch as practically all nonlinear phenomena of this kind—the relaxation oscillations—could be treated uniformly on this basis [12]. The objection to the discontinuous theory is mostly that it is not a purely analytical theory but is a kind of a hybrid theory mixing up a mathematical treatment with a physical postulate regarding the invariance of energy *during* the rapid transitions. Strictly speaking, this objection is hardly justified because the very nature of the problem shows that the discontinuous treatment here is just as logical to apply as it is in the classical theory of shocks where it is now universally accepted. The essential point is that the theory based on this idealization gives a simple and consistent explanation of all known relaxation phenomena and is thus justified on this basis.

The second line of endeavor originated principally in the important publications of Cartwright and Littlewood [13] and led to the establishment of the conditions for the existence of a periodic solution for (1) in the case when μ is not small. The nonautonomous solution was also considered in a similar manner. One interesting detail of this work is that the problem is possible if, in addition to the d.e. in question, one has to take into account also two additional equations—the "integrated" one and the "energy" equation which thus seem to play the role of the "additional information" just mentioned in connection with the theory of shocks. It is to be noted that this theory gives means of investigating the qualitative behavior of the solution but does not permit its actual calculation (e.g., in a form of a series). In spite of a considerable theoretical interest, these developments did not progress at present far enough to offer a practical tool for applied problems, as is the case of the discontinuous theory.

Aside from these major problems of the general theory, a number of other problems appeared recently which are more or less outside the scope of the principal methods. One of these is the problem concerning oscillations arising from retarded actions [14], which are very often encountered in connection with modern control systems as well as in a number of some special problems of econometrics and in studies of the biological fluctuations. The essential feature of these "retarded actions" is that they always lead to a transcendental problem which is outside the scope of the general methods of Poincaré but can be occasionally reduced to them under some additional restrictions [15]. One must admit, however, that the whole situation here is far less explored than in the classical "nearly linear" theory and leads, generally, to the nonlinear difference-differential equations the theory of which did not progress enough to permit establishing a definite connection with these phenomena.

Another interesting domain for which there exists a certain experimental evidence, but so far no theory, is that of the so-called inertial nonlinearities [16]. By this one means generally the nonlinear characteristics which are not fixed but vary as a function of a parameter. This feature appears, for instance, in the case of a heat-responsive conductor (the thermistor) inserted in an oscillating system. As the nonlinear characteristic here is affected by the heat, and the latter is a function of time, one can see that the behavior of such a system is ultimately reduced to a nonlinear integro-differential equation from which one has to establish the usual conditions for the stationary state. Here one encounters again a problem in which the analytical solution is not yet available.

We have indicated a few of these problems, taken more or less at random, in order to emphasize the fact that not all gaps between the mathematical theories and the observed facts have been bridged as yet, which is, of course, a natural state of things for a science which is still in full evolution.

Referring more specifically to the contents of this review, Chapter 1 gives a brief outline of the general topological methods as viewed from the standpoint of the oscillation theory. This subject can be found in any textbook on the theory of d.e. (references [*e*] and [*f*]), and we mention it here merely for the purpose of emphasizing some of the most important connections with the theory of oscillations.

Chapter 2 deals with the fundamentals of the theory of approximations used extensively in the quantitative work, particularly in connection with the first approximation beyond which the applied problems generally do not go if the parameter is sufficiently small. In Art. 5 of Chapter 2 is outlined a method of approximation, worked out recently in collaboration with M. Schiffer, which is often useful for the formal reduction of nonautonomous problems to the autonomous form, as was previously mentioned; some applications of this method are given in Chapter 3. In Art. 7 of Chapter 2 are indicated some of the latest developments in the theory of approximations.

Chapter 3 is devoted to the survey of the principal phenomena in nearly linear systems. Strictly speaking, the contents of Art. 6 are outside the scope of the general methods, but it is shown that it is still possible to use the perturbation theory if one assumes the smallness of time lag and considers the problem under some other limitations.

The last chapter gives an account of the discontinuous theory of relaxation oscillations, following closely the original presentation of Chaikin and Lochakov [11] but omitting a number of other examples of relaxation oscillators which can be found in Andronov and Chaikin's book [12]. Article 7 of Chapter 4 contains the elements of the asymptotic approach to those phenomena which are at present less developed, as far as their applications are concerned, than the discontinuous theory.

CHAPTER 1

GENERAL METHODS

1. Phase plane; singular points

We consider an autonomous system of d.e. of the form

$$\dot{x} = \frac{dx}{dt} = P(x, y), \qquad \dot{y} = \frac{dy}{dt} = Q(x, y) \tag{1}$$

In applications this represents a system with one degree of freedom. A d.e. of a second order can be reduced to this form by setting $y = \dot{x}$.

If one considers x and y as coordinates of a certain plane (x, y)—the phase plane—the theorem of Cauchy may be stated in this way: Through every *ordinary* point (x, y) of the phase plane passes one, and only one, integral curve defined by the equation

$$\frac{dy}{dx} = \frac{Q(x, y)}{P(x, y)} \tag{2}$$

It is also possible to consider the system (1) as defining the components $\dot{x}$ and $\dot{y}$ of a velocity vector $\mathbf{V}(x, y)$.

The two descriptions (1) and (2) yield the same integral curve, but (1), in addition, gives information on the motion of the "representative point" R on it. In what follows we shall call "integral curve" the curve defined by (2) and "trajectory," the same curve with the corresponding motion of R as given by (1). In this terminology the term "the representative point" R means the instantaneous state of the system defined by (1).

A special situation arises for a point (x_0, y_0) for which $P(x_0, y_0) = Q(x_0, y_0) = 0$; such a point is called a *singular point*, and it is clear that for this point $\mathbf{V}(x_0, y_0) = 0$, i.e., a singular point is a point of *equilibrium*, and this is the reason why the concept of a singular point plays an important role in the theory of oscillations.

Since the system (1) is autonomous, one can replace t by $t + t_0$, t_0 being arbitrary, so that the solution can be written as

$$x(t) = x(t - t_0, x^0, y^0), \qquad y(t) = y(t - t_0, x^0, y^0) \tag{3}$$

where x^0, y^0 are the initial conditions.

A convenient way of investigating the nature of singular points is to consider first the system (1) in the form

$$\frac{dx}{dt} = -x, \qquad \frac{dy}{dt} = -ay \tag{4}$$

or the d.e.

$$\frac{dy}{dx} = \frac{ay}{x} \tag{5}$$

It is seen that the origin $x = y = 0$ is a singular point according to the above definition, and from (5) one also notes that at this point the direction field is indeterminate inasmuch as the singular point may be regarded as a special trajectory reduced to one single point.

The integration of (5) gives the integral curve

$$y = Cx^a \tag{6}$$

C being a constant of integration. This curve may have different forms, depending on the value of the parameter a.

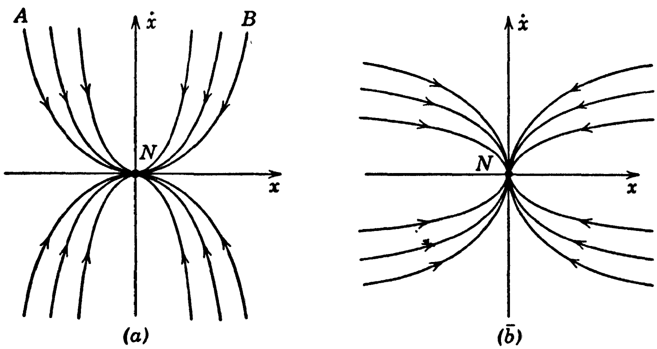


Fig. 1. Trajectories in the neighborhood of a node.

If $a > 0$, (6) represents a family of parabolic curves shown in Figs. 1(a) and (b). If $a > 1$, the curves of the family approach the singular point along the abscissa axis; if $a < 1$, this approach occurs along the axis of ordinates. The essential point is that the parabolic curves approach the singular point from a *definite direction.* A singular point of this nature is called the *nodal point* or the *node* N.

If $a = 1$, one has a special case. The integral curves are, in this case, straight lines $y = Cx$ through the origin which is still a singular point, called a degenerate node or a *star.* The difference with the general case of the node is that the integral curves enter such a special node from *all* directions.

If one wishes to consider these curves as *trajectories,* one must refer to (4) and investigate the sign of $\dot{x}$ and $\dot{y}$ in the various quadrants. Thus, for instance, in the first quadrant $x > 0$, $y > 0$ which gives $\dot{x} < 0$, $\dot{y} < 0$; this specifies direction on the curves shown by arrows. If one changes the sign on the right-hand side of equations (4), the directions on the integral curves are reversed.

It must be noted that a trajectory stops at the node so that the integral curve ANB (Fig. 1(a)), considered as *trajectory,* comprises two distinct trajectories AN and BN excluding the point N and a third trajectory comprising only one point N.

If $a < 0$, a similar discussion shows that the integral curves are given by

$$yx^{|a|} = C \tag{7}$$

They are thus hyperbolic curves (reducing to the ordinary hyperbola referred to its asymptotes for $|a| = 1$). The equivalent system in this case is

$$\frac{dx}{dt} = -x, \qquad \frac{dy}{dt} = |a|y \tag{8}$$

The determination of positive directions on trajectories is again obtained by considering the sign of $\dot{x}$ and $\dot{y}$ in different quadrants, which gives the family of hyperbolic trajectories shown in Fig. 2. The point $x = y = 0$ is a singular point called the *saddle point* S. In this case there are four asymptotic trajectories DS, BS, SA, and SC. Along the first two, the representative point R "enters" the saddle point for $t \to \infty$; along the last two, it, on the contrary, leaves it. All other hyperbolic trajectories do not enter S and are traversed with a finite velocity.

Consider further a differential system of the form

$$\dot{x} = -ax + y, \qquad \dot{y} = -x - ay \tag{9}$$

The origin $x = y = 0$ is a singular point called the *focal point,* or the focus F. In polar coordinates ($x = r\cos\phi$, $y = r\sin\phi$) these equations reduce to

$$r = Ce^{-at}, \qquad \dot{\phi} = -1 \tag{10}$$

representing a logarithmic spiral. The point R following the spiral approaches the focus without any definite direction as the spiral turns around F indefinitely for $t \to \infty$ (Fig. 3). Through every point A of

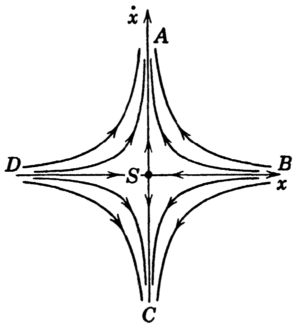


Fig. 2. Trajectories near a saddle point.

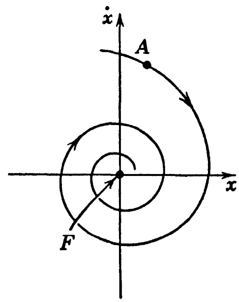


Fig. 3. Trajectories near a focal point.

the phase plane (i.e., for some initial conditions) passes one, and only one, such a spiral.

Finally, for the d.e. of the harmonic oscillator $\ddot{x} + x = 0$, the system (1) is

$$\dot{x} = y, \qquad \dot{y} = -x \tag{11}$$

The trajectories in this case form a family of concentric circles, as shown in Fig. 4. The origin here is again a singular point, called the *center* C; all trajectories turn around C and none of them "enters" C (Fig. 4).

It can be shown that a d.e. of the form

$$\ddot{x} + 2b\dot{x} + \omega^2 x = 0 \tag{12}$$

can be always reduced to one of the preceding forms corresponding either to a node or to a focus. The node appears if $b^2 - \omega^2 > 0$ and the focus if $b^2 - \omega^2 < 0$. Thus, the node is a singular point (position of equilibrium) to which tends an aperiodically damped motion, and the focus is a singular point to which approaches an oscillatory damped motion. The center arises if $b = 0$ (conservative system). As to the saddle point, it characterizes an essentially unstable motion like the one that appears around the unstable position of equilibrium ($\theta = \pi$) of a pendulum.

The reduction of (12) to the previously indicated form is carried out by means of certain linear (affine) transformation of variables (which we omit here) [17]. Owing to this, the form of the curves so reduced differs from those shown in Figs. 1, 2, 3, and 4, but the essential properties of trajectories in the neighborhood of these singular points remain the same.

A more general study of singular points is made starting from the system

$$(13) \qquad \dot{x} = ax + by + P_2(x, y), \qquad \dot{y} = cx + dy + Q_2(x, y)$$

where P_2 and Q_2 are polynomials in x and y beginning with terms of the second degree or higher. In a great majority of cases encountered in

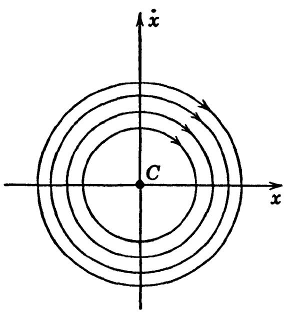


Fig. 4. Trajectories near a center.

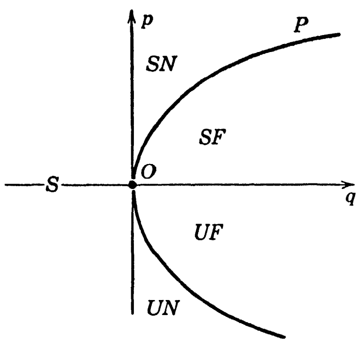


Fig. 5. Distribution of singular points.

applications, the study of singular points can be made from the linear approximation by dropping P_2 and Q_2 in (13). One can see intuitively the reason for this simplification by the following argument: the equilibrium of the system is a *local* property depending on the conditions near the equilibrium point (theoretically, infinitely near to that point); hence x and y are small, which justifies the dropping P_2 and Q_2. If we assume for the moment that this assumption is legitimate, then the standard procedure [18] consists in reducing the system to the *canonical form* by a linear transformation of variables, in which case, instead of (13), one has

$$(14) \qquad \dot{\xi} = S_1\xi, \qquad \dot{\eta} = S_2\eta$$

where S_1 and S_2 are certain constants.

From the general theory it follows that this reduction is possible if S_1 and S_2 are the roots of the *characteristic* equation

$$S^2 - (a + d)S + (ad - bc) = 0 \tag{15}$$

It is supposed that the roots are distinct, which imposes certain conditions on the coefficients of the linear transformation [18].

Leaving out these details, we mention merely the conclusions, namely: the nature of singular points (and, hence, of equilibrium) is related to that of the roots of (15):

1. If S_1 and S_2 are real and of the same sign, the singular point is a node.
2. If S_1 and S_2 are real and of opposite signs, the singular point is a saddle point.
3. If S_1 and S_2 are conjugate complex, the singular point is a focus.
4. If S_1 and S_2 are purely imaginary, the singular point is a center.

This suggests a simple representation of *the distribution* of singular points in the (p, q) plane (Fig. 5); $p = -(a + d)$; $q = ad - bc$. If one draws in this plane a parabola P of equation $p^2 = 4q$ the axes p and q and the curve P define five regions: SN, SF, UF, UN, and S, corresponding to stable nodal, stable focal, unstable focal, unstable nodal, and saddle points respectively. The origin O appears thus as a *branch point* around which permute different equilibria. It is observed that the focal points (stable or unstable) cannot be transformed into saddle points without passing first through the regions of nodes.

The above conclusions are valid if (15) has distinct roots. If this is not the case, or if one of the roots is zero, a special investigation is needed.

Although the investigation of singular points is very simple in cases generally encountered in applications where the linear approximation is valid, it may become complicated in some special "pathological" cases. Thus, for instance, the existence of purely imaginary roots does not necessarily indicate a center; in some cases it may also indicate a special focus, and the linear approximation does not permit ascertaining which of the two singularities exists in a given case [19] also [*e*] and [*f*]. Still more complicated are cases where the linear terms are missing in (13), thus leading to singular points of a higher order. Although there exists a considerable mathematical literature on this subject [*e*], [20], so far, practically, no connections have been established in regard to some special conditions of equilibria (particularly in electric circuits) which are reported from time to time from the experimental evidence.

Summing up, practically in all nonlinear oscillations of the self-sustained type, the singular points are either foci or nodes (stable or unstable); occasionally, one encounters saddle points which always mean instability. In conservative systems the only singularities encountered are centers and saddle points, but these systems do not play an important role in applications. More important are systems which may be considered as "conservative on the average," but these, in reality, are nonconservative systems in a stationary state and we shall investigate their properties in the following article.

2. Limit cycles; topological configurations [21]

The second basic concept in the new theory of oscillations is that of *limit cycles* identified with *stationary periodic motions.*

As this constitutes the most important point of the new theory, it is useful to contrast it with the pattern of such motion—the harmonic oscillator—of the old theory. As we saw already, the periodic motions around the center, form a continuous family of closed curves enclosing the center in its interior. In the linear case these curves are either circles, (Fig. 4) or, more generally, ellipses, if the d.e. of the harmonic oscillator is taken in the form of $\ddot{x} + \omega^2 x = 0$. In a still more general case, where the nonlinearity is taken into account, the closed curves are certain ovals enclosed inside each other. However, as long as the singular point is a center, the general feature remains the same, namely, there exists a two-parameter family of closed curves surrounding the center. If one selects some particular initial condition (a point in the phase plane), this specifies a particular trajectory out of the infinity of possible trajectories.

In contrast with this well-known property of trajectories of a conservative system, we shall consider now an entirely new type of a stationary motion—the limit cycle—which can be defined as follows.

We call the limit cycle a closed trajectory C (in the phase plane) having the following property: all other trajectories C' are not closed and approach C either for $t \to \infty$ (stable limit cycle) or for $t \to -\infty$ (the unstable cycle). We use the term "trajectory approaches C" in the sense that R moving on C' approaches some point on C.

Figures 6(a) and (b) illustrate a stable and an unstable limit cycle respectively.

This definition needs an additional information which is supplied by the theorem of Poincaré which states: In order for a closed curve C to be a *trajectory,* it is necessary (but not sufficient) that there be at least one singular point in its interior the stability of which be opposite to that of C.* Thus, a stable limit cycle must have inside it an unstable singular point, Fig. 6(a), and vice versa. This establishes the important concept of a *topological configuration* in its simplest form, e.g., of an unstable (stable) singular point surrounded by a stable (unstable) limit cycle. This, in turn, permits visualizing the situation almost intuitively, if one considers a trajectory as a line of flow (of a "fluid" of trajectories)

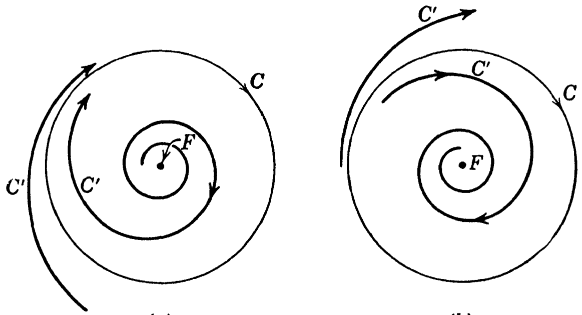


Fig. 6 (a) Stable limit cycle; (b) unstable limit cycle.

originating at an unstable element (singular point, limit cycle) and disappearing at a stable one. The unstable elements appear thus as *sources* and the stable ones as *sinks*. Such a hydrodynamical analogy is useful in appraising more complicated topological configurations. In view of the asymptotic approach to the elements of configuration, such configuration describes the process between $t = -\infty$ and $t = +\infty$. It is more convenient to operate with intervals $(t_0, +\infty)$, $(-\infty, t_0)$ by assigning t_0 to the instant when R coincides with some point A. In such a case it is customary to speak about *half-trajectories* originating at A.

The comparison of properties of a limit cycle with those of trajectories of a conservative system leads to the following remarks:

1. A limit cycle is a unique trajectory in the sense that there exists no other limit cycle at least in its neighborhood.†

2. Motion on the limit cycle does not depend on the initial conditions inasmuch as from any point on C' the ultimate motion establishes itself on C. Thus, for instance, the initial conditions do not play any role in

* We give here only a simplified definition. For more complete information, see the theory of indices in references [e] or [f].

† This statement relates to normal cases generally encountered in applications. There are, however, some special "pathological" cases in which limit cycles accumulate, forming a kind of "cluster." In [18], an example of this kind is indicated but, as far as known, it has no physical significance.

determining the stationary oscillation of an electron-tube oscillator or a clock, this oscillation being determined uniquely by the parameters of the d.e. itself.

The simple topological configuration of Fig. 6 was generalized later for *polycyclic* configurations which (for the first approximation, Chapter 2) may be considered in the form of concentric circles. In such a case the stability of cycles always alternates, the singular point at the center of configuration being regarded as a cycle reduced to one point. Figure 7 gives a frequently encountered *bicyclic* configuration which may be designated for short as SUS (reading: stable singular point S is surrounded by an unstable cycle C_1 which, in turn, is surrounded by a stable one C_2.

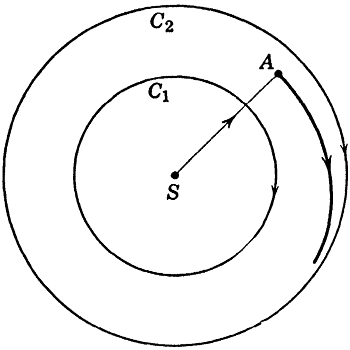


Fig. 7. Bicyclic configuration.

A system of this kind does not become self-excited from rest but, if an impulse is communicated to the system so as to bring R (the "initial conditions") to some point A outside the unstable cycle C_1, the self-excitation develops normally, R tending to approach the stable cycle C_2.

This phenomenon is called sometimes "hard" self-excitation (to distinguish it from the "soft" self-excitation occurring spontaneously from rest, as in Fig. 6(*a*)), and is well known in the electron-tube circuits under the appropriate form of the polynomial $f(x)$ in the generalized van der Pol equation

$$\ddot{x} + \mu f(x)\dot{x} + x = 0$$

as we shall see later.

As examples of d.e. possessing limit cycles, one can indicate the following systems

$$(16) \qquad \dot{x} = y + \frac{x}{\sqrt{x^2 + y^2}}[1 - (x^2 + y^2)] = P(x, y),$$

$$\dot{y} = -x + \frac{y}{\sqrt{x^2 + y^2}}[1 - (x^2 + y^2)] = Q(x, y)$$

$$(17) \qquad \dot{x} = -y + x(x^2 + y^2 - 1), \qquad \dot{y} = x + y(x^2 + y^2 - 1)$$

The discussion of both systems is very simple in polar coordinates and leads in the case of (16) to the d.e.

$$\dot{r} = 1 - r^2, \qquad \dot{\theta} = 1 \tag{18}$$

yielding

$$r = \frac{Ae^{2t} - 1}{Ae^{2t} + 1}, \qquad \theta = \theta_0 + t$$

where $A = (1 + r_0)/(1 - r_0)$, r_0 being the initial value. For $t \to \infty$, $r \to 1$, both for $r_0 > 1$ and $r_0 < 1$. The stability condition is seen directly from the first equation (18). The system (16) yields thus a stable limit cycle $r = 1$.

In the case of (17) in polar coordinates, one has

$$\dot{r} = r(r^2 - 1), \qquad \dot{\theta} = -1$$

and the integration gives

$$r = \frac{1}{\sqrt{1 - Ae^{2t}}}$$

where the constant $A = (r_0^2 - 1)r_0^2$. For $r_0 < 1$, $A < 0$, so that $r = 1/\sqrt{1 + |A|e^{2t}}$ which shows that, for $t = -\infty$, $r = 1$; the same result is obtained for $r_0 > 1$. This corresponds to the unstable limit cycle shown in Fig. 6(*b*) with radius $r = 1$. These examples are formed from relatively simple relations in polar coordinates and by transforming the d.e. into cartesian coordinates.

In reality, from a given *form* of a d.e. it is impossible to say whether it has a limit cycle or not. Aside from the van der Pol equation and a few others, very little is known about the determination of limit cycles on the basis of the general topological methods.

Poincaré established [22] the necessary conditions for the existence of limit cycles in his theory of indices, which are, however, not sufficient. There exists also another criterion in the form of the Poincaré-Bendixson theorem [23], which gives both necessary and sufficient conditions for the existence of a limit cycle but its use is limited. As it is impossible to enter into these questions here, we merely refer to the existing publications [24] and limit ourselves to the statement of the Poincaré-Bendixson theorem which has almost an intuitive significance on the basis of the hydrodynamical analogy previously mentioned; consider a domain D limited by two closed curves C_1 and C_2, as shown in Fig. 8, and assume that no singular points exist either in D or on its boundaries; then the Poincaré-Bendixson theorem states: If it is possible to show that the trajectories enter (leave) D through *every point* of the bounding curves, then one can assert that there exists at least one stable (unstable) cycle in D.

The consideration of sources and sinks under the specified conditions renders the significance of the Poincaré-Bendixson theorem obvious. The principal difficulty of this criterion is in the selection of bounding curves C_1 and C_2 for which the imposed condition holds. In fact, in order to be able to select these curves, one must know already the solution of the d.e., so that one is caught in a vicious circle.

In some cases it is possible to obtain an answer from a relatively simple choice of C_1 and C_2; in such a case the theorem settles the matter at once.

As an example, consider the differential system

$$(19)\quad \dot{x} = -ay + x[1 - (x^2 + y^2)], \qquad \dot{y} = ax + y[1 - (x^2 + y^2)]$$

which we shall encounter later.

As $x = y = 0$ is the only singular point at a finite distance, one can try a domain D limited by two circles C_1 and C_2, the first having small radius, and the second, on the contrary, a large radius.

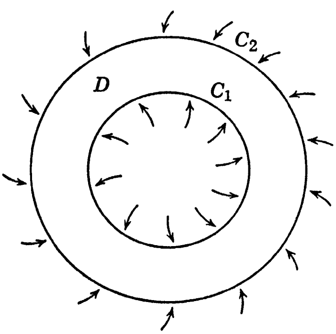


Fig. 8. Illustrating the Poincaré-Bendixson theorem.

As the origin is an unstable focus, trajectories enter D through every point of C_1 which surrounds the unstable focus as one ascertains from (19) in which one keeps only the linear terms.

For the large circle C_2, on the contrary, one keeps only the nonlinear terms in (19) (by considerations of the order of magnitude). A simple examination of the direction field of the vector $\mathbf{V} = \mathbf{x} + \mathbf{y}$ shows that the trajectories enter D through every point of C_2.

In this case the Poincaré-Bendixson theorem gives an immediate answer concerning the existence of a stable limit cycle in D. In general, such simple situations are rather exceptions than the rule. Thus, for instance, in the case of the van der Pol equation the application of the Poincaré-Bendixson theorem requires a great deal of ingenuity to construct the domain satisfying the stated requirements. The reader is referred to a paper by La Salle [25] in connection with the application of the Poincaré-Bendixson theorem to the van der Pol equation (large μ).

In view of a rather limited use of these general methods in applied problems, we shall not elaborate them here, particularly because in the theory of approximations we shall encounter much simpler means for detecting the existence of limit cycles.

3. Stability

We have encountered already the question of stability in Art. 1 in connection with singular points on a more or less intuitive basis. Thus, for instance, if R following on trajectories associated with nodal and focal points approaches these points, we defined them as stable; if, on the contrary, it moves away from them, these singularities are unstable. The same question appeared in the definition of stability of limit cycles.

It is useful, therefore, to consider the question of stability from a broader point of view. It turns out that this matter is far from being simple in view of several definitions of stability leading to different topological aspects of the problem.

As this matter in its generality is beyond the scope of this review, we refer to the corresponding references [26] and mention here only some of the definitions of stability most frequently used in the theory of oscillations:

Variational equations. Assume that a differential system

$$\frac{dx_i}{dt} = X_i(x_1, x_2 \cdots x_n) \qquad i = 1, 2, \cdots n \tag{20}$$

has a set $x_{i0}(t)$ of known periodic solutions which represent a closed curve in the n space (for $n = 2$, we have a planar representation with which we are concerned mostly in this review).

The general problem of equilibrium consists in considering a neighboring solution

$$x_i(t) = x_{i0}(t) + \xi_i(t) \tag{21}$$

where $\xi_i(t)$ is a set of functions, called *perturbations*. If one replaces (21) into the d.e. (20) and develops the functions X_i in Taylor's series around x_{i0}, keeping only the linear terms in ξ_i, one obtains a system of the *variational equations*

$$\frac{d\xi_i}{dt} = \sum_{j=1}^{n} \frac{\partial X_i}{\partial x_{j0}} \xi_i \tag{22}$$

where $\partial X_i/\partial x_{j0}$ are the partial derivatives of X_i with respect to x_j into which the known periodic solutions are replaced after the differentiation. According to the problems, the coefficients of ξ_j may be either constants or periodic functions of t.

The most frequently used definition of the *asymptotic stability* consists in the requirement that all $\xi_i(t) \to 0$ as $t \to \infty$.

It is clear that the problem so formulated is sufficiently broad to include also the problem of stability of equilibrium which may be regarded as a special case of a periodic motion reduced to one single point.

As an example, we consider the problem of equilibrium of a system with one degree of freedom (we use here a rather applied term which the mathematicians have a habit of calling the two-dimensional problem)

$$\dot{x} = P(x, y), \qquad \dot{y} = Q(x, y) \tag{23}$$

If x_0, y_0 is a singular point (position of equilibrium), the method of the variational equations consists in replacing x and y by $x_0 + \delta x$, $y_0 + \delta y$, where δx and δy are the small perturbations playing the role of ξ_i in (21).

The variational equations in this case are

$$\frac{d\,\delta x}{dt} = P_x(x_0, y_0)\,\delta x + P_y(x_0, y_0)\,\delta y \tag{24}$$

$$\frac{d\,\delta y}{dt} = Q_x(x_0, y_0)\,\delta x + Q_y(x_0, y_0)\,\delta y$$

where P_x, P_y, Q_x, and Q_y play the role of $\delta X_i/\delta x_{j0}$ of the general case (22). It is noted that (24) are of the form of the abridged linear system (13).

The characteristic equation in this case has the form

$$S^2 - (P_x + Q_y)S + (P_xQ_y - P_yQ_x) = 0 \tag{25}$$

which is, in fact, (15). Conditions for stability are obviously $P_x + Q_y < 0$ and

$$\begin{vmatrix} P_xP_y \\ Q_xQ_y \end{vmatrix} > 0$$

The second condition merely expresses that the singular point *is not* a saddle point (and is either a node or a focus), and the first one specifies that the real parts of the roots S_1, and S_2 are negative, which is a general condition.

In the case of more than two d.e., the procedure is similar. The characteristic equation in such a case has the form

$$\Delta(S) = \begin{vmatrix} a_{11} - S & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - S & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} - S \end{vmatrix} = 0 \tag{26}$$

If one develops the determinant, one obtains an algebraic equation of the nth degree in S, and it is necessary to formulate the conditions under which all real parts of the roots are negative. This requires the application of the well-known Hurwitz's theorem.

In problems concerning the stability of stationary motion, the variational equations have generally periodic coefficients. As is known, the general solution $x(t)$ in this case has the form

$$x_i(t) = C_1 e^{h_1 t}\lambda_{i1}(t) + \cdots + C_n e^{h_n t}\lambda_{in}(t) \tag{27}$$

where h_i are the so-called *characteristic exponents* and $\lambda_{ij}(t)$ are absolutely and uniformly convergent trigonometric series. The determination of the characteristic exponents constitutes generally a very difficult problem [27] but, for a system of the second order not containing time explicitly, the matter is considerably simplified owing to a remark of Poincaré [28] which states that, in this case, one of the characteristic exponents is zero. In such a case the second characteristic exponent h is given by the formula [29]

$$h = \frac{1}{T}\int_0^T [P_x + Q_y]\, dt \tag{28}$$

where P_x and Q_y are the coefficients of abridged equations (13). The condition of orbital stability is then

$$h < 0 \tag{29}$$

If $h > 0$, the periodic motion is unstable. If one applies the criterion (28) to the d.e. of harmonic oscillator, one finds $P_x = Q_y = 0$, which gives $h = 0$. This shows that, from the standpoint of the orbital stability, the trajectories of a harmonic oscillator are neither stable nor unstable but *indifferent*. We had already an occasion to note this point inasmuch as any perturbation (that is, any new initial condition) results in a new trajectory which has no tendency to return to its original path upon the disappearance of the perturbation. Thus, from the standpoint of orbital stability, trajectories of the harmonic oscillator appear as a very special case of motion when both characteristic exponents vanish. This results also from another theorem of Poincaré which states that if a d.e. admits a first integral, one of the characteristic exponents vanishes [30]. In fact, in this example, in view of the existence of the integral of energy, the second characteristic exponent also vanishes.

If one applies criterion (28) to the d.e. (16) and (17) possessing limit cycles, one verifies that $h < 0$ for (16) and $h > 0$ for (17).

Very often it is convenient to reduce a differential system to two equations, one of which concerns the amplitude and the other the phase. In the latter appear on the right-hand side trigonometric functions. As an example consider a d.e. $d\phi/dt = \cos\phi$. Clearly there is a position of equilibrium for $\cos\phi_0 = 0$, i.e., for $\phi_0 = \pi/2$, and $\phi_0 = 3\pi/2$. In order to see which of these two values is a stable phase, one applies the variational equation which is

$$\frac{d(\phi_0 + \delta\phi)}{dt} = \cos(\phi_0 + \delta\phi) \cong \cos\phi_0 - \sin\phi_0\,\delta\phi$$

since $\delta\phi$ is small. This results in $d\,\delta\phi/dt = -\sin\phi_0\,\delta\phi$. Hence, of the two values $\sin\phi_0 = +1$ and $\sin\phi_0 = -1$, one must take $\sin\phi_0 = +1$, inasmuch as $\delta\phi \to 0$ for $t \to \infty$ in this case. This means that $\phi_0 = \pi/2$ is the stable phase and $\phi_0 = 3\pi/2$ is unstable. We shall use this form of the variational equations later.

In addition to the classical definition of asymptotic stability ($\xi_i(t) \to 0$, for $t \to \infty$) with which we were concerned so far, there is also a more general definition of stability given by Liapounov [31] which is based on the comparison of two neighboring motions.

A motion is stable in the sense of Liapounov (or L stable) if, to a given $\epsilon > 0$, one can find $\eta(\epsilon) > 0$ such that, given $\bar{x}_0 - x_0 < \epsilon$, $\bar{y}_0 - y_0 < \epsilon$ implies that

$$(30) \quad |x(t, x_0, y_0) - x(t, \bar{x}_0, \bar{y}_0)| < \eta, \qquad |y(t, x_0, y_0) - y(t, \bar{x}_0, \bar{y}_0)| < \eta$$

for all t where x_0, y_0 are the initial conditions, and $\bar{x}_0$ and $\bar{y}_0$ are slightly modified initial conditions.

This means that, if two motions represented by trajectories issuing for $t = 0$ from two neighboring points (of the phase plane) remain not far apart from each other in the course of time, they are stable L.

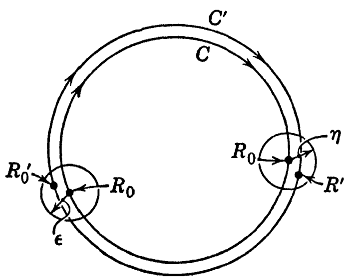


Fig. 9. Illustrating stability of motion.

To illustrate the meaning of this definition, consider two motions on two closed trajectories C and C', Fig. 9, and assume that for $t = 0$, R_0 (on C) and R_0' (on C') are close to each other (i.e., R_0' is inside the circle drawn from R_0 as center of radius ϵ). If it is possible to show that after an arbitrary time interval R and R' are still near to each other (circle of radius η), the motion is stable L. Clearly, the stability L requires an exact isochronism of the two motions.

If one takes, for instance, the motion of a pendulum, the isochronism is only approximate for small deviations when one can assume that $\sin \theta \cong \theta$. In reality the two neighboring motions will depart from each other after a sufficiently long time. Thus, the two neighboring motions of a conservative system are never L stable.

The L stable motions exist, however, in the case of a stable limit cycle. In fact, if one reproduces this argument when R_0 and R_0' are respectively on C and C', Fig. 6(a), one ascertains easily that R and R' will be still within a circle of radius η (however small) for $t \to \infty$.

In addition to this concept of *stability of motion*, there is another concept of *orbital stability*, as was shown in connection with the criterion (28); in this case, one is interested in the inherent stability of the *orbit* (i.e., closed trajectory) without specifying on it any particular motion.

One readily sees the difference between these two concepts if one considers, for instance, a d.e. of a conservative system (e.g., a frictionless pendulum). As was explained in connection with Fig. 9, two neighboring motions are always *unstable* since their periods are slightly different. On the other hand, from the standpoint of orbital stability, the closed trajectories in this case have an *indifferent* stability since, if a trajectory is perturbed which results in a neighboring trajectory, the latter does not have any tendency to return to the trajectory which existed prior to the occurrence of the perturbation.

There exists thus a slight difference in the appraisal of stability in both cases according to the concept of neighborhood one has in mind.

It is to be noted that in these two cases *the form* of the d.e. is supposed to be the same and the perturbation is due to a change of the initial conditions at some instant $t = t_0$.

One obtains still another manifestation of stability, *the structural stability*, if the perturbation is due to a *change of the form* of the d.e. This concept of stability [82] has been investigated recently and it is useful to say a few words in this connection. The term "stability" used here means really a certain *permanence* of the topological configuration or "structure" in spite of certain changes of the form of the d.e. arising, for instance, from a variation of a certain parameter. In particular, if the structure *qualitatively* remains the same, it is called *a stable structure*; if, for a small (infinitely small) variation of the parameter the topological aspect of the structure changes qualitatively, it is unstable.

It is seen that the word "stability" is somewhat overtaxed in its various manifestations although the fundamental definition (30) remains the same. Summing up, the concept of orbital stability characterizes, as we saw, oscillations of the limit cycle type in which case the orbit as a whole is stable (for a stable limit cycle). On the other hand, for conservative systems, as was also mentioned, the orbital stability is *indifferent.* Here the criteria of stability are related to the sign of the characteristics exponents. Finally, the concept of structural stability is defined by the property of invariance of the *qualitative* aspect of trajectories under small variations of a parameter in d.e.

In applied problems one encounters mostly the concept of stability of motion or stability of equilibrium (or of singular points). The latter problem generally is considerably simpler than the former, and the principal advantage of the stroboscopic method (Chapter 2, section 4) as was mentioned, is that it replaces a more complicated problem of stability of motion by a simpler one of stability of a singular point.

Before closing this brief outline of the various aspects of stability, it is useful to mention also an entirely different approach to the problem of stability which is called the *second method of Liapounoff.* In all problems of stability so far considered, the problem reduces ultimately to the integration of the variational equations, and this is possible only in a relatively few simple cases.

In the second method of Liapounoff the question of variational equations does not appear at all, and the criteria of stability (or instability) are derived from the properties of definiteness of certain algebraic forms.

These properties appear in connection with a certain algebraic function—the function of Liapounoff, $V(\xi_i, \cdots, \xi_r)$—in which ξ_i are the quantities which have been previously designated as perturbations. By definition, the function V is positive in a certain domain G, except at the origin where it is zero.

Liapounoff shows that if one can determine such a function V so that

$$\frac{dV}{dt} = \sum_{i=1}^{n} \frac{\partial V}{\partial \xi_i}\frac{d\xi_i}{dt} \tag{31}$$

is either negative or identically zero, the equilibrium is stable; if $\dfrac{dV}{dt} < 0$, the equilibrium is asymptotically stable. A number of theorems cover various cases of stability and instability.

The most important part in this method is the determination of this function V. If one succeeds in determining it, the problem reduces to ascertaining the sign of (31).

The essential feature of the method is in that it gives at the same time *stability in the large* in contrast to the classical method in which stability is determined only in small (theoretically, infinitely small) regions around the equilibrium point (either of motion or of a singular point).

In view of these features, the second method of Liapounoff offers a definite advantage as compared to the classical treatment of stability problems on the basis of the variational equations.

The whole question hinges thus on the point: can one determine the function V for a given differential system? This problem undergoes a continuous generalization, mostly on the part of the Russian mathematicians in recent years [33, 79]. It can be stated that, at present, this problem has been solved in the case of d.e. with constant as well as with periodic coefficients, i.e., in cases in which the theory of oscillations is primarily interested. Most of this work is directed toward the "inversion" of the original theorems of Liapounoff and, at present, there remain relatively few cases where the existence of the function V has yet to be proved. All this seems to indicate that the formal justification of the method is definitely in sight. In the meantime the applications of this method increase continuously and we merely refer to the existing literature (mostly in Russian) [33] regarding these developments.

CHAPTER 2

METHODS OF APPROXIMATIONS

1. Introductory remarks

The general methods outlined in the preceding chapter give a qualitative approach to problems of the theory of oscillations. For the quantitative results it is necessary to make use of the methods of approximations inasmuch as practically all nonlinear d.e. encountered in applications do not admit exact integration.

At the early stage of these developments the use was made of another treatise of Poincaré "Les méthodes nouvelles de la mécanique céleste," [26] which we call here (II), reserving (I) to his paper "*Sur les courbes définies par une équation différentielle*," [5] which constitutes the basis of methods outlined in Chapter 1. In the presentation of Poincaré there is no relation between (I) and (II), whereas for the development of the new theory of oscillations it became necessary to correlate these two independent developments.

A great deal of this work of correlation was due to the Russian physicists, particularly to Andronov [34]. In the following article we outline the method of Poincaré regarding the existence of a periodic solution of autonomous systems in a form given by Andronov with a view to adapting it primarily to the theory of oscillations. Two other methods of approximation appeared later; one due to van der Pol [35] and the other to Kryloff-Bogoliuboff [36]. Both these applications are based on the method resembling that of the variation of parameters. Originally, these methods did not go beyond the theory of the first approximation, but later the last-mentioned authors applied the astronomical method of Lindstedt [37] for the approximations of higher orders [38].

All these applications were adapted in the beginning mostly for the autonomous systems. Although for the nonautonomous systems some extensions were also made (particularly by Mandelstam and Papalexi), the question of stability remained a difficult point, to say nothing of the impossibility of using the phase-plane representation, as in this case the matter is outside the scope of (I).

Finally, in collaboration with M. Schiffer a method was developed (Chapter 2, Art. 4) which permits a formal reduction of nonautonomous systems to the form (1) in which the question of stability is reduced to that of a singular point, which simplifies nonautonomous problems as is shown on a number of examples treated in Chapter 3 [40].

In all these methods the smallness of the parameter μ (as, for instance, in equation (1) of the "Introduction") is essential.

2. Condition of periodicity (Poincaré) [40]

In what follows we consider an autonomous d.e. of the form

$$\ddot{x} + x = \mu f(x, \dot{x}) \tag{1}$$

where $f(x, \dot{x})$ is a certain nonlinear function of x, and $\dot{x}$ and μ is a parameter. The van der Pol equation (equation (1) of the "Introduction") is a particular case of (1) for $f(x, \dot{x}) = (1 - x^2)\dot{x}$.

Very often this type of equations is called "quasi-linear" or "nearly linear" when $\mu \ll 1$. Likewise, the method which will be outlined is frequently called also as "the method of small parameter (or parameters)" of Poincaré.

If $\mu = 0$, (1) becomes the d.e. of the harmonic oscillator which has a two parameter family of periodic solutions, as was previously explained in connection with Fig. 4.

The problem presents itself in the following manner: Under which conditions (1) may still have a periodic solution (i.e., a stationary oscillation) when $\mu \neq 0$ but small?

It is to be observed that the problem so formulated is by no means obvious since, for instance, if $f(x, \dot{x}) = b\dot{x}$, clearly, there is no periodic solution for any b however small. On the other hand, as (1) is "in the neighborhood" of the d.e. of the harmonic oscillator, its periodic solution (if it exists) is also in the neighborhood of the solution of the latter.

The special solution of the d.e. of the harmonic oscillator in the neighborhood of which *may* appear a periodic solution of (1), for $\mu \neq 0$, is called the *generating solution.* The crux of the problem is in formulating the conditions under which a periodic solution of (1) *may* exist.

If $\mu = 0$, the solutions of the d.e. of the harmonic oscillator can be written as

$$x = x_0(t, K), \qquad y = \dot{x} = y_0(t, K) \tag{2}$$

which merely expresses the fact that x and $\dot{x}$ besides t depend also on the amplitude K, i.e., on the initial conditions.

If a periodic solution exists when $\mu \neq 0$, one can represent it in the same manner, adding μ in brackets, but if, in addition, one wishes to consider the initial conditions as well, Poincaré introduces them by means of equations of definition:

$$(3) \qquad x(0, \mu, K) = x_0(0, K) + \beta_1, \qquad y(0, \mu, K) = y_0(0, K) + \beta_2$$

which means that the "neighborhood" of two equations (one for $\mu = 0$ and the other for $\mu \neq 0$) manifests itself by slight discrepancies β_1 and β_2 in the initial conditions of motions governed by these equations. The simplest would be to start the comparison from the same point for which one of β is zero and to write the condition of periodicity in the form

$$(4) \qquad \begin{aligned} x(T + \tau, \mu, 0, \beta, K) - x(0, \mu, 0, \beta, K) &= \phi(\tau, \mu, \beta, K) = 0 \\ y(T + \tau, \mu, 0, \beta, K) - y(0, \mu, 0, \beta, K) &= \psi(\tau, \mu, \beta, K) = 0 \end{aligned}$$

which means that after one period both x and $y = \dot{x}$ resume their original values but, as we now consider a nonlinear problem ($\mu \neq 0$), the original period T is also changed and is now $T + \tau$, τ being the *nonlinear correction for the period.**

It is to be noted, however, that if $\mu = 0$, one has also $\tau = 0$, $\beta = 0$, and also, obviously, $\phi \equiv 0$ and $\psi \equiv 0$. It is clear, therefore, that $\beta = \beta(\mu)$ and $\tau = \tau(\mu)$ are such that $\beta \to 0$ and $\tau \to 0$ when $\mu \to 0$. Besides this, μ is obviously a factor both in ϕ and in ψ so that one can write $\phi = \mu\phi_1$; $\psi = \mu\psi_1$, which takes care of the identical vanishing of ϕ and ψ for $\mu = 0$.

One reaches thus the following conclusion: if there are any periodic solutions when $\mu \neq 0$, this requires that

$$(5) \qquad \phi_1(\tau, \mu, \beta, K) = 0, \qquad \psi_1(\tau, \mu, \beta, K) = 0$$

On the other hand, in view of the above condition regarding the form of the functions $\tau(\mu)$ and $\beta(\mu)$, one can represent them by power series

$$(6) \qquad \tau(\mu) = d\mu + e\mu^2 + \cdots, \qquad \beta(\mu) = d_1\mu + e_1\mu^2 + \cdots$$

It is observed that there is one independent parameter μ and two others τ and β are functions of μ. If one expands (5) in Taylor's series,

* This is possible in the case of the autonomous systems possessing the translation property (in t). For nonautonomous systems, β is present in the equations but, on the contrary, $\tau = 0$ because the period is imposed owing to the external periodic excitation. The following concerns only autonomous systems.

keeping only the first order, one has

(7) $\phi_1 = \phi_{10} + a\mu + b\tau + c\beta = 0, \quad \psi = \psi_{10} + a_1\mu + b_1\tau + c_1\beta = 0$

Replacing for the first approximation $\tau(\mu) = d\mu$, and $\beta(\mu) = d_1\mu$ in (7), one gets

(8) $\phi_{10} + \mu(a + bd + cd_1) = 0, \qquad \psi_{10} + \mu(a_1 + b_1d + c_1d_1) = 0$

These two equations are to be solved for the unknowns d and d_1 for any arbitrary μ, provided it is small. This is possible only if $\phi_{10} = \psi_{10} = 0$ and

(9) $$a + bd + cd_1 = 0, \qquad a_1 + b_1d + c_1d_1 = 0$$

which requires that

$$\begin{vmatrix} b & c \\ b_1 & c_1 \end{vmatrix} \neq 0$$

but as $$b = \frac{\partial\phi_1}{\partial\tau}, \qquad c = \frac{\partial\phi_1}{\partial\beta} \cdots \text{etc.}$$

it is clear that the preceding condition requires that the Jacobian

(10) $$J = \partial(\phi_1, \psi_1)/\partial(\tau, \beta) \neq 0$$

Hence, if the functions ϕ_1 and ψ_1 do not contain terms independent of μ, τ, and β, and if the Jacobian $J \neq 0$, it is possible to fulfill the condition (5) of periodicity and, at the same time, determine the functions $\tau(\mu)$ and $\beta(\mu)$ to the first order.

In order to go beyond this point, Poincaré uses as formal solution a series arranged according to the ascending powers of the parameters μ, β_1, and β_2, namely,

(11) $$x = \phi_0(t) + A\beta_1 + B\beta_2 + C\mu + D\beta_1\mu + E\beta_2\mu + F\mu^2 + \cdots$$

It is important to note that this series solution proceeds according to the powers of *parameters* and the coefficients A, B, C, $\cdots$ are unknown *functions of time* and, from now on, the problem centers on the determination of these functions.*

It is clear that the series converges only if μ is sufficiently small, which is precisely the condition previously set forth.

The rest of the calculation is simple but too long to be produced here and we merely indicate the procedure as well as the final result. With the form (11), one calculates $\dot{x}$ and $\ddot{x}$ and develops $f(x, \dot{x})$ into Taylor's series around $f(x_0, \dot{x}_0)$, where $x_0 = \phi_0(t)$ and $y_0 = \dot{\phi}_0(t)$ are ultimately

* This important theorem appears in the first volume, p. 58, first reference [26].

the generating solutions. The identification of the coefficients of μ, β_1, β_2, $\beta_1\mu$, $\beta_{2\mu}$, $\cdots$ results in a number of d.e., namely,

$$(12)\quad \ddot{A} + A = 0, \qquad \ddot{B} + B = 0$$
$$\ddot{C} + C = f(x_0, \dot{x}_0), \qquad \ddot{D} + D = \left(\frac{\partial f}{\partial x}\right)_0 A + \left(\frac{\partial f}{\partial \dot{x}}\right)_0 \dot{A}$$
$$\ddot{E} + E = \left(\frac{\partial f}{\partial x}\right)_0 B + \left(\frac{\partial f}{\partial \dot{x}}\right)_0 \dot{B}, \qquad \ddot{F} + F = \left(\frac{\partial f}{\partial x}\right)_0 C + \left(\frac{\partial f}{\partial \dot{x}}\right)\dot{C}$$

where $\left(\frac{\partial f}{\partial x}\right)_0$ and $\left(\frac{\partial f}{\partial \dot{x}}\right)_0$ are partial derivatives of f with respect to x and $\dot{x}$ into which the generating solution $x_0 = \phi_0(t)$ and $y_0 = \dot{\phi}_0(t)$ is substituted after the differentiation.

From the definition of $\beta_1 = x - x_0$, $\beta_2 = \dot{x} - \dot{x}_0$, and the series (11), one obtains the initial conditions $A(0) = \dot{B}(0) = 1$, all other initial conditions being zero.

The integration of (12) is given by the usual formula

$$(13)\qquad v = \int_0^t V(u) \sin (t - u)\, du$$

where $V(u)$ are the functions on the right-hand side of (12). For the conditions of periodicity the integrals (13) are to be taken between 0 and 2π, which yields a sequence of functions, $C(2\pi)$, $\dot{C}(2\pi)$, $D(2\pi)$, $\dot{D}(2\pi)$, $\cdots$.

As the last step, one calculates the expressions $x(2\pi + \tau)$, $\dot{x}(2\pi + \tau)$ which give the expressions

$$(14)\qquad x(2\pi + \tau) - x(0) = C(2\pi)\mu + \cdots$$
$$\dot{x}(2\pi + \tau) - \dot{x}(0) = -K\tau + \dot{C}(2\pi)\mu + \cdots$$

In these expressions we have written only the terms of the first order in small parameters τ and μ, leaving out the second-order terms such as those with τ^2, μ^2, $\tau\mu$, $\cdots$.

If the condition of periodicity is to be determined to the first order, the expressions (14) yield the conditions

$$(15)\quad C(2\pi) = -\int_0^{2\pi} f(K \cos u, -K \sin u) \sin u\, du = \Phi(K) = 0$$
$$\tau = \dot{C}\frac{(2\pi)\mu}{K} = \frac{\mu}{K}\int_0^{2\pi} f(K \cos u, -K \sin u) \cos u\, du = \mu\Psi(K)$$

The first of these equations yields the stationary amplitude K_0 and the second gives the correction for the period, the generating solutions in this case being $x_0 = K_0 \cos t$, $\dot{x} = y = -K_0 \sin t$, K_0 being the root of $\Phi(K_0) = 0$.

If one applies this procedure to the van der Pol equation (1) of the "Introduction," one has $f(K \cos t, -K \sin t) = -(1 - K^2 \cos^2 t)K \sin t$, and the first equation (15) yields after a simple integration

$$C(2\pi) = K \int_0^{2\pi} \sin^2 t \, dt - K^3 \int_0^{2\pi} \sin^2 t \cos^2 t \, dt = K\pi\left(1 - \frac{K^2}{4}\right) = 0$$

Rejecting the solution $K = 0$ (which is a position of equilibrium), one obtains for the stationary solution $K = 2$, which is the well-known result.

If one carries out this calculation for the second equation, one finds that $\Psi(K) = 0$, which shows that in the first approximation the period remains the same as for the harmonic oscillator.

Thus, in the first approximation the periodic solution of the van der Pol equation is simply

$$x = 2 \cos t \tag{16}$$

If one wishes to go to higher approximations, more terms are to be taken in (14), but the calculations become too complicated to be reproduced here.

We shall not go into a further elaboration of this method inasmuch as it is seldom used in applications in view of its complexity. We have outlined it here because historically it was the first method which opened an entirely new approach to a systematic study of periodic solutions encountered in the theory of nonlinear oscillations.

3. Theory of the first approximation by the van der Pol-Kryloff-Bogoliuboff method *

These two methods in reality are identical. The only difference is in that van der Pol [41] tries to satisfy the d.e. (1) by a solution of the form $x = M \sin t + N \cos t$ and K.B. (Kryloff-Bogoliuboff) [42] assume the solution $x = a \sin (\omega t + \phi)$.

Since the problem is nearly linear, the method is obviously derived from the known procedure of a variation of parameters M, N (for the van der Pol case) and a, ϕ (for the K.B. case). As the K.B. form is a little more convenient, we outline it here using the notations of these

* The matter outlined here belongs to the early theory of K. B. In a recent theory [80], the so-called *asymptotic method*, the derivation is somewhat different; see Art. 7.

authors, who write (1) in the form

$$\ddot{x} + \omega^2 x + \mu f(x, \dot{x}) = 0 \tag{17}$$

Since in the process of "fitting" the solution $x = a \sin (\omega t + \phi)$ into (17) both a and ϕ are variable, if one imposes the requirement that the solution be of the form

$$x = a \sin (\omega t + \phi), \qquad \dot{x} = a\omega \cos (\omega t + \phi) \tag{18}$$

one has an additional condition

$$\dot{a} \sin (\omega t + \phi) + a\dot{\phi} \cos (\omega t + \phi) = 0 \tag{19}$$

If one differentiates the second equation (18) and replaces x, $\dot{x}$, and $\ddot{x}$ into (17), one has

$$\dot{a}\omega \cos (\omega t + \phi) - a\omega\dot{\phi} \sin (\omega t + \phi) + \mu f(a \sin (\omega t + \phi), a\omega \cos (\omega t + \phi)) = 0 \tag{20}$$

From (19) and (20) one gets

$$\frac{da}{dt} = -\frac{\mu}{\omega} f[\ ] \cos (\omega t + \phi), \qquad \frac{d\phi}{dt} = \frac{\mu}{a\omega} f[\ ] \sin (\omega t + \phi) \tag{21}$$

where by $f[\ ]$ we have designated the last term in (20). As μ is assumed to be small and the other terms are bounded, the functions $a(t)$ and $\phi(t)$ vary slowly and, as an approximation, one can assume that during one period T of the trigonometric functions a and ϕ remain constant but vary from one period to the other. This amounts to replacing $\dot{a}$ and $\dot{\phi}$ by expressions

$$\frac{a(t + T) - a(t)}{T}, \qquad \frac{\phi(t + T) - \phi(t)}{T} \tag{22}$$

On the other hand $f[\ ] \cos \gamma$ and $f[\ ] \sin \gamma$; ($\gamma = \omega t + \phi$) are periodic and can be represented by trigonometric series of the form

$$f[\ ] \cos \gamma = P_0(a) + \sum_{n=1}^{\infty} [P_n(a) \cos n\gamma + P_n'(a) \sin n\gamma] \tag{23}$$

$$f[\ ] \sin \gamma = Q_0(a) + \sum_{n=1}^{\infty} [Q_n(a) \cos n\gamma + Q_n'(a) \sin n\gamma]$$

As during one period T, the quantity a is assumed to be constant, the trigonometric series (23)* is approximated by the Fourier series and the integration of these expressions over the period eliminates all trigonometric terms.

The d.e. (21) are thus replaced by two sequences of the difference equations

$$\frac{a(t+T)-a(t)}{T} = -\frac{\mu}{\omega}P_0(a) \tag{24}$$

$$\frac{\phi(t+T)-\phi(t)}{T} = \frac{\mu}{a\omega}Q_0(a)$$

If the oscillatory phenomenon lasts long enough in terms of one period T, the left-hand sides of these expressions can be written as $\frac{\Delta a}{\Delta T}$ and $\frac{\Delta\phi}{\Delta T}$ respectively, and, at the limit, as $\frac{da}{dT}$, $\frac{d\phi}{dT}$, or $\frac{da}{dt}$, $\frac{d\phi}{dt}$, recalling that t here is not the same as in (21).

As to P_0 and Q_0, they are given by the usual Fourier procedure. One obtains thus the d.e. of *the first approximation* of the K.B. theory, namely:

$$\frac{da}{dt} = -\frac{\mu}{\omega}\frac{1}{2\pi}\int_0^{2\pi} f(a\sin\gamma, a\omega\cos\gamma)\cos\gamma\, d\gamma = \Phi(a) \tag{25}$$

$$\frac{d\gamma}{dt} = \omega + \frac{\mu}{2\pi a\omega}\int_0^{2\pi} f(a\sin\gamma, a\omega\cos\gamma)\sin\gamma\, d\gamma = \Omega(a) \tag{26}$$

The stationary amplitude a_0 is obtained from $\Phi(a_0) = 0$ and it is noted that this is precisely the same result as that given by the first equation (15) of the theory of Poincaré. As to the equation (26), the term with μ is the same as in the second equation (15). The only difference is in that the roles of $\sin\gamma$ and $\cos\gamma$ are interchanged because of a different selection of the generating solution. Although this method is much simpler than that of Poincaré, it is less general than the latter inasmuch as it concerns only the first approximation, whereas the higher approximations develop indefinitely in the Poincaré method. In reality, as we saw, the practical difficulties of obtaining higher approximations

* This series is, in fact, an almost periodic function (see Art. 7 below) but in the first approximation, it is approximated by a Fourier series, so that the question of almost-periodicity is not involved here.

in Poincaré's method are so great that the method ceases to be a useful tool in applications. In Art. 7 we indicate a recent work of Malkin which simplifies the use of Poincaré's method in applied problems.

At a later stage K.B. adopted the astronomical method of Lindstedt to obtain higher order approximations by their method [37].* In applications the need for approximations of higher orders is very remote inasmuch as the accuracy with which the various parameters (particularly the nonlinear ones) are known does not warrant generally this additional complication in calculations.

4. Nonautonomous systems

The methods of approximations outlined in the preceding two sections are applicable to quasi-linear d.e. of the form (1) not *containing the variable t (time) explicitly*. Such d.e. or differential systems are called often the *autonomous systems* in order to distinguish them from the *nonautonomous ones*, in which, on the contrary, t enters explicitly. Thus, for instance, if one adds to the d.e. (1) a term $a \sin t$ or $a \sin nt$, the new system becomes then nonautonomous.

In Chapter 1, Art. 1, we have indicated the general form of autonomous systems with one degree of freedom [equation (1), Chapter 1] and mentioned that the *implicit* dependence on t is easily eliminated by merely dividing the second equation (1) in Chapter 1 by the first one, which gives the d.e. (2) of Chapter 1 of integral curves. It was through this simple transformation that it was possible to reduce the problem to a pure geometry and thus to establish a contact with paper I of Poincaré dealing with the geometrical properties of integral curves. It is true, that, by taking into account (1) of Chapter 1 as well, one can still gain the insight into the motion of R on these curves (which are then "trajectories"), but this is merely a slight extension, not changing in any manner the topological aspect of the problem.

The advantages of a topological representation in the phase plane disappear however for nonautonomous systems. In fact, in this case, the d.e. are of the form

$$\dot{x} = P(x, y, t), \qquad \dot{y} = Q(x, y, t) \tag{27}$$

and it is obviously impossible to form (2) of Chapter 1 not containing t. This means that all concepts such as singular points, limit cycles, etc., become impossible for the nonautonomous systems.

* The above method of Kryloff and Bogoliuboff (1937) was generalized much later (1955) by Bogoliuboff and Mitropolsky [80] into what is known as the *asymptotic method* mentioned in Art. 7. In this general method approximations of higher orders result directly without any reference to Lindstedt's procedure.

One can see this circumstance also from the fact that, whereas the d.e. (1) of Chapter 1 of autonomous systems (for one degree of freedom) represent a system of two d.e. of the first order, a nonautonomous system (27) in reality reduces to *three* such d.e., namely,

$$\frac{dx}{dt} = P(x, y, z), \qquad \frac{dy}{dt} = Q(x, y, z), \qquad \frac{dz}{dt} = 1 \tag{28}$$

In other words the variable t may be regarded as the third dimension of the three-space (x, y, z), if one wishes to consider t as a parameter. If one disregards this circumstance and continues using the phase-plane representation in the case of nonautonomous systems, one may encounter absurd situations; for instance, an integral curve may intersect itself at some points, etc. In reality this does not happen if one takes into account the third dimension z.

In the case of nonautonomous systems it is impossible to use the topological representation (Chapter 1) and it is necessary to follow the general theory leading to the use of the characteristic exponents for the determination of stability conditions, which is generally a more complicated problem.

Such a course was followed, for instance, by Mandelstam and Papalexi in their studies of the nonlinear resonance [43]. It would be too long to go into this matter here, particularly because this constitutes the classical procedure in the theory of d.e. which can be found in any textbook on d.e. (see, for instance [*e*], Chapter VII), and also because this general procedure did not find any extended use in the theory of oscillations, mainly on account of the above-mentioned complications.

Inasmuch as we are interested here primarily in applied problems, it is preferable to outline a method which reduces formally a system of the form (27) to that of (1) of Chapter 1 on the basis of the transformation theory. Such a method was evolved recently [44] and, in Chapter 3, we shall indicate a number of examples of application of this method to the nonautonomous systems.

5. Stroboscopic method

We consider the d.e. (27) in polar variables ρ and ψ * defined by the relations $\rho = r^2 = x^2 + \dot{x}^2 = x^2 + y^2$; $\psi = \arctan\ (y/x)$ with $x = r\cos\psi$, $y = r\sin\psi$. This transformation is carried out easily by observing that $x\dot{x} + y\dot{y} = \frac{1}{2}(d\rho/dt)$ and $y\dot{x} - x\dot{y} = -\rho\dfrac{d\psi}{dt}$. One ob-

* The use of polar variables is not essential; however, it is convenient in the following discussion.

tains thus (27) in the form

$$\frac{d\rho}{dt} = F(\rho, \psi, t), \qquad \frac{d\psi}{dt} = G(\rho, \psi, t) \tag{29}$$

We assume that the functions F and G are periodic with period 2π and that the problem is nearly linear, i.e., the system (29) "is in the neighborhood" of that of the harmonic oscillator for which (29) becomes

$$\frac{d\rho}{dt} = 0, \qquad \frac{d\psi}{dt} = -1 \tag{30}$$

We note first that $\rho = x^2 + \dot{x}^2$ may be regarded (with a proper normalization and in an appropriate scale) as a measure of total energy stored in the oscillator represented by (29). The first equation (29) merely expresses the law of conservation of energy. As to the second equation (29), it shows that the angular velocity of the point R is constant and the rotation takes place clockwise (since positive or trigonometric direction for counting angles is counterclockwise). If one wishes to consider also a third dimension, as was explained in connection with (28), it is obvious that the trajectory is a circular helix H with a constant pitch proceeding along the t axis. The projection of H on the $(x, \dot{x})$ plane is a circle C. Nothing essential is gained in this manner except the representation in space of the trajectory of the harmonic oscillator usually considered in the plane $(x, \dot{x})$.

Since we wish to consider the quasi-linear problem, the d.e. (29) are in the neighborhood of (30), and one can write (30) in the form, assuming that the d.e. contain a parameter μ,

$$\frac{d\rho}{dt} = \mu f(\rho, \psi, t), \qquad \frac{d\psi}{dt} = -1 + \mu g(\rho, \psi, t) \tag{31}$$

since (31) becomes (30) for $\mu = 0$. One can apply the classical procedure for the integration of (31) by the series (Art. 2):

$$\rho(t) = \rho_0(t) + \mu\rho_1(t) + \mu^2\rho_2(t) + \cdots,$$

$$\psi(t) = \psi_0(t) + \mu\psi_1(t) + \mu^2\psi_2(t) + \cdots$$

which obviously gives here

$$\rho_0(t) = \rho_0, \qquad \psi_0(t) = \phi_0 - t \tag{32}$$

ρ_0 and ϕ_0 being some arbitrary constants. We follow now the usual perturbation procedure which consists in building up successive approximations starting with the *zero-order* approximation (32) which is here merely the solution for the harmonic oscillator.

The following (the first) approximation is then: $\rho(t) = \rho_0 + \mu\rho_1(t)$, $\psi(t) = \phi_0 - t + \mu\psi_1(t)$, where the corrective terms $\rho_1(t)$ and $\psi_1(t)$ are calculated as usually, namely,

$$(33) \qquad \rho_1(t) = \int_0^t f(\rho_0, \psi_0, t)\, dt, \qquad \psi_1(t) = \int_0^t g(\rho_0, \psi_0, t)\, dt$$

If one limits the solution to the first approximation only, it is clear that nothing indicates that such solution will actually converge to the true solution if $t \to \infty$. The reason for this is due to the presence of *secular terms* in higher order terms (with μ^2, μ^3, $\cdots$) left out in the first approximation.

In order to avoid this, we proceed as follows: instead of letting $t \to \infty$, we increase t from 0 to 2π and determine

$$(34) \qquad \rho_1(2\pi) = K(\rho_0, \phi_0), \qquad \psi_1(2\pi) = L(\rho_0, \psi_0)$$

This determines a transformation

$$(35) \qquad \rho' = \rho + \mu K(\rho, \phi), \qquad \phi' = \phi - 2\pi + \mu L(\rho, \phi)$$

which can be continued indefinitely as the functions f and g have period 2π, and the terminal conditions of one interval 2π become by (35) the initial conditions for the next interval. It is noted that the explicit dependence on t has already disappeared in this procedure and what plays the role of "time" is now the number of the subsequent iteration steps as the transformation (35) proceeds from one interval 2π to the following one. Moreover, the cumulative error caused by the terms with μ^2, μ^3 is avoided as the Poincaré series is used only during a limited time interval 2π so that the error does not accumulate as it does when $t \to \infty$.

The transformation (35) is nothing but two sequences of difference equations which can be written as

$$(36) \qquad \frac{\Delta\rho}{\Delta\tau} = K(\rho_0, \phi_0), \qquad \frac{\Delta\phi}{\Delta\tau} = L(\rho_0, \phi_0)$$

where $\Delta\rho = \rho' - \rho$, $\Delta\phi = \phi' - \phi$ and $\Delta\tau = \mu$, ρ_0, ϕ_0 being discrete values of ρ and ϕ at the beginning of each interval 2π.

It is clear that if the period 2π is short enough in comparison with the total duration of the process, one can more conveniently use continuous variables and, thus, pass to the limit of two d.e.

$$(37) \qquad \frac{d\rho}{d\tau} = K(\rho, \phi), \qquad \frac{d\phi}{d\tau} = L(\rho, \phi)$$

These d.e. are fundamental in the sequel and, for the reasons which will appear in the following article, we shall call the system (37), *the stroboscopic system.* It is noted that the system (37) is now of the form (1) of Chapter 1, so that the topological argument can be still used.

6. Stroboscopic image; planes (ψ) and (ϕ)

Call (ψ) the plane of the variables (ρ, ψ). In this plane the trajectory of the harmonic oscillator is a circle with some radius ρ_0 described by the point R with a uniform velocity $\dot{\psi} = -1$. For a neighboring non-

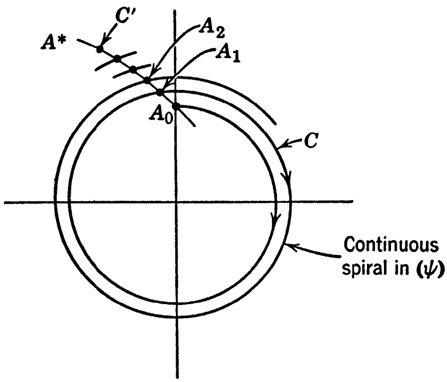


Fig. 10. Stroboscopic method; planes (ψ) and (ϕ).

linear d.e., this trajectory is a spiral C (Fig. 10) with a small decrement (increment), since μ is assumed to be small.

Instead of a continuous set of points defining C, the procedure outlined in Art. 5 considers a subset C' which consists only of the points separated in time by one period 2π. Clearly, for a harmonic oscillator, this subset reduces to one single point fixed in the plane (ϕ) which we may, for the sake of convenience, regard as superimposed on the plane (ψ). For the neighboring nonlinear oscillator this subset consists of a discrete sequence of points $A_0, A_1, A_2, \cdots$ (Fig. 10) since both the amplitude ρ and the phase ϕ vary slowly after each period 2π.

A simple physical interpretation is obvious, namely: if one illuminates the motion of R on C by the stroboscopic flashes separated in time by the period 2π, the sequence of points $A_0, A_1, A_2, \cdots$ is obviously the sequence which the eye sees under this condition. The polygon $A_0, A_1, A_2, \cdots$ approaches a curve C' at the limit when the period 2π becomes small enough in comparison with the total duration of the process. This is precisely the relation which connects the d.e. (36) with the d.e. (37) of the stroboscopic system. The curve C' appears thus in the plane (ϕ) which may be called the *stroboscopic plane.*

We reach thus the following simple relation between the d.e. (29) and the corresponding stroboscopic system (37). The trajectory of (29) is a continuous spiral curve C in the plane (ψ) and that of (39) is a polygon joining the points A_0, A_1, $\cdots$ approximated as a continuous curve C' of the stroboscopic plane (ϕ).

If the sequence A_0, A_1, A_2, $\cdots$ of "the stroboscopic points" in (ϕ) approaches a limit A^*, the latter point becomes a *fixed point* in (ϕ) or, which is the same, becomes a singular point of (37). In such a case, in the (ψ) plane, in each period and for the same phase, the trajectory passes always through the same point, which means that the motion is periodic with period 2π. On the basis of this intuitive argument we reach thus the following conclusion. *The existence of a stable singular point of (37) is the criterion for the existence of a stable periodic solution of the original system (31) or (27).**

In this manner a more complicated nonautonomous problem is reduced to a simpler autonomous problem.

In the following chapter we shall make use of this method in connection with a number of applied problems.

It may be useful to add that, when there exists a nonlinear frequency (or period) correction already in the first approximation, the limit point A^* moves (in ϕ) on a circle with a constant velocity $\Delta\omega$. In such a case, the fixed point is obtained in a plane (ϕ') rotating relatively to (ϕ) with the same velocity $\Delta\omega$, the latter thus constituting the so-called *nonlinear frequency correction.* It is clear that $\Delta\omega = \dfrac{\tau}{2\pi}$, τ being the correction for the period of the Poincaré theory.

7. Recent developments

It is useful to supplement the above review of quantitative methods by a brief mention of some recent advances in this field, among which the asymptotic theory of Bogoliuboff and Mitropolsky [80] represents an important generalization of the earlier approach of Kryloff and Bogoliuboff mentioned in Art. 3 of this chapter. In view of the fact that this work (together with Mitropolsky's extension of the method for nonstationary processes [81]) occupies nearly 700 pages, we mention only a few salient points of this important development.

* This heuristic conclusion was proved later (1952) by Gomory (in an unpublished communication) and still later by Minoru Urabe, *J. of Sci.*, Hiroshima University (1956).

Given a differential equation (d.e. for short) of the form

$$\ddot{x} + \omega^2 x = \mu f(\nu t, x, \dot{x}) \tag{38}$$

which is sufficiently general for nearly linear systems with one degree of freedom, its solution is sought in the form:

$$x = a \cos \psi + \mu u_1(a, \psi, \nu t) + \mu^2 u_2(a, \psi, \nu t) + \cdots + \mu^n u_n(a, \psi, \nu t) + \cdots \tag{39}$$

where a and ψ have the same meaning as in Art. 3, νt is the argument of the external periodic excitation and the u_i are unknown functions to be determined from a system of d.e.; these d.e. have a different form according to whether the oscillation is of a nonresonance form or, on the contrary, is a resonance oscillation. In the first case these d.e. are:

$$\dot{a} = \mu A_1(a) + \mu^2 A_2(a) + \cdots; \qquad \dot{\psi} = \omega + \mu B_1(a) + \mu^2 B_2(a) + \cdots \tag{40}$$

and in the second case, they are

$$\dot{a} = \mu A_1(a, \theta) + \mu^2 A_2(a, \theta) + \cdots; \qquad \dot{\psi} = (p/q)\nu + \mu B_1(a, \theta) + \mu^2 B_2(a, \theta) + \cdots \tag{41}$$

where the variable $\theta = \psi - (p/q)\nu t$, p and q being relatively prime small integers.

It is seen that the resonance oscillation is more complicated to deal with than a nonresonance one, since, in the latter case, no relative phase difference between the autoperiodic (free) and the heteroperiodic (forced) oscillations is to be considered, so that the variable θ does not appear in this case which simplifies the problem.

The series (39) is generally divergent, but its asymptotic character manifests itself in that, by breaking off the series at its nth term, the sum (39) remains bounded for all values of the angular variables ψ and θ (both with period 2π) and approximates the solution of (38) with the accuracy $0\ (\mu^{r+1})$. If μ is small enough, only two or three terms of the series are generally sufficient.

The problem consists in determining the functions u_i, A_i, and B_i in such a manner that x so determined is to be the desired approximation of the solution of (38).

If one replaces x, $\dot{x}$, and $\ddot{x}$ by their values and develops the function f around the generating solution, $x_0 = a \cos \psi$ and $\dot{a}_0 = -a\omega \sin \psi$, one obtains (after rather long calculations) a recursive system of d.e. from which the unknown functions u_i, A_i, and B_i are gradually determined.

In these calculations use is made of the conditions:

$$(42) \qquad \int_0^{2\pi} u_i(a, \psi) \cos \psi \, d\psi = 0; \qquad \int_0^{2\pi} u_i(a, \psi) \sin \psi \, d\psi = 0$$

in order to guarantee that the first approximation contains *the full* amplitude of the fundamental harmonic, which is important for the theory of the first approximation.

This asymptotic method was further extended by Mitropolsky for systems with slowly varying parameters. This is accomplished by introducing a second independent variable τ, the "slow time," and by carrying out calculations in terms of two independent variables: t (the time) and τ (the slow time). This generalization broadens considerably the scope of the method. Thus, for instance, it is customary to investigate the subharmonic resonance (see the next chapter) for a fixed value of the "detuning." It becomes now possible to investigate what happens when the region of resonance *is traversed* at a certain rate; clearly, the effect of the resonance is different in these two cases. Besides this, the modulated oscillations which were so far beyond the reach of the theory can be treated now on this basis inasmuch as it is sufficient to take as the scale of the slow time the period of modulation, and, as that of the time t, the period of the oscillation itself, and proceed on this basis with the approximations.

Besides its numerous applications, the method is interesting also on account of the connection between the periodic and almost-periodic phenomena which appear directly in the theory of approximations.

A little later (1956) appeared another treatise [79] devoted to further extensions of the theory of Poincaré. This work covers separately nonautonomous and autonomous systems and in each of these two cases investigates the nonresonance and the resonance oscillations with numerous examples. Here again almost periodic oscillations play an important role; as in the case of the periodicity, the almost-periodicity is characterized by an approach of the solution to the generating solution when $\mu \to 0$, but the criteria of existence here are different; Malkin shows that these criteria are generally connected with the existence of certain "critical" and "noncritical" roots in the characteristic equation. For autonomous systems a transformation of the independent variable of the form $t = \tau(1 + \alpha_1\mu + \alpha_2\mu^2 + \cdots)$ considerably simplifies calculations and permits avoiding a previously existing difficulty when a series approximating a periodic solution with period 2π has as its coefficients periodic functions of other (generally incommensurate) periods. Likewise, it is shown that the method of Poincaré becomes much simpler if one takes as generating solution certain special functions introduced by Liapounoff. All these modifications render the theory of Poincaré developed originally for celestial mechanics better adapted for the theory of oscillations. A considerable amount of developments relative to the mathematical part of the theory of oscillations was published by Prof. S. Lefschetz's Seminar [83].

Aside these basic developments in the theory of oscillations itself, the nonlinear theory finds a gradually extended use in numerous applications, of which the theory of nonlinear control systems is perhaps most important at present. Here the situation is less definite and one can only note certain trends without being able to form any definite opinion. In this particular field the problems are just opposite to those of the theory of oscillation. In fact, the latter is primarily interested in establishing conditions for the existence of stable stationary oscillations. In the non-linear control theory the aim is to avoid such oscillations which are always of undesirable (parasitic) nature. The question of stability is, on the contrary, of a fundamental importance. Two distinct trends are noticed in the last question.

In one of them one tries to generalize the Nyquist diagram which turned out to be such a valuable tool for the investigation of linear systems with a feedback; the aim, therefore, is to try to adapt it for nonlinear systems.

The other trend is to deal directly with the d.e. of the control system, and to use the second method of Liapounoff for the determination of stability which, as we saw already, does not require integration of the variational equation, which is always a difficult problem and, very often, even an impossible one.

Each of these two trends has certain advantages and disadvantages.

As regards the generalization of the Nyquist diagram, the difficulty seems to be in the theoretical justification of this generalization. In fact, this diagram has been established on the basis of the theorem of Cauchy (linear d.e. in the complex domain), and it is desired to transfer the results into the real domain. In recent years some physicists adopted a purely heuristic approach by considering the Nyquist diagram as expressing certain physical facts and disregarding the above-mentioned mathematical difficulties. If one proceeds on this heuristic basis, certain conclusions are immediate: namely, in the linear case the amplitude and frequency do not stand in any relation to each other, whereas in a nonlinear case, they are always interrelated. This means that, instead of one single diagram (linear case), in a nonlinear case there exists always *a family* of Nyquist's diagrams depending on amplitude as parameter of the family. This assumption turned out apparently to be correct inasmuch as the majority of nonlinear phenomena established on the basis of the classical theory (next chapter) could be also traced out on the basis of the Nyquist diagram so generalized, but a justification of this fact is lacking at present.

The other trend proceeding along the line of the classical theory and using the second method of Liapounoff for the investigation of stability leads also to similar results in a more straightforward manner, but the procedure is generally complicated in view of several degrees of freedom usually present.

In addition to these two approaches to the problems of nonlinear control systems, there is also a tendency to consider a control system as an *operator* operating on the d.e. of the original (uncontrolled) system and transforming it into the d.e. of the system under control. For linear systems this amounts to the introduction of a *linear operator* which modifies the coefficients of the d.e., leaving its form unchanged. On this basis a nonlinear control system would amount to the introduction of a *functional* (nonlinear operator) changing *the form* of the d.e. It is seen that in the latter case there is a far greater number of possibilities which have not been yet investigated.

In the "Introduction" we mentioned some new problems in which connection with the mathematical analysis is either far from being well established or missing entirely. It is likely that problems of this nature will result in the corresponding extensions in the mathematical methods. Thus, for instance, in the theory of relaxation oscillations where the analytical procedure breaks down, there appears a gradually increasing trend to use the asymptotic expansions, but, here, the complications are such that the method is still outside the reach of applied problems.

This somewhat confused situation indicates that the theory of oscillations has not reached yet the stage of its final codification in spite of its enormous progress during the last two decades.

CHAPTER 3

OSCILLATIONS IN NEARLY LINEAR SYSTEMS

1. Introductory remarks

In this chapter are reviewed the principal oscillatory phenomena of the "nearly linear" type in which case the theory of Poincaré offers a convenient tool for their treatment.

The arrangement of this material by the phenomena is convenient in applications and, historically, the progress in this field was accomplished precisely in this manner. It must be noted, however, that from a mathematical point of view the situation appears in a somewhat different way. In fact, generally, there exists no definite single-valued relation between a certain "phenomenon" and the d.e. which governs it inasmuch as it frequently occurs that the same d.e. may account for several "phenomena" according to the various regions of its parameter space.

In order to make this statement more definite, we may consider, for instance, the van der Pol equation in its general form with a "forcing term," namely:

$$\ddot{x} + \mu f(x)\dot{x} + x = a \sin \omega t \tag{1}$$

where $f(x)$ is a function satisfying certain conditions and μ is a parameter.

We may consider different "phenomena" according to the various values of parameters of this d.e.:

1. If $a = 0$, (1) is an autonomous system the behavior of which has been already investigated previously.
2. If $a \neq 0$ and ω is an integer (say 2, 3, $\cdots$), (1) yields the phenomenon of the subharmonic resonance (Art. 4 below).
3. If $a \neq 0$ and ω is in *the neighborhood* of an integer value, another nonlinear phenomenon—the synchronization—takes place. This phenomenon consists in that the free frequency (of the left-hand term) is

brought exactly to the nearest subharmonic value. As a matter of fact, the two physical phenomena 2 and 3 appear merely as two *aspects* of a more complicated "mathematical phenomenon." One of the aspects concerning the amplitude yields phenomenon 2 and the other concerning the phase yields phenomenon 3.

4. If $a \neq 0$ and ω is very large, otherwise quite arbitrary, one comes across another phenomenon—the asynchronous quenching which consists in suppressing of the "free" oscillation by the "forced" one.

The above considerations determine to some extent the attitude of investigators. In purely mathematical work the various "phenomena" appear often merged together and become more distinct when the d.e. are expressed in terms of the amplitude and the phase, as was just mentioned in connection with the combined effects of the resonance and synchronization. In the theory of oscillations proper, it is still customary to proceed with the investigation of the "phenomena," and in this chapter we adopt this procedure.

As to the material contained in this chapter, Art. 2 relates to the theory of synchronization as was originally formed by van der Pol [45] with the topological extension by Andronov and Witt [46]. In Arts. 3, 4, and 5 the phenomena of parametric excitation, nonlinear resonance, and asynchronous action respectively are treated.

Article 4, as just mentioned, deals with the theory of subharmonic resonance. The early theory of these phenomena is due to Mandelstam and Papalexi [47] who treated it directly as a nonautonomous problem, which resulted in somewhat complicated stability calculation. Moreover, the theory was applied mainly to the potentially self-excited systems. There exists also another approach to this problem through the iteration procedure but this was done in connection with problems amenable to Duffing's equation which is of a nonself-excited type; we mention only briefly this method as it is outlined in a great detail in Stoker's book reference [*c*]. Finally, quite recently the stroboscopic method was used with a view to applying it to these questions; a brief outline of this method was given in the preceding chapter.

Article 6 relates to the oscillation produced by retarded actions which are frequently encountered in control systems. Strictly speaking this matter is outside the scope of the general methods (of Poincaré), and we have included it here only because in an important special case when the time lag is small, it can still be considered within the scope of the nearly linear theory. In the general case, this problem has not been solved as yet as it leads to the nonlinear difference-differential equations the theory of which did not progress to a point at which it could be used in applications.

2. Synchronization

The phenomenon of synchronization (or of "entrainment" of frequency) is perhaps the best known among the others. In fact, at present this is the only nonlinear phenomenon the study of which progressed sufficiently so that it can be used in engineering applications.

It is interesting to note that this phenomenon was apparently known to Huyghens [48] who reports that two clocks slightly out of synchronism, when suspended on a wall, become synchronized when fixed on a thin wooden board. More than three centuries elapsed, however, before the modern theory of this effect was formulated by van der Pol [49] in connection with electron-tube circuits.

This phenomenon can be best demonstrated in the following way. If one impresses on the grid of an electron-tube oscillator oscillating, say, with frequency ω_0 an electromotive force (emf) of frequency ω, the oscillation exhibits an interference or "beats" of the two frequencies in accordance with the linear theory. One may expect, therefore, that, if $\omega \to \omega_0$, the period of the beats should increase indefinitely. This is true, however, only up to a certain limit of the difference $|\omega - \omega_0|$, for which the beats disappear suddenly and there remains only one frequency ω. Everything happens as if the "free" frequency ω_0 of the oscillator were "entrained" by the extraneous frequency ω.

A simple application of Kirkhoff's law to such a circuit with a cubic approximation for the characteristic of the tube results in the d.e.

$$\ddot{v} - \alpha\dot{v} + \gamma\dot{v}^3 + \omega_0^2 v = B\omega_0^2 \sin \omega t \tag{2}$$

where v is the grid voltage, α and γ are the coefficients of the polynomial representing the nonlinear characteristic, and B is the amplitude of the external emf.

In accordance with his theory (Chapter 2, Art. 3), van der Pol assumes the solution of (2) in the form

$$v = b_1 \sin \omega t + b_2 \cos \omega t \tag{3}$$

where $b_1(t)$ and $b_2(t)$ are slowly varying functions of time if ω is sufficiently near to ω_0. If one substitutes (3) into (2) and neglects the terms with $\ddot{b}_1$ and $\ddot{b}_2$, in view of their smallness, one obtains the system

$$\begin{aligned} 2\dot{b}_1 + zb_2 + \alpha b_1(1 - b^2/a_0^2) &= 0 \\ 2\dot{b}_2 - zb_1 - \alpha b_2(1 - b^2/a_0^2) &= -B\omega_0^2 \end{aligned} \tag{4}$$

where $z = 2(\omega_0 - \omega)$; $b^2 = b_1^2 + b_2^2$; $a_0^2 = 4\alpha/3\gamma$. This system reduces to the form (1) of Chapter 1, namely,

$$\frac{db_1}{dt} = P(b_1, b_2), \qquad \frac{db_2}{dt} = Q(b_1, b_2) \tag{5}$$

If $P(b_1, b_2) = Q(b_1, b_2) = 0$, this system has a singular point, but from the fact that b_1 and b_2 are constant, it also follows that the oscillation takes place with only one frequency ω without any beats as is seen from (3). Thus the existence of a synchronized state is characterized by the corresponding existence of a singular point of (5) and this led Andronov and Witt [46] to an interesting topological analysis of the synchronization effect.

These authors transform (4) by setting $x = b_1/a_0$; $y = b_2/a_0$; $a = z/\alpha$; $A = -B\omega_0/\alpha a_0$; $r^2 = x^2 + y^2$; $\tau = t\alpha/2$, which gives the system

$$\frac{dx}{d\tau} = x(1 - r^2) - ay, \qquad \frac{dy}{d\tau} = ax + y(1 - r^2) + A \tag{6}$$

Following the procedure outlined in Chapter 1, Art. 1, one obtains the coordinates of the singular point of (6), namely,

$$x_0 = -\alpha\rho/A, \qquad y_0 = -\rho(1 - \rho)/A \tag{7}$$

where $\rho = r_0{}^2$. It is sufficient, therefore, to transfer the origin to the point (x_0, y_0) and to reduce the system to the canonical form. This gives the characteristic equation (15) of Chapter 1 of the form

$$S^2 - 2(1 - 2\rho)S + [(1 - \rho)(1 - 3\rho) + a^2] = 0 \tag{8}$$

The two equations (7) can be combined into one, namely,

$$a\rho^2 + \rho(1 - \rho)^2 = A^2 \tag{9}$$

If one considers a and ρ as variables and A as a parameter, (9) determines a family of cubic curves shown in Fig. 11. A simple discussion of the nature of the roots of (8) [compare with equation (15) of Chapter 1] permits ascertaining the nature of singular points in the various regions of the plane (a, ρ) and, therefore, the possibility of synchronization for the various values of the parameters. As the full discussion of the Andronov-Witt analysis can be easily found in the textbooks on this subject (reference [a], Chapter XVIII), we omit it here inasmuch as the principal point of the theory is already sufficiently clear, once the system (5) is reached.

It is interesting to note that the phenomenon of synchronization is not limited to the electron-tube circuits. Phenomena of an acoustic synchronization of musical pitches have been also ascertained [50] as well as those of a purely mechanical synchronization. In this latter connection the important work of Haag [51] must be mentioned. This author studied these phenomena in clocks in great detail, that is, precisely in the domain where synchronization effect was discovered by Huyghens more than three centuries ago.

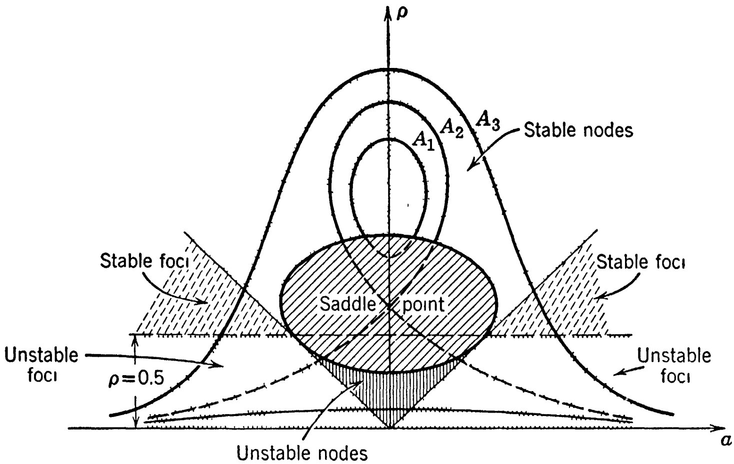


Fig. 11. Distribution of singular points in the case of synchronization.

This interesting effect gave rise to a number of schemes for producing an *artificial synchronization* of a sufficient intensity to be able to secure an accurate control of speed of electric motors synchronized with a source of a constant frequency (e.g., quartz oscillators). Thus, in this particular case, the study of the phenomenon has progressed so far that not only its analysis was completely explored but the problem of synthesis (i.e., the engineering application) becomes also possible.

3. Parametric excitation

Another nonlinear phenomenon explored recently on the basis of the new theory is the so-called *parametric excitation*.

A simple demonstration of this effect can be made by attaching a rope at one end and by pulling the other end at a certain frequency; it is observed that the rope starts oscillating *laterally* in such a case. Lord Rayleigh produced [52] this experiment by attaching a string A to a prong of a tuning fork B the oscillations of which with frequency f were maintained by an electromagnet (Fig. 12). It is observed that A begins to oscillate laterally with frequency $f/2$.

Similar phenomenon can be also produced electrically, as was shown by Mandelstam and Papalexi [53]. These authors provided means for changing periodically one of the parameters (either L or C) of an oscillating circuit. If the frequency of the circuit is f and that of the parameter variation $2f$, the phenomenon is particularly well marked. Owing to the presence of residual charges in condenser, a small initial electrostatic energy is sufficient to start the oscillation with increasing amplitudes. If the circuit is linear, the oscillation grows up indefinitely until the circuit is ultimately destroyed by the breakdown of insulation. If, however, a nonlinear conductor is provided in series with the circuit, the amplitude reaches a stationary value.

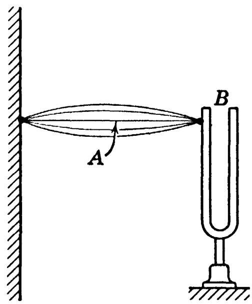


Fig. 12. Lord Rayleigh's experiments.

From a physical point of view, the energy is put into the system through the degree of freedom associated with the parameter variation and reappears with a half-frequency in the principal degree of freedom.

These phenomena are amenable to the d.e. of Mathieu and, according to whether the system is linear or nonlinear, they have different aspects. In the first case the problem can be treated rigorously in terms of the Mathieu functions but, in the second, it is necessary to make use of the theory of approximations.

As the d.e. of Mathieu contains t explicitly, it is convenient to treat both cases by the stroboscopic method which, at the same time, permits to gain an insight into the physical nature of this effect [54].

Consider a d.e. of Mathieu in a reduced form

$$\ddot{x} + (1 + a \cos \omega t)x = 0 \tag{10}$$

where a is small.

It is well known that any linear d.e. of the second order with a periodic coefficient in the last term can be reduced to this form by proper transformations of the dependent as well as the independent variables. Written as a system of two equations and with the polar variables $\rho = r^2 = x^2 + \dot{x}^2 = x^2 + y^2$; $\psi = \arctan (y/x)$; $x = r \cos \psi$; $y = r \sin \psi$, (10) is replaced by the system

$$\frac{d\rho}{dt} = -a\rho \cos 2t \sin 2\psi, \qquad \frac{d\psi}{dt} = -1 - a \cos 2t \cos^2 \psi \tag{11}$$

In these equations we assume $\omega = 2$ since the phenomenon is particularly well marked in such a case.* Since a is assumed to be small,

* This can be also proved by calculation but would be too long to be presented here.

the approximation of the zero order [see equation (32), Chapter 2] is here $\rho_0(t) = \rho_0$; $\psi_0(t) = \phi_0 - t$.

The d.e. for the first-order corrections $\rho_1(t)$ and $\psi_1(t)$ are then

$$\frac{d\rho_1}{dt} = -\rho_0 \sin 2\phi_0 \cos^2 t, \tag{12}$$

$$\frac{d\psi_1}{dt} = -\frac{1}{2}\cos 2t - \frac{1}{2}\cos 2t \cos 2(\phi_0 - t)$$

and the increments of $\rho_1(t)$ and $\psi_1(t)$ during one period 2π are

$$\rho_1(2\pi) = -\tfrac{1}{2}\rho_0 \sin 2\phi_0 \cdot 2\pi, \qquad \psi_1(2\pi) = -\tfrac{1}{4}\cos 2\phi_0 \cdot 2\pi$$

The corresponding variations for $\rho(2\pi)$ and $\psi(2\pi)$ are obtained by multiplying these expressions by a (which plays the role of μ of Art. 5, Chapter 2) since $\rho(t) = \rho_0 + a\rho_1(t)$. If one sets $2\pi a = \Delta\tau$ and writes $\Delta\rho$ and $\Delta\phi$ instead of $\rho(2\pi)$ and $\psi(2\pi)$, one obtains the difference equations [equations (36), Chapter 2] in the form

$$\frac{\Delta\rho}{\Delta\tau} = -\frac{1}{2}\rho_0 \sin 2\phi_0, \qquad \frac{\Delta\phi}{\Delta\tau} = -\frac{1}{4}\cos 2\phi_0 \tag{13}$$

so that the stroboscopic system [equations (37), Chapter 2] is here

$$\frac{d\rho}{d\tau} = -\frac{1}{2}\rho \sin 2\phi_0, \qquad \frac{d\phi}{d\tau} = -\frac{1}{4}\cos 2\phi_0 \tag{14}$$

From the second equation (14) it is seen that the phase ϕ has the point of equilibrium for $\cos 2\phi_0 = 0$, that is, for $\sin 2\phi_0 = \pm 1$. The variational equation in this case yields $d\delta\phi/dt = \frac{1}{2}\sin 2\phi_0$ which shows that the phase is at the point of a stable equilibrium if $\sin 2\phi_0 = -1$. In such a case the first equation (14) shows that the amplitude grows beyond any bound, which is in accordance with the Mandelstam-Papalexi experiment for a linear system.

More interesting is, however, the nonlinear case as, according to the above experiment, a stationary periodic oscillation with a finite amplitude is then possible.

We start therefore with a nonlinear Mathieu equation of the form

$$\ddot{x} + b\dot{x} + (1 + a\cos 2t)x + ex^3 = 0 \tag{15}$$

It is to be noted that the classical linear transformation by which the term $b\dot{x}$ is removed in the linear case does not apply here so that the term $b\dot{x}$ is to be retained; the last term ex^3 is a nonlinear term which, in this case, represents the effect of the nonlinear conductor.

We assume again that all three parameters a, b, and e are small in order to be within the limits of the nearly linear theory.

If one applies the method of Art. 5, Chapter 2, one obtains the following stroboscopic system

$$(16) \quad \frac{d\rho}{d\tau} = -\frac{1}{2}\rho(A \sin 2\phi + 2B), \qquad \frac{d\phi}{d\tau} = -\frac{1}{4}\left(A \cos 2\phi + \frac{3}{2}E\rho\right)$$

where $A = a/\mu$, $B = b/\mu$, and $E = e/\mu$, μ being the parameter of the series solution both for ρ and ϕ.

The system (16) admits a singular point given by relations

$$(17) \qquad \sin 2\phi_0 = -2B/A, \qquad \cos 2\phi_0 = -3E\rho_0/2A$$

As $|\sin 2\phi_0| < 1$, one must have $2B \leq A$. On the other hand, from $\sin^2 2\phi_0 + \cos^2 2\phi_0 = 1$ and (17), one gets

$$(18) \qquad \rho_0 = 2\sqrt{(A^2 - 4B^2)/9E^2}$$

The reality of ρ_0 requires again the condition $2B \leq A$.

For the investigation of stability, one forms the variational equations by replacing ρ and ϕ in (16) by $\rho_0 + \delta\rho$, $\phi_0 + \delta\phi$, where ρ_0 and ϕ_0 are the coordinates of the singular point given by (17) and (18) and $\delta\rho$ and $\delta\phi$ are small perturbations. If one carries this calculation (Art. 1, Chapter 1) and forms the characteristic equation, one finds easily that the roots are stable if $2B \leq A$.

Thus, if

$$(19) \qquad 2B \leq A$$

the d.e. (15) has a stable stationary solution with amplitude given by (18) and phase given by $\phi_0 = \arctan(4B/3E\rho_0)$.

A still more general form of the d.e. of parametric excitation is

$$(20) \qquad \ddot{x} + b\dot{x} + x + (a - cx^2)x \cos 2t + ex^3 = 0$$

of which (15) is a special case when $c = 0$. This equation gives rise to a number of special cases investigated recently [55], but in its general form it represents an interesting case of a parametric excitation known as "phenomenon of Bethenod" [56] which can be specified as follows: if one provides a physical pendulum with a piece of soft iron on its lowest part and places a coil coaxial with the pendulum in its equilibrium position, the pendulum begins to oscillate and reaches a certain stationary amplitude.

The phenomenon of parametric excitation has one special feature worth mentioning. In its normal form the phenomenon *always appears* as soon as the parameter begins to oscillate. As it was mentioned already, this situation may be characterized, from the standpoint of energy relations, by the fact that the energy is *always absorbed* in the degree of freedom of the parameter, and is introduced into the principal degree of freedom where the oscillation appears with the frequency twice smaller than that with which the parameter varies. This effect, as we saw, is due to the stability of the phase for such a value, for which the amplitude is unstable.

One may visualize an *inverse parametric effect* which would consist in transferring the energy of the existing oscillation (in the principal degree of freedom) into the degree of freedom of the parameter where it might be dissipated somehow (e.g., in friction). If such inverse effect could be produced, it is clear that, instead of a *parametric generation*, as is always the case, one would have a kind of a *parametric damping*. The only reason why no such effect has been discovered so far is due to the *instability of the phase* in this case. In fact, in the *natural* parametric effect the stable phase corresponds always to $\sin 2\phi_0 = -1$ (i.e., $\phi_0 = 3\pi/4$), the phase corresponding to $\sin 2\phi_0 = +1$ (i.e., $\phi_0 = \pi/4$) being unstable. If, however, by means of a special arrangement (e.g., an electron-tube circuit with a phase-shifting network) the unstable phase $\phi_0 = \pi/4$ could be made *artificially stable*, a device would work as a parametric damper.

For the time being the parametric actions appear merely as curious effects, but it is quite probable that their further study may lead to some useful applications.

4. Nonlinear resonance

Phenomena of resonance in nonlinear systems appear in a great variety of forms of which only relatively few have been explored completely.

The simplest way of approaching this subject is through ascertaining the existence of the so-called *subharmonics*.

If a linear system is acted on by an external periodic force of frequency ω, it is well known that, in the response of the system, in addition to the oscillation with frequency ω, appear also *harmonics* of higher frequencies 2ω, 3ω, $\cdots$. In the nonlinear systems, in addition to these, lower frequencies (e.g., $\omega/2$, $\omega/3$, $\cdots$) may occur occasionally, and these are called *subharmonics*.

Helmholtz was the first to ascertain this circumstance in connection with the problem of the physiological hearing [57] by noting that the ear sometimes hears the sounds which are not contained in the incoming acoustic radiation, and he explained this on the basis of the nonlinearity of the tympanic membrane.

It is easier to ascertain this fact by considering the nonlinearity of an electron tube. Assume that the nonlinear characteristic of the tube $i_a = f(v)$, (i_a is the plate current, v is the grid voltage) is approximated by a cubic polynomial

$$i_a = a_1 v + a_2 v^2 + a_3 v^3 \tag{21}$$

and that $v = K(\cos \omega_1 t + \cos \omega_2 t)$, i.e., it consists of two oscillations of the same amplitude K but of different frequencies. If one replaces this value of v into (21) and carries out the trigonometric transformations so as to have only the terms with multiple frequencies, one finds that they are of the form

$$\omega_1,\ \omega_2,\ 2\omega_1,\ 2\omega_2,\ 3\omega_1,\ 3\omega_2\ \|\ \omega_1 + \omega_2,\ \omega_1 - \omega_2,\ 2\omega_1 + \omega_2,\ 2\omega_1 - \omega_2,\ 2\omega_2 + \omega_1,\ 2\omega_2 - \omega_1$$

The terms to the left of the bar are harmonics and those to the right are the so-called *combination tones;* those with frequencies smaller than the smallest of the two fundamental frequencies (ω_1 or ω_2) are precisely the *subharmonics*. The reader can easily ascertain this by taking different polynomials as well as different values of ω_1 and ω_2.

If one wishes to have only a superficial insight into the problem of nonlinear resonance, the matter is very simple: if the subharmonics are known, it is easy to see that, if conditions for exciting a particular subharmonic are favorable, a corresponding oscillation will develop, dominating other oscillations and this constitutes the phenomenon in question.

In reality the theory of subharmonic resonance is more complicated than that of other nonlinear phenomena, mostly on account of long and tedious stability calculations. The fundamental investigation concerning the theory of the subharmonic resonance was carried out by Mandelstam and Papalexi [58] and is based on an extension of the Poincaré method for nonautonomous systems. This approach is rather simple as far as the *existence* of a subharmonic solution is concerned, but is more involved in the establishment of the *stability* condition. As it is impossible to give here a complete account of this theory, we shall limit ourselves to outlining its salient points.

In the method of Poincaré for the autonomous systems (Art. 2, Chapter 2), one tries to ascertain the existence of a generating solution in the neighborhood of which a unique nonlinear solution exists and is stable. It is logical, therefore, to attempt to apply the same argument to a nonautonomous d.e. of the form

$$\ddot{x} + a = \mu f(x, \dot{x}) + \lambda_0 \sin nt \tag{22}$$

for which it is proposed to establish the existence of the subharmonic solution with period 2π.

If $\mu = 0$, (22) has an infinity of solutions of the form

$$x(t) = a_0 \sin t + b_0 \cos t + \frac{\lambda_0}{1 - n^2} \sin nt \tag{23}$$

where the first two terms represent the free oscillation and the third the forced one, a_0 and b_0 being the integration constants.

For the neighboring nonlinear problem ($\mu \neq 0$ but small), it is necessary to determine $a(\mu)$ and $b(\mu)$ such that $a(\mu) \to a_0$ and $b(\mu) \to b_0$, when $\mu \to 0$. If it is possible to determine the functions $a(\mu)$ and $b(\mu)$ having this property, the solution $x(t)$ is called the *principal solution* and it is clear that such a solution is unique.

Through a double transformation of variables, namely,

$$z = x + \frac{\lambda_0}{1 - n^2} \sin nt \quad \text{and} \quad u = \dot{z} \cos t + z \sin t, \tag{24}$$

$$v = \dot{z} \sin t - z \cos t$$

(22) is reduced to a system

$$\dot{u} = \mu\psi(u, v, t) \cos t, \qquad \dot{v} = \mu\psi(u, v, t) \sin t \tag{25}$$

where

$$\psi(u, v, t) = f\left(u \sin t - v \cos t + \frac{\lambda_0}{1 - n^2} \sin nt, \right.$$
$$\left. u \cos t + v \sin t + \frac{n\lambda_0}{1 - n^2} \cos nt\right)$$

$$x(t) = u \sin t - v \cos t + \frac{\lambda_0}{1 - n^2} \sin nt \tag{26}$$

It is noted that (26) has the form (23), so that the problem consists now to express the conditions that $u(\mu) \to a_0$, $v(\mu) \to b_0$ when $\mu \to 0$. One can now follow the argument of Poincaré (Art. 2, Chapter 2), assuming that

$$u_{t=0} = a_0 + \alpha, \qquad v_{t=0} = b_0 + \beta \tag{27}$$

α and β being small numbers which are to be determined by this procedure [compare with d and d_1 in equation (8), Chapter 2]. The integration of (25) yields

$$u(t) = u(0) + \mu \int_0^t \psi(u, v, \tau) \cos \tau \, d\tau,$$

$$v(t) = v(0) + \mu \int_0^t \psi(u, v, \tau) \sin \tau \, d\tau$$

In view of (27) one can develop $\int_0^t \psi(u, v, \tau) \cos \tau \, d\tau$ and $\int_0^t \psi(u, v, \tau)$ $\sin \tau \, d\tau$ around a_0 and b_0, the series proceeding according to the powers of α, β, and μ.

On the other hand, these expansions are at the same time the series of Poincaré of the form

$$(28) \quad \begin{aligned} u &= a_0 + \alpha + \mu\, C_1(t) + \mu\alpha\, D_1(t) + \mu\beta\, E_1(t) + \mu^2\, G_1(t) + \cdots \\ v &= b_0 + \beta + \mu\, C_2(t) + \mu\alpha\, D_2(t) + \mu\beta\, E_2(t) + \mu^2\, G_2(t) + \cdots \end{aligned}$$

and the identification of the expansion with the series solutions (28) yields

$$(29) \quad C_1(t) = \int_0^t \psi(a_0, b_0, \tau) \cos \tau \, d\tau, \qquad C_2(t) = \int_0^t \psi(a_0, b_0, \tau) \sin \tau \, d\tau$$

$$D_1(t) = \int_0^t [\psi_u] \cos \tau \, d\tau, \qquad D_2(t) = \int_0^t [\psi_u] \sin \tau \, d\tau$$

$$E_1(t) = \int_0^t [\psi_v] \cos \tau \, d\tau, \qquad E_2(t) = \int_0^t [\psi_v] \sin \tau \, d\tau$$

where the quantities $[\psi_u]$ and $[\psi_v]$ designate the partial derivatives of ψ with respect to u and v into which one sets $\mu = \alpha = \beta = 0$ after the differentiation (compare with notations $\left(\frac{\partial f}{\partial x}\right)_0$, $\left(\frac{\partial f}{\partial \dot{x}}\right)_0$ in Art. 2, Chapter 2).

The condition of periodicity follows from (28) and is

$$(30) \quad \begin{aligned} C_1(2\pi) + \alpha\, D_1(2\pi) + \beta\, E_1(2\pi) + \mu\, G_1(2\pi) &= 0 \\ C_2(2\pi) + \alpha\, D_2(2\pi) + \beta\, E_2(2\pi) + \mu\, G_2(2\pi) &= 0 \end{aligned}$$

This system with the unknowns α and β is similar to equation (8), Chapter 2, so that the argument remains the same. Inasmuch as it is desired to determine $\alpha = \alpha(\mu)$ and $\beta = \beta(\mu)$—which can be taken for instance in the form $\alpha(\mu) = K_1\mu$; $\beta(\mu) = K_2\mu$ for the first approximation—the first condition is that

$$C_1(2\pi) = \int_0^{2\pi} \psi(a_0, b_0, \tau) \cos \tau \, d\tau = 0, \tag{31}$$

$$C_2(2\pi) = \int_0^{2\pi} \psi(a_0, b_0, \tau) \sin \tau \, d\tau = 0$$

and, moreover

$$\Delta = \begin{vmatrix} D_1(2\pi) & E_1(2\pi) \\ D_2(2\pi) & E_2(2\pi) \end{vmatrix} \neq 0 \tag{32}$$

If it is possible to determine α and β in this manner, one can assert that there exists a subharmonic solution for the nonlinear d.e. ($\mu \neq 0$).

The proof of *existence* of the subharmonic solution is thus relatively simple and represents merely an extension of Poincaré's method for the generating solution of the form (23).

The proof of stability is, however, more difficult because, in this case, one operates directly with a nonautonomous system. The variational equations in this case are obtained by replacing u and v by $u_0 + \gamma$ and $v_0 + \delta$ respectively (where γ and δ are small perturbations and are functions of t).

The variational system in this case is of the form

$$\frac{d\gamma}{dt} = (\mu\psi_u \sin t)\delta + (\mu\psi_v \sin t)\gamma$$

$$\frac{d\delta}{dt} = (\mu\psi_u \cos t)\delta + (\mu\psi_v \cos t)\gamma$$

These are the d.e. with periodic coefficients, the notations ψ_u and ψ_v being the same as before. We omit the intermediate steps leading to the calculation of the characteristic exponents and mention only the final conclusion, namely, that the stability condition requires that

$$D_1(2\pi) + E_2(2\pi) < 0, \qquad \begin{vmatrix} D_1(2\pi) & D_2(2\pi) \\ E_1(2\pi) & E_2(2\pi) \end{vmatrix} > 0 \tag{33}$$

The second condition (33) satisfies also the condition (32) of the existence of the subharmonic solution.

Mandelstam and Papalexi carried out these calculations for the resonance of the order 2, i.e., $n = 2$ in (22), in application to an electron-tube circuit; we shall indicate merely the results.

For $n = 2$, (23) has the form

$$(34)\quad x(t) = a \sin t - b \cos t - \frac{\lambda_0}{3} \sin 2t = X \sin (t - \phi) - \frac{\lambda_0}{3} \sin 2t$$

where $X^2 = a^2 + b^2$. Assuming the nonlinear characteristic of the electron tube of the form

$$(35)\qquad I = f(x) = I_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5$$

I being the plate current and x the grid voltage, it is necessary to replace x in (35) by its expression (34). In the case of the soft self-excitation $a_4 = a_5 = 0$. A detailed calculation of the nonlinear function $f(x, \dot{x})$ in

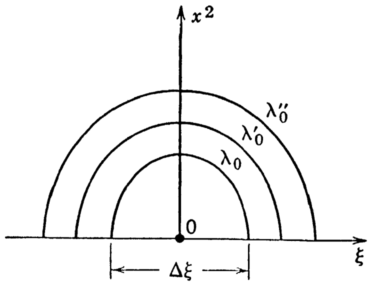


Fig. 13. Subharmonic resonance curves; "soft" self-excitation.

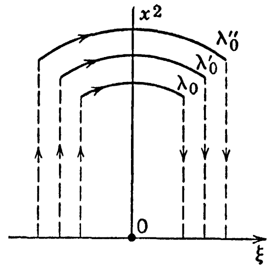


Fig. 14. Subharmonic resonance curves; "hard" self-excitation.

terms of the parameters of the electron tube results finally in the following conclusions. If one introduces the new variable $\xi = \omega^2/(\omega_0{}^2n^2 - 1)$, where ω is the actual frequency and $\omega_0 = 1$ is the frequency of the left-hand term in (22) for $\mu = 0$ and for the "soft" characteristic, the family of curves $X^2 = \phi(\xi)$, with λ_0 as the parameter, has the form shown in Fig. 13. The resonance curves exist in a certain interval $\Delta\xi$ which grows with the increasing λ_0 up to a certain value of λ_0, after which, on the contrary, it begins to shrink so that the resonance oscillation disappears for a certain critical value of λ_0. For the "hard" characteristic, it is necessary to take the full polynomial (35) which leads to long calculations resulting in the appearance of resonance curves, as shown in Fig. 14. The resonance oscillation appears suddenly with a finite amplitude for a definite value of ξ, goes through a maximum, and disappears again in a similar manner for another value of ξ. The behavior of the phenomenon as a function of λ is the same as in the "soft" case.

In spite of its considerable complexity arising mostly from the fact that one deals here directly with a nonautonomous system, the theory of Mandelstam and Papalexi found a complete experimental confirmation in a series of papers published in U.S.S.R. between 1934 and 1940. It is to be noted that this theory deals yet with a particular case when the nonlinear d.e. has a limit-cycle solution for $\lambda_0 = 0$.

A different procedure was followed by other investigators [59] for the case where the nonlinear d.e. has no limit-cycle solution for $\lambda_0 = 0$. The most important case of this kind is the so-called Duffing equation with forcing term, namely,

$$\ddot{x} + b\dot{x} + x + \beta x^3 = H(t) \tag{36}$$

The solution of this d.e. has been investigated by Duffing. A detailed analysis of this equation can be also found in Stoker's book [60]. For

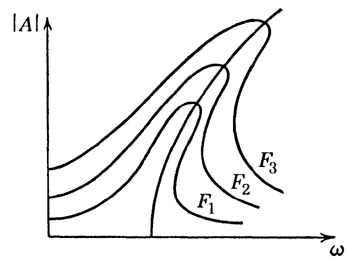


Fig. 15. Curves of nonlinear resonance; hard spring force.

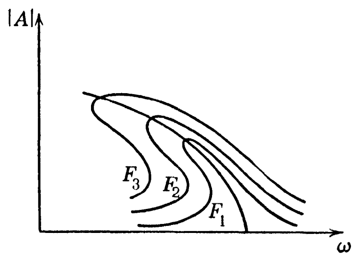


Fig. 16. Curves of nonlinear resonance; soft spring force.

these reasons we merely mention the conclusions of Stoker's analysis. If $b = 0$ and $H(t) = F \cos \omega t$, there exist oscillations of frequency ω with amplitudes depending on ω. For some values of ω several such oscillations may occur. For $\omega = 1$ (which corresponds to the resonance in the linear case), there exist oscillations of the same frequency with bounded amplitudes. Although the general solution is not known, various types of periodic solutions are possible (subharmonics, etc.) as well as nonperiodic solutions. The behavior of periodic solutions depends on the sign of β. If $b \neq 0$, substantially the same results are obtained but calculations are longer.

Interesting phenomena of "jumps" are observed in this case, as shown in Figs. 15 and 16 where the amplitude A is considered as a function of ω. If one considers, for instance, the case of Fig. 15 reproduced in Fig. 17 for a particular value of the parameter F, it is noted that by increasing ω beginning with some point 1, the branch 1,2 of the curve is followed. At the point 2, corresponding to the vertical tangency to the curve C, the amplitude suddenly drops to point 3 and continues following the branch 3,4 if ω continues increasing. If, however, ω decreases from 4, the amplitude remains on the branch 4,3 up to the point 5, where the tangent to C is vertical and another jump 5,6 takes place if ω continues to decrease. We shall limit ourselves to these few conclusions referring for the detailed information regarding this approach to Chapter IV of Stoker's book [60].

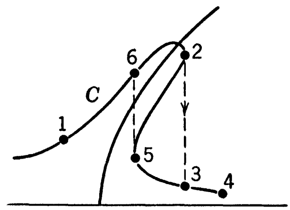


Fig. 17. Nonlinear resonance curve; oscillation hysteresis.

It is to be noted that the two theories just outlined relate to two different classes of the resonance phenomena and follow different methods. In fact, the Mandelstam-Papalexi theory relates to the subharmonic resonance effects in the self-excited systems acted on by an external periodic excitation (22), whereas the studies of this form of resonance for not self-excited systems (36) follow a different argument.

One can obtain a more uniform treatment of the subharmonic resonance problem by an extension of the stroboscopic method. The important point in this procedure is in that the difficult part of the problem concerning the determination of the characteristic exponent is eliminated and the problem of stability is reduced to that of the singular point in the stroboscopic plane (Art. 5, Chapter 2). In this way the remaining calculation involves only the standard perturbation procedure. We shall outline briefly this method omitting details.

Given a d.e. of the form

$$\ddot{x} + x_1 + \mu f(x, \dot{x}) = e \sin nt \tag{37}$$

which is the same as (22), one can always change the dependent variable $x = e\xi$ so that, in the new equation for ξ, the amplitude $e = 1$. Moreover, in the transformed equation (for ξ, but we shall continue using x as dependent variable) for $\mu = 0$, the solution is

$$x_0(t) = A_0 \sin t + B_0 \cos t + \frac{1}{1 - n^2} \sin nt \tag{38}$$

$$y_0(t) = A_0 \cos t - B_0 \sin t + \frac{n}{1 - n^2} \cos nt$$

We look for the solution of the form $x(t) = x_0(t) + \mu\, x_1(t) + \cdots$, in which we know already the "zero-order" solution $x_0(t)$, $y_0(t)$ given by (38). If the subharmonic solution exists, for instance in the first approximation, this means that the replacement of $x(t)$ by $x_0(t) + \mu\, x_1(t)$ in (37) must give a periodic solution. Since $x_0(t)$ is already periodic, this will impose a certain condition on $x_1(t)$, $y_1(t)$ in order to obtain the periodicity. The perturbation procedure yields

$$\ddot{x}_1 + x_1 = -\mu f(x_0, \dot{x}_0) \tag{39}$$

The integration of this d.e. gives

$$x_1(t) = D \sin t + E \cos t - \int_0^t \sin (t - \tau) f(x_0, \dot{x}_0)\, d\tau \tag{40}$$

$$y_1(t) = D \cos t - E \sin t - \int_0^t \cos (t - \tau) f(x_0, \dot{x}_0)\, d\tau$$

With the initial conditions $x_1(0) = y_1(0) = 0$, the constants D and E are zero so that the condition for the periodicity [compare with equation (4), Chapter 2] for the first-order solution is

$$x(2\pi) - x(0) = x_0(2\pi) + \mu\, x_1(2\pi) - x(0) = 0 \tag{41}$$

$$y(2\pi) - y(0) = y_0(2\pi) + \mu\, y_1(2\pi) - y(0) = 0$$

Hence

$$x_1(2\pi) = \int_0^{2\pi} \sin \tau\, f(x_0, \dot{x}_0)\, d\tau \tag{42}$$

$$y_1(2\pi) = -\int_0^{2\pi} \cos \tau\, f(x_0, \dot{x}_0)\, d\tau$$

In view of (41), the expressions (42) are

$$\int_0^{2\pi} \sin \tau\, f(x_0, \dot{x}_0)\, d\tau = 0, \qquad \int_0^{2\pi} \cos \tau\, f(x_0, \dot{x}_0)\, d\tau = 0 \tag{43}$$

Thus if the subharmonic solution exists, in the first-order approximation, *the Fourier development of* $f(x_0, \dot{x}_0)$ *does not contain terms with sin t and cos t*. The actual calculation begins from this point and leads to certain algebraic (in terms of A and B) expressions of the form

$$a_1(A, B) = 0, \qquad b_1(A, B) = 0$$

which expresses the vanishing of the coefficients of $\sin t$ and $\cos t$ in the Fourier development.

To bring the problem within the scope of the stroboscopic method, we set $x(2\pi) - x(0) = \Delta x(0)$ and $y(2\pi) - y(0) = \Delta y(0)$ so that equations (41) result in the difference equations

$$\Delta x(0) = \mu \int_0^{2\pi} \sin \tau\, f(x_0, \dot{x}_0)\, d\tau, \tag{44}$$

$$\Delta y(0) = -\mu \int_0^{2\pi} \cos \tau\, f(x_0, \dot{x}_0)\, d\tau$$

The approach to the subharmonic solution is characterized by $\Delta x(0) \to 0$ and $\Delta y(0) \to 0$. On the other hand, during this approach the coefficients a_1 and b_1 also approach zero. Setting $x(0) = x_0(0) = B$; $y(0) = y_0(0) = A + n/(1 - n^2)$, and considering a_1 and b_1 as functions of A and B, the difference equations [compare with equation (36), Chapter 2] appear in the form

$$\Delta x(0) = 2\pi\mu\, a_1(A, B), \qquad \Delta y(0) = 2\pi\mu\, b_1(A, B) \tag{45}$$

Changing the notations: $\Delta x(0) = \Delta\xi$; $\Delta y(0) = \Delta\eta$; $\Delta\tau = 2\pi\mu$ and passing to the limit $\Delta\tau \to d\tau$, one obtains the stroboscopic system

$$\frac{d\xi}{d\tau} = a_1\left(\xi, \eta - \frac{n}{1 - n^2}\right), \qquad \frac{d\eta}{d\tau} = b_1\left(\xi, \eta - \frac{n}{1 - n^2}\right) \tag{46}$$

where we set $B = \xi$; $A = \eta - n/(1 - n^2)$. Once this point is reached, the problem does not present any further difficulty, at least in principle. There may be yet *practical* difficulties of finding the real roots ξ_0 and η_0 of the polynomials $a_1(\xi_0, \eta_0) = b_1(\xi_0, \eta_0) = 0$, but the remaining part of the problem reduces to numerical calculations.

This somewhat lengthy survey of methods dealing with the subharmonic resonance shows at the same time the difficulties of this problem. In the Mandelstam-Papalexi method the problem is restricted to the self-excited systems but the essentially nonautonomous system is treated directly by the classical method without any attempt to reduce it first to the autonomous form. As the result of this, the calculation of stability reduces to that of the characteristic exponents, leading to equation (33).

The method of dealing with Duffing's equation is simpler but, by its nature, this equation is a rather special one, being outside the scope of equations of the type (22) most frequently encountered in electrical applications.

It may seem that the stroboscopic procedure leads more rapidly to the conclusions but, in reality, it contains also long calculations in connection with the determination of the functions $a_1(A, B)$ and $b_1(A, B)$. In fact, if $f(x_0, \dot{x}_0)$ is, for instance, a polynomial of the third degree, it is necessary to raise the expression (38) to the third power; for a "hard" characteristic the same expressions are to be raised to the fifth power, etc. Ultimately the *practical* difficulties of carrying out these calculations are still considerable, particularly when the common roots of two algebraic equations $a_1(\xi_0, \eta_0) = b_1(\xi_0, \eta_0) = 0$ are to be determined. However, the advantage of the stroboscopic method is in that it deals equally with the self-excited and with nonself-excited systems.

Summing up, the phenomena of the subharmonic resonance present a rather striking contrast between their simple intuitive meaning on the basis of subharmonics exciting corresponding resonances in a system and long and tedious calculations, even in a simplest possible case of the resonance 1/2 in a system with a "soft" characteristic. At present there exist no applications of these phenomena, presumably, because of the above-mentioned difficulties of actual numerical calculations.

5. Asynchronous actions

In the preceding article and also in Art. 2 of this chapter we outlined separately the phenomena of the subharmonic resonance and synchronization. In reality, these two physical manifestations form merely two different aspects of a more complicated picture when the autoperiodic frequency is "entrained" by the heteroperiodic one (the synchronization aspect) with the incidental relations affecting the behavior of the amplitude (the resonance aspect). Thus, because the global effect is too complicated, it is customary to split it into the phase and the amplitude relations and to speak about the former as "synchronization" and about the latter as "resonance."

In the case of an *exact* resonance, the matter is simpler since the synchronization effect vanishes and there remains only the resonance effect in its pure form.

Thus, for instance, if one considers the van der Pol equation wi... forcing term, namely,

$$\ddot{\xi} - \mu(\alpha - \beta\xi^2)\dot{\xi} + \xi = e \sin \omega t \tag{47}$$

where α and β are positive constants, μ is a small parameter, and $e \sin \omega t$ is a forcing term, one can consider the following situation. If, for a constant e, for instance, ω varies continuously, the oscillation will pass through a series of subharmonic resonances as ω will pass through a series of integer values; not all of these resonances exist however and, moreover, if they do, this happens only when ω is a relatively small integer, say, 2, 3, etc., [61].

Around each of these *exact* resonances there exists a *zone* of resonance in which the amplitude varies and, at the same time, the synchronization operates so as to "entrain" the autoperiodic frequency by the heteroperiodic one. In this way, as ω varies, subsequent resonance zones are traversed.

Between these zones there are regions of the so-called *asynchronous action* which do not play an important role compared to the resonance regions and which, for that reason, have not been investigated as completely as the latter, except in some special cases, one of which we shall mention.

If the amplitude e in (47) is finite, the change of the variable $\xi = ex$ transforms (47) into the d.e.:

$$\ddot{x} - \mu(\alpha - \gamma x^2)\dot{x} + x = \sin \omega t, \qquad \gamma = \beta e^2 \tag{48}$$

The generating solution ($\mu = 0$) in this case is:

$$\begin{aligned} x_0(t) &= A \sin t + B \cos t + M \sin \omega t, \\ y_0(t) &= \dot{x}_0(t) = A \cos t - B \sin t + M\omega \cos \omega t \end{aligned} \tag{49}$$

where A and B are constants of integration, and $M = 1/(1 - \omega^2)$.

In view of the assumed smallness of μ, we propose to investigate only the approximation of the first order: $x(t) = x_0(t) + \mu x_1(t)$; $y(t) = y_0(t) + \mu y_1(t)$.

The usual perturbation procedure yields:

$$\begin{aligned} x_1(t) &= \int_0^t \sin (t - \tau)\, f(x_0, \dot{x}_0)\, d\tau, \\ y_1(t) &= \int_0^t \cos (t - \tau)\, f(x_0, \dot{x}_0)\, d\tau \end{aligned} \tag{50}$$

where $f(x_0, \dot{x}_0) = (\alpha - \gamma x_0{}^2)\dot{x}_0$ and the initial conditions are:

$$x_0(0) = B, \qquad y_0(0) = A + M\omega, \qquad x_1(0) = y_1(0) = 0 \tag{51}$$

If a heteroperiodic oscillation with period $2\pi/\omega$ exists, the theory of Poincaré yields the following conditions of periodicity in the first approximation:

$$\begin{aligned} x(2\pi/\omega) - x(0) &= x_0(2\pi/\omega) - x_0(0) + \mu x_1(2\pi/\omega) \\ y(2\pi/\omega) - y(0) &= y_0(2\pi/\omega) - y_0(0) + \mu y_1(2\pi/\omega) \end{aligned} \tag{52}$$

On the other hand, from (49) and (51) one has

$$\begin{aligned} x_0(2\pi/\omega) - x_0(0) &= A \sin (2\pi/\omega) + B[\cos (2\pi/\omega) - 1] \\ y_0(2\pi/\omega) - y_0(0) &= A[\cos (2\pi/\omega) - 1] - B \sin (2\pi/\omega) \end{aligned} \tag{53}$$

Hence, if $\mu = 0$, the system (52) has only a trivial solution $A = B = 0$ since the determinant is different from zero.

If $\mu \neq 0$, but small (and this is the case in which we are interested), the condition of periodicity may be fulfilled if one determines A and B from the equations:

$$(54) \quad A \sin (2\pi/\omega) + B[\cos (2\pi/\omega) - 1]$$
$$+ \mu[\sin (2\pi/\omega) \int_0^{2\pi/\omega} \cos t f(\ ) \, dt - \cos (2\pi/\omega) \int_0^{2\pi/\omega} \sin t f(\ ) \, dt] = 0$$
$$A[\cos (2\pi/\omega) - 1] - B \sin (2\pi/\omega)$$
$$+ \mu[\cos (2\pi/\omega) \int_0^{2\pi/\omega} \cos t f(\ ) \, dt + \sin (2\pi/\omega) \int_0^{2\pi/\omega} \sin t f(\ ) \, dt] = 0$$

where $f(\ ) = f(x_0, \dot{x}_0)$ as before.

It is obvious that, since $A = B = 0$ for $\mu = 0$, they are small if μ is small. The actual calculation is long and tedious but, in the asymptotic case when ω is large, one can obtain the result from a simple discussion of the order of magnitude of the different terms. In fact, if ω is very large, $2\pi/\omega$ and $\sin (2\pi/\omega)$ are small and we can assume them $0(\mu)$ * in (54); hence $[\cos (2\pi/\omega) - 1]$ is small $0(\mu^2)$. Moreover, the integrals are also $0(\mu)$ since the integrands are finite and the upper limit is $0(\mu)$. One ascertains in this manner that both A and B are to be $0(\mu)$ in order to make the system (54) possible.

This shows that if one impresses a sufficiently high frequency on the existing autoperiodic (self-excited) oscillation, the latter is extinguished or "quenched"; this phenomenon is called *asynchronous quenching*, the term "asynchronous" in this case meaning that the extraneous frequency ω does not stand in any rational ratio with respect to that of the autoperiodic oscillation which is quenched in this manner.

This phenomenon is well known in electronics where the frequency ω is applied to the grid of the electron tube working as oscillator through the inductive coupling.

No corresponding phenomenon has been discovered as yet in mechanics. In fact, to be able to discover it in a mechanical system, the latter should have a *predetermined* nonlinear characteristic, but such a predetermination of characteristics in advance is yet impossible with our present knowledge of this question.

There is another phenomenon of the asynchronous type, the so-called *asynchronous excitation* which consists in releasing a potentially possible but initially nonexisting autoperiodic oscillation by applying an oscilla-

* The symbol $0(\mu)$ means, as usually, "of the order of μ."

tion of an asynchronous frequency ω. This phenomenon is, however, of an entirely different nature as compared to the phenomenon of the synchronous quenching inasmuch as it is possible only in systems of the "hard" type, that is, such whose nonlinear characteristic is expressible by a polynomial of, at least, fifth degree; moreover it occurs only when the relative values of the parameters of the nonlinear characteristic are in certain intervals.

In view of the complicated nature of this phenomenon we shall not elaborate this question here and refer to publications on this subject [62].

6. Oscillations caused by retarded actions [63]

In recent years there appeared a considerable interest in a special type of oscillations the theory of which is amenable to the so-called *difference-differential equations* (d.d.e. for short). Most of these oscillations occur in control systems and are generally very objectionable. The theory of these oscillations is far less advanced than the classical procedure of Poincaré on the basis of which we have studied the previous phenomena. In fact, to be able to account for these oscillatory phenomena, one should treat them in terms of the nonlinear d.d.e., but the advance in this field is yet insufficient to be able to develop a theory on that basis.

Only in a very special case of the smallness of time lag and under some additional plausible assumptions concerning the order of magnitude of certain quantities encountered in applications, it is possible to reduce the problem to a point where one can use the perturbation theory to the extent of obtaining the first approximation. The results so obtained are generally in agreement with the experimental facts but the whole situation is yet in a rather qualitative stage and, as was just stated, is still limited to a rather special case.

What constitutes an entirely different feature here, is *the linear problem.* In all that precedes the linear problem was reducible to the simplest possible equation $\ddot{x} + x = 0$, of the harmonic oscillator. Here one encounters an essentially transcendental problem the significance of which is necessary to establish in the first place.

For our purposes here it is sufficient to define a d.d.e. as a d.e. in which the various terms *do not relate to the same values of the independent variable t* (time). Thus, for instance, the following equations the physical significance of which is obvious, namely,

$$(55) \qquad a\ddot{x}(t) + b\dot{x}(t) + cx(t-h) = 0$$

$$a\ddot{x}(t) + b\dot{x}(t-h) + cx(t) = 0$$

are examples of the d.d.e. We use here the notations $\ddot{x}(t)$, $\dot{x}(t)$, and $x(t)$ to emphasize that these dependent variables relate to the value t of the independent variables in order to contrast these terms with the *retarded terms* $cx(t - h)$, $b\dot{x}(t - h)$ which are related to the *retarded* value $(t - h)$ of the independent variable, h being *the time lag.*

The essential feature of the d.d.e. is that they always result in a transcendental problem, which may be specified as follows. In an ordinary d.e. (e.g., linear with constant coefficients) the degree of its characteristic equation (and, hence, the number of its roots) is always equal to the order of the d.e.

In a d.d.e. the degree of its characteristic equation is always infinite whatever the order of the d.d.e. may be, as is obvious if one replaces the retarded terms such as $cx(t - h)$ and $b\dot{x}(t - h)$ in (55) by their Taylor's expansions which thus account for the infinite order of (55) *considered as a d.e.*

As the number of roots of the characteristic equation is also infinite, one readily sees that there is always a possibility that some of these roots may have positive real parts, thus accounting for the self-excitation of parasitic oscillations. It is precisely in this connection that the interest for these d.d.e. appeared in applications.

Instead of operating with the d.e. of an infinitely high order, it is preferable to modify the characteristic equation. In fact, since the d.e. is linear, one may try to satisfy it with a solution $x = x_0e^{zt}$, $z = \alpha + i\omega$. On the other hand, one has

$$x(t - h) = x(t) - \frac{h}{1!}\dot{x}(t) + \frac{h^2}{2!}\ddot{x}(t) - \cdots$$
$$= x(t)\left[1 - \frac{h}{1!}\frac{\dot{x}(t)}{x(t)} + \frac{h^2}{2!}\frac{\ddot{x}(t)}{x(t)} - \cdots\right]$$

With this solution $\dot{x}/x = z$; $\ddot{x}/x = z^2$, $\cdots$ so that one has

$$x(t - h) = x(t)\left[1 - \frac{hz}{1!} + \frac{h^2z^2}{2!} - \cdots\right] = xe^{-hz} \tag{56}$$

Consider, for instance, the first equation (55), making use of (56); it is clear that its characteristic equation becomes

$$az^2 + bz + ce^{-hz} = f(z) = 0 \tag{57}$$

In a similar way, for the second equation (55) the characteristic equation is

$$az^2 + bze^{-hz} + c = f(z) = 0 \tag{58}$$

The problem of determination of the roots reduces thus to the determination of zeros of the corresponding analytic functions $f(z)$. This question has been recently studied by Pontriagin [64].

In applications the problem presents itself in a somewhat special way which renders it somewhat simpler on one hand but, on the other hand, introduces certain new features. In fact, the simplification consists in that in applications there is no necessity of investigating the whole (infinite) spectrum of frequencies, but only those frequencies which are not far from the frequency with which the system is purported to operate, as only a few parasitic frequencies are generally objectionable. In a similar way, the quantities α (increments if $\alpha > 0$; decrements if $\alpha < 0$) cannot be large either, since α is a measure of energy (either absorbed if $\alpha > 0$ or dissipated if $\alpha < 0$) stored in the oscillation. This reduces the problem to the investigation of zeros only in a limited area of the complex plane, in which only a few such zeros are generally present.

As to the complication just mentioned, it consists in that the coefficients associated with e^{-hz} (e.g., c in 57 or b in 58) are generally *variable parameters*. One recalls that these *retarded terms* are generally certain quantities introduced by the control schemes, and the intensity of these control actions is usually subject to adjustments within certain limits.

The practical procedure consists to replace z in (57) (or 58) by its value $\alpha + i\omega$ and to separate the real and the imaginary parts, giving rise to two algebraic equations leading (after some elementary transformations) to certain transcendental equations which can be solved graphically when the parameter λ varies.

If one writes, for instance, the first equation (55) as

$$\ddot{x} + a\dot{x} + \lambda x_h = 0$$

where $x_h = x(t - h)$, its characteristic equation is

$$f(z) = z^2 + az + \lambda e^{-hz} = 0$$

and the above procedure of separating the real and the imaginary quantities leads finally to equations

$$\operatorname{cotan} \beta = (\omega^2 - a\alpha - \alpha^2)/(a + 2\alpha)\omega,$$
$$(\omega^2 - a\alpha - \alpha^2)^2 + (a + 2\alpha)^2\omega^2 = \lambda^2 e^{-2h\alpha}$$

where $\beta = \omega h$. If λ varies, the zeros of $f(z)$ vary and, for some special, *harmonic* values $\lambda = \lambda_1$, the path of some zero in the complex plane may cross the $\alpha = 0$ axis at which point there appears a purely sinusoidal solution. If one sets $\alpha = 0$ in the preceding equations and attaches the subscript 1 to such *harmonic values*, one has two equations to be fulfilled simultaneously

(59) $$\cot \beta_1 = \frac{\omega_1}{a} = \left(\frac{1}{ah}\right)\beta_1 \qquad \omega_1^4 + a^2\omega_1^2 - \lambda^2 = 0$$

with conditions

(60) $$\cos \beta_1 = \omega_1^2/\lambda > 0, \qquad \sin \beta_1 = a\omega_1/\lambda > 0$$

Equations (59) permit determining the "harmonic roots" graphically. We refer to an earlier publication [65] in this connection and merely mention here that "harmonic roots" correspond to a spectrum of discrete frequencies $\omega_1, \omega_1', \omega_1'', \cdots$ to which corresponds also a discrete sequence of the parameter values $\lambda_1, \lambda_1', \lambda_1'', \cdots$.

If one takes one of these values, for instance ω_1, to which corresponds the parameter value λ_1, one finds that for $\lambda = \lambda_1 + \Delta\lambda$, $\Delta\lambda$ being small, the harmonic condition ($\alpha = 0$) disappears and one has $\Delta\alpha$, the sign of which is related to that of $\Delta\lambda$. ($\Delta\alpha > 0$ for $\Delta\lambda > 0$ and $\Delta\alpha < 0$ for $\Delta\lambda < 0$.)

Although the determination of ω_1 by this procedure is in accordance with observed facts, the above-mentioned result regarding α (the decrement) is obviously absurd, inasmuch as it shows that, for a departure from a harmonic value $\lambda = \lambda_1$, the oscillation either disappears ($\alpha < 0$) or grows beyond any bound (if $\alpha > 0$). Thus, although the linear theory gives correct results as regards the determination of the transcendental spectrum of frequencies, it fails to explain the existence of the stationary state of these oscillations.

As the general theory of nonlinear d.d.e. did not progress sufficiently for treating this problem completely, it is necessary to limit the problem to a *local* analysis of one such oscillation in the neighborhood of some harmonic oscillation with frequency ω_1. Under this restriction it is still possible to use the perturbation method. Thus, for instance, if, instead of the first d.d.e. (55), we consider the following

(61) $$\dot{x} + a\dot{x} + x_h + \epsilon x_h^3 = 0$$

where $x_h = x(t - h)$, it is possible to show that this equation admits a stable periodic solution of a limit-cycle type in a certain region where $\alpha > 0$. We refer to a paper on this subject [65], as it would be too long to establish this result here. It is sufficient to mention that, in addition to small parameters a and ϵ as usually, one has to assume also that h and α in these notations are also small. With these assumptions, the perturbation procedure leads to the above-mentioned results for the d.e. of the first approximation, and these yield the results in agreement with the experimental facts.

It is interesting to note that, if one sets $h = 0$ in (61), it becomes an ordinary nonlinear d.e. not having any stationary periodic solution (except $x = 0$), but it becomes a self-excited system (of a special transcendental type) if $h \neq 0$. In this manner the presence of a time lag in the nonlinear "restoring force" term $x_h + \epsilon x_h^3$ is sufficient to convert a nonself-excited (dissipative) physical system into a self-excited one of a "limit-cycle type." Besides this, such a system maintains still its transcendental character inasmuch as several oscillations (with incommensurate frequencies) may exist in a stationary state under certain conditions.

As an example of a case in which the manifestation of these curious effects was particularly striking [66], we can mention the problem of antirolling stabilization of a ship by the so-called activated-tanks method. In this method the displacement of the water ballast between the port and the starboard tanks A and B is effected by means of an axial variable pitch pump P placed in the channel connecting the tanks. The actuation of the pitch of the blades of P is effected through a servo system controlled by instruments responsive to the ship's angular motion. As long as P works in a normal condition (i.e., without cavitation, etc.), the system works also normally in accordance with the d.e. of stabilization. If, however, one forces P to transfer more ballast, a moment is reached when the cavitation sets in and, owing to the time lag so introduced, the operation of the system begins to be governed by a d.d.e. (in fact, the second equation (55)) with the incidental appearance of one transcendental frequency in the neighborhood of the frequency of rolling on which the system is purported to react. Needless to say that in this particular example the appearance of "fluttering" of the pump's blades is highly objectionable in view of the disturbance on the flow of ballast which is thus produced.

At present the whole subject of oscillations (of a generally parasitic type) caused by retarded (or advanced, if $h < 0$) actions acquires a great deal of interest in connection with problems of automatic control as well as with some biological and econometric problems. The reader is referred to a recent monograph by Bellman [63] outlining these problems in a very interesting manner and appending an extensive bibliography on the subject.

CHAPTER 4

RELAXATION OSCILLATIONS

1. Introductory remarks

The methods of approximations, outlined in Chapter 2, cease to hold when the parameter μ is not small. In fact, in all these methods use is made of the series solutions arranged according to the ascending powers of μ, and it is obvious that, if μ is not small, the series ceases to converge. This is the reason why all attempts to obtain a solution of the van der Pol equation (when μ is large) in a form of a series were not successful although in the asymptotic case ($\mu \to \infty$) some qualitative conclusions could be still obtained.

As was already mentioned, the principal reason for this difficulty lies in that, when μ is large, the solution of the van der Pol equation (obtained graphically) exhibits two points of an extremely bad analyticity across which the analytical continuation is virtually impossible. From a physical point of view this circumstance is analogous to the effect of a shock for which the continuity is preserved but the analyticity is lost. These phenomena characterizing a strong nonlinearity (if one thinks in terms of the large μ in the van der Pol equation), or *relaxation oscillations*, have been known for a long time but their study began only after the van der Pol equation had been explored in the domain of the small parameter, the large-parameter domain having been attempted for study as a natural extension of this problem.

At present the whole situation seems to undergo a certain "parting of the ways" between the physicists and engineers on one hand, and the mathematicians on the other. The former, pursuing the applied problems, seem to lean more and more to a purely discontinuous treatment of the relaxation oscillations following the pattern of the theory of shocks in classical mechanics, whereas the latter still persist in the search of an exact solution of the van der Pol equation as evidenced by the work of Cartwright and Littlewood and their school [73].

From the point of view of the discontinuous treatment outlined below, the van der Pol equation (with large μ) is not involved at all and, instead, the d.e. are of the form [67]:

$$\frac{dx}{dt} = \frac{P(x, y)}{T(x, y)}; \qquad \frac{dy}{dt} = \frac{Q(x, y)}{T(x, y)} \tag{1}$$

In this form P, Q, and T may be regarded as analytic functions of x and y; the "relaxation range" begins at the points (x_0, y_0) for which $T(x_0, y_0) = 0$.

However, in order to be able to reduce the d.e. to the form (1), certain idealizations of physical problems of this nature are necessary. These idealizations resemble closely similar ones used in the classical theory of mechanical shocks. Once this point is understood, the formation of such d.e. does not present any difficulty, as we shall see in an example below.

The delicate point of such a theory is in the nature of this idealization and, for that reason, we insist somewhat on this point in the following four articles.

The discontinuous theory is the work of the Russian mathematical physicists (Mandelstam, Chaikin, Lochakov, and a few others) [68]; practically the same conclusions were reached independently by Vogel in France [69].

The asymptotic theory has been recently investigated by Wasow [70], Flanders and Stoker (U.S.A.) [71], and Haag (France) [72]. Still broader mathematical aspects are contained in important contributions of Cartwright and Littlewood [73].

As far as applied results are concerned, in all cases they are mostly qualitative, in the sense that none of these theories yield any quantitative solution of the d.e. (except, possibly, the determination of the period) in the usual sense of the word. As to the qualitative results, they are also somewhat differently presented in these different approaches.

The discontinuous theory deals with a very broad class of the d.e. (1) in which the van der Pol equation generally does not figure at all. The relaxation range is characterized by the vanishing of $T(x, y)$, as was just mentioned, and not by any "large value of μ." The theory is in all respects analogous to the classical theory of shocks. The regions of rapid transitions on the integral curves are idealized by discontinuities. Likewise, the intentional ignoration of the d.e. "during" the discontinuities is compensated for by an "additional information" not contained directly in the d.e. but appears in the form of the so-called *condition of Mandelstam* regarding the invariance of energy during a discontinuity (compare this with "additional information" appearing in the theory of shocks in the form of the theorems of momentum and kinetic energy).

To be able to build a theory on this basis, it was necessary to prepare a theoretical ground in the form of *degenerate* d.e. admitting quasi-discontinuous solutions.

In a final shape the discontinuous theory turned out to be eminently successful as a practical tool of exploration of all known relaxation phenomena. The usual reproach, that this theory "mixes up" so to speak the analytical approach with a physical postulate (the condition of Mandelstam) is no more justified than it is in the classical theory of shocks. In fact, in both cases the situation is practically the same and the underlying philosophy of the question can be best stated in the following words of Boussinesq [74]:

"Si la continuité simplifie les choses quand elle en relie plusieurs qui suivent la même loi, elle les complique, au contraire, le plus souvent, lorsqu'elle établit la transition entre deux catégories d'objets ou de faits régis par deux lois simples différentes; et c'est alors une discontinuité fictive, un passage brusque de la première catégorie à la seconde, qui rend les questions abordables." *

As regards the asymptotic theory, its purpose is to avoid the short cuts offered by the degeneration theory; it prefers to deal with the d.e. as it stands in spite of the complications at some points of a bad analyticity. In its asymptotic form $\mu \to \infty$ (Liénard [75], Le Corbeiller [76], Flanders and Stoker [71], and others), the results are practically the same as in the discontinuous theory but the reach of the method is shorter than in the latter, for the reason that the discontinuous theory deals with a system of one degree of freedom on the basis of a d.e. of the first order whereas the asymptotic theory treats the same problem on a basis of a d.e. of the second order since it does not make use of degenerate form of the d.e. One of the advantages of the discontinuous theory is that it permits using the phase-plane representation even in the case of a degenerate system with two degrees of freedom and is thus able to cover an additional number of cases of relaxation oscillations such as symmetrical multivibrators, etc.

* "If the continuity simplifies the matter when it connects several phenomena following similar laws, it complicates, on the contrary, the relations when it is used for the purpose of connecting phenomena following different laws. It is precisely here that an idealized discontinuous passage from one law to the other which renders the study possible."

In addition to these two trends, there appeared also a third one in the form of the theory of Cartwright and Littlewood [73]; it may be regarded as an intermediate theory inasmuch as it deals with the van der Pol equation in which μ is large but not infinitely large as assumed in the asymptotic theory. As it is difficult to give an account of this theory here, it is sufficient to mention that it considers the behavior of the solution (the term "solution" in this theory may be regarded as "representative point" R in this review) in the various domains of the integral curve supposed to be known (e.g., from the van der Pol graphical curve). It is shown that this behavior can be followed up from one domain to the other if, in addition to the d.e. itself, one takes into account also the *integrated* and the *energy* equations. This analysis results in a number of lemmas which ultimately lead to the establishment of the condition of periodicity and to the calculation of the period of the oscillation. As was previously mentioned, the actual solution (e.g., in a form of a power series) is not determined by this method; this theory gives, however, a more precise calculation of the period as compared to the earlier estimates. Although this theory was applied later by Cartwright to the analysis of a number of applied problems [77], in the theory of relaxation oscillations proper, its application did not go beyond the van der Pol equation and, for that reason, it is impossible to compare its ultimate reach with that of the discontinuous theory, which, on the contrary, deals with all known cases of relaxation oscillations on a uniform basis.

In view of this somewhat unsettled situation concerning the question of *strong nonlinearities* (we intentionally avoid the use of the term "large-parameter" range for the reason mentioned in connection with (1)), we outline in this chapter mostly the fundamentals of the discontinuous theory inasmuch as it has apparently reached a certain stage of completeness.

In Art. 7 of this chapter are indicated also some fundamentals of the asymptotic theory. As to the basic investigations of Cartwright and Littlewood, we merely refer to the papers of these authors as their presentation would be too difficult here in view of the elementary character of this review dealing primarily with the physical aspects of the theory of oscillations.

2. Piecewise analytic phenomena

Consider the following theoretical example due to Andronov [12]. A ball C is projected with a velocity v on a horizontal plane limited by two vertical walls A and B, Fig. 18(a). It is assumed that the ball and the walls are perfectly elastic and there is no friction. It is clear that C will be bounced back and forth owing to the collisions with the walls and the phase-plane diagram, Fig. 18(b), in this case is a rectangle $abcd$. On the sides ab and cd the motion occurs with a constant velocity $\pm v$ and is governed by a simple d.e. $\ddot{x} = 0$; the stretches bc and da are traversed in no time if we assume the classical theory of shocks, namely: the velocity v changes discontinuously from $+v$ to $-v$ or vice versa, when $x = \pm\alpha$. One has thus a kind of a periodic motion (since the trajectory $abcd$ is closed) but *it is not an analytic* motion because of discontinuities at $x = \pm\alpha$. One can, however, call it *piecewise analytic* motion, emphasizing the fact that on the stretches ab and cd it is still governed by the d.e. $\ddot{x} = 0$.

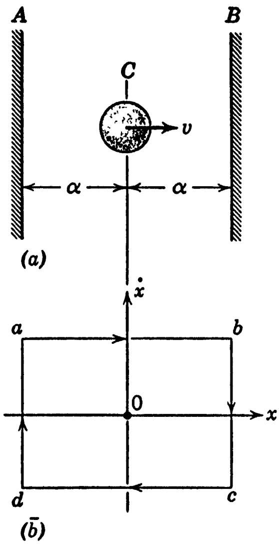


Fig. 18. Mechanical piecewise analytic cycle.

As a second example one may consider the theory of a clock. In a clock there are two principal parts: (a) a torsional pendulum with a small damping and (b) an escapement mechanism which delivers periodically timed impulses applied to the pendulum. The trajectory of (a) is obviously a logarithmic spiral approaching a stable focus. Figure 19 shows an arc AMB of this spiral, the points A and B being separated in time by one period T. When the point B is reached the escapement communicates an impulse to the pendulum which increased in a quasi-discontinuous manner its velocity from B to A. If the curve $AMBA$ is closed, the motion is periodic. It is shown easily that if one starts the clock either from a larger or from a smaller amplitude, it will finally approach the stationary amplitude for which the curve $AMBA$ becomes closed. The motion exhibits in all respects the properties of a *limit cycle* but, in this case, this cycle is not analytic on account of the stretch BA. One can call it again a *piecewise analytic* limit cycle, inasmuch as the analytic arc AMB is traversed by R with a finite velocity in accordance with the d.e. of the part (a) of the mechanism, whereas the segment BA is traversed with infinite (practically) velocity and the d.e. has nothing to do with this stretch BA, at least under the assumed idealization.

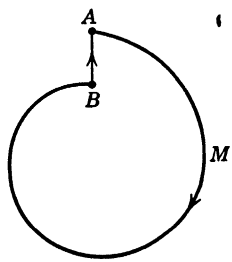


Fig. 19. Piecewise analytic trajectory of a clock.

Already from these two examples the piecewise analyticity is apparent, namely, on the analytic arcs the motion takes place according to the d.e. governing the process, but the discontinuous stretches "closing" a piecewise analytic cycle of this kind are traversed with infinite velocity and are due to some other cause which does not stand in a *direct* relation to the d.e.

3. Degeneration theory

If the pattern of the theory of shocks is to be used, it is clear that it is necessary to find its essential element—discontinuities—in the solutions of the d.e. describing an oscillatory process of this nature. This feature can be ascertained in the so-called *degenerate* d.e.

Consider, for instance, a simple d.e. of the form

$$a\ddot{x} + b\dot{x} + kx = 0 \tag{2}$$

with constant coefficients and assume that the coefficient a is very small; we call in this case the d.e. *degenerescent* (from the second to the first order). If, however $a \equiv 0$, one has the d.e. of the first order which may be regarded as a completely *degenerate* one. The question arises as to how the solutions of the two d.e.—the degenerescent and the degenerate—compare to each other. If one calls the solution in the first case $x(t)$ and in the second $\bar{x}(t)$, a simple calculation (12), taking into account the order of magnitude of a, shows that $\alpha(t) = x(t) - \bar{x}(t)$ approaches zero uniformly when $a \to 0$ for all t. As to the derivative $\dot{\alpha}(t) = \dot{x}(t) - \dot{\bar{x}}(t)$, this convergence to zero is uniform everywhere except for very small values of t where the uniformity of convergence is lost.

This result is of interest for the following reason: very often in applications one encounters the cases where the coefficient of the second derivative is very small. This happens for instance in an electric circuit with a negligible inductance, in which case one assumes approximately $L \cong 0$, so that instead of the d.e. $L(di/dt) + Ri + 1/C\int idt = 0$, one considers the degenerate equation $Ri + 1/C\int idt = 0$. If, instead of the variable i, the current, one introduces $x = \int (di/dt)\, dt$, the charge on capacitor, these two equations are respectively, $L\ddot{x} + R\dot{x} + (1/C)x = 0$ and $R\dot{x} + (1/C)x = 0$.

We assume that the circuit is initially "dead" (i.e., $x = 0$, $\dot{x} = 0$ for $t \leq 0$) but, for $t = 0$, a constant emf E is suddenly applied to the circuit. In the degenerescent case we have the d.e. of the second order

$$LC\ddot{x} + RC\dot{x} + x = EC \tag{3}$$

with two initial conditions: for $t = 0, x = 0, \dot{x} = 0$, which gives the well-known case of the variable x approaching asymptotically the constant value $x_0 = EC$, Fig. 20. The essential point is that curve $x(t)$ starts with the zero slope since $\dot{x} = 0$ for $t = 0$.

Consider now the degenerate equation of the first order $RC\dot{x} + x = EC$ in which case there is only one constant of integration corresponding to the initial condition: $t = 0$, $x = 0$. The integration gives $x = EC[1 - \exp(-t/RC)]$. Differentiating this expression and setting $t = 0$, one gets $\dot{x}(0) = E/R$. On the other hand, since it is assumed that the circuit is "dead" initially, one has two initial conditions $x(0) = \dot{x}(0) = 0$. One has thus a contradiction for the second initial condition $\dot{x}(0)$ which can be removed only by the assumption that, for $t = 0$, $\dot{x}$ jumps discontinuously from 0 to E/R. In this way, if one operates with a degenerate equation, the variable $\dot{x}(t)$ has to admit a discontinuity in order to reconcile the result with the physical initial conditions: for $t = 0$, $x(0) = \dot{x}(0) = 0$.

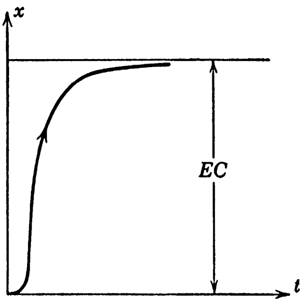


Fig. 20. Transient state of a physical system describable by a d.e. of the second order.

4. Conditions imposed by invariants

The theory of degeneration merely states that certain variables of a degenerate d.e. admit discontinuities arising from the inconsistency in their initial conditions. It is necessary to specify this further in connection with physical problems. It was said already that the discontinuous theory of relaxation oscillations follows the pattern of the theory of shocks. In the latter one ignores intentionally what kind of a d.e. governs the shock *during* a very short period when the colliding bodies are in contact with each other. Instead of this, such a theory merely connects the initial and final values of the variables (before and after the shock) on the basis of the invariance of the momentum and of the kinetic energy during this short time.

In the discontinuous theory of relaxation oscillations a similar approach was suggested by Mandelstam who pointed out that during the short time of a rapid transition of a variable from one value to the other (which is idealized by a mathematical discontinuity), the *energy stored in a system remains invariant*, which may be regarded as a plausible physical postulate. In fact, since in the assumed idealization the "duration" of a discontinuous process is zero, the only way to vary the energy by a finite quantity is to admit that the power (the rate of change of energy) is infinite which is obviously ruled out on physical grounds.

Consider for instance "the L degeneration," in the previously considered case where L is so small that the corresponding term is neglected. The only form of the stored energy in this case is electrostatic energy $E = \frac{1}{2}CV^2 = \frac{1}{2}xV$, where $x = CV$ is a charge in capacitor. Since E is conserved during the discontinuity, V cannot vary discontinuously either. Nothing prevents, however, dV/dt and, therefore, $i_c = C(dV/dt)$ from changing discontinuously as we saw already from the example of the preceding section. Likewise, in the case of the "C degeneration" (i.e., when $C = \infty$), the stored energy is purely electromagnetic $E = \frac{1}{2}Li^2$ and, since E remains invariant, i cannot change discontinuously, but nothing prevents di/dt and, therefore, $L(di/dt)$ from so doing.

It is thus seen that the condition imposed by the invariant in a given degenerate problem permits ascertaining "in which direction" (in the phase plane) a discontinuity can occur.

5. Discontinuous theory of relaxation oscillations

Under the conditions specified in the last two articles, the discontinuous theory can be outlined in a simple manner.

Suppose that the d.e. are of the form

$$\frac{dx}{dt} = P(x, y)/T(x, y), \qquad \frac{dy}{dt} = Q(x, y)/T(x, y) \tag{4}$$

where P, Q, and T are analytic functions. If $T(x, y) \neq 0$, this system is a normal one and the general theory (Art. 1, Chapter 1) is applicable. Thus, for instance, if a point A of a phase plane is given (which means certain initial conditions), a trajectory, represented by an analytic arc AB (Fig. 21) will begin at this point and will continue up to the point B of coordinates x_B, y_B for which $T = 0$. We call such a point a *critical point*. At this point B the d.e. lose their meaning and the analytic continuation of solution is impossible. If, however, one takes into account the condition of Mandelstam, a *physical continuation* is still possible. In fact, the point B in this theory is the "beginning" of the discontinuity BC traversed in no time, provided C is on another analytic arc CD representing the solution of the d.e.

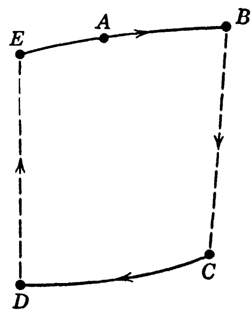


Fig. 21. Piecewise analytic cycle.

Assume that the arc CD ends at the point D for which $T = 0$ again. This determines another discontinuity DE which ends at the point E which is on the arc EB, and so on. The cycle consists thus of the two analytic arcs EB and CD on which the motion of R occurs with a finite velocity, joined by discontinuous stretches BC and DE traversed instantaneously.

We give first this result for the purpose of specifying the general aspect of a piecewise analytic phenomenon before entering into its detailed analysis.

It is useful to note the following points:

1. The form of the d.e. (4) appears practically in all relaxation problems and it is generally impossible to reduce it to the van der Pol equation with a large parameter value. In fact, the parameter does not figure at all in these equations and the "critical points" B and D as defined above appear when the denominator T vanishes.
2. The oscillatory phenomenon is governed by the d.e. as long as R moves continuously on analytic arcs but, on arriving at a critical point, the phenomenon ceases to be governed by the d.e. during its rapid (instantaneous) transition until another analytic arc is encountered on which the motion takes place again in accordance with the d.e. The instantaneous transitions occur in accordance with the condition of Mandelstam.

Although all this may appear as a somewhat *a priori* statement from the standpoint of the analytic theory with which we were dealing so far, in reality the matter is just of the same nature as in the classical theory of shocks where precisely the same argument is used. The fact that the shocks are capable of maintaining periodicity in cases where the d.e. itself has no periodic solution is quite clear from Andronov's example (Fig. 18).

The role of the invariants is also clear in both cases. In the theory of shocks these invariants appear in the form of momentum and kinetic energy "conserved" during a very short duration of the shock idealized by a discontinuity. In the discontinuous theory of relaxation oscillations, a similar invariant appears in the form of the condition of Mandelstam which states that the total energy stored in an oscillatory process is likewise "conserved" during a similar rapid transition.

Summing up, the discontinuous theory of relaxation oscillations is nothing but a generalized theory of shocks intended to represent periodic phenomena with quasi-discontinuous portions in their cycle.

The following article describes an example of application of this theory.

6. Application of the discontinuous theory to a multivibrator

We consider a multivibrator according to the scheme of Fig. 22 with notations and positive directions on circuits shown [12]. The electron tube V_2 is the nonlinear element of the circuit whereas V_1 acts merely as a linear amplifier, amplifying the voltage ri across r and applying the amplified voltage $e_g = kri$ to the grid of V_2, k being the amplification

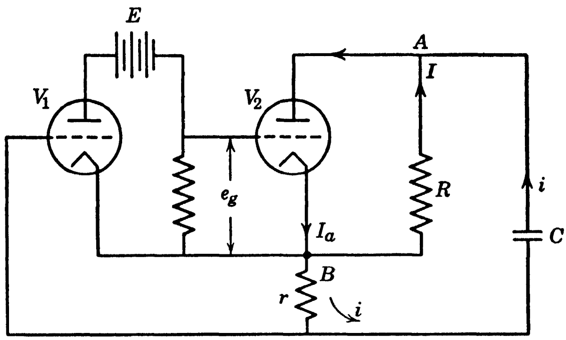


Fig. 22. Asymmetrical multivibrator circuit.

factor. If the nonlinear characteristic of V_2 is $I_a = \phi(e_g) = \phi(kri)$, the equations of the circuit are

$$(5) \qquad (R + r)i + V = R\phi(kri), \qquad i = C\frac{dV}{dt}$$

These equations can be reduced to one equation of the first order

$$(6) \qquad \frac{di}{dt} = \frac{i}{C[krR\phi'(kri) - (R + r)]} = \frac{i}{CT(i)}$$

which is of the form (4), with $\phi'(kri) = [d\phi\,(kri)]/(di)$. The critical point i_1 of (6) is given by the root i, of $T(i_1) = 0$. One can proceed either analytically if one approximates $\phi(kri)$ by a polynomial, or graphically. Assume the latter procedure, shown in Fig. 23, where the curves C and C' are $C{:}R\phi(kri)$ and $C'{:}(R + r)i$. The difference of ordinates of C and C' is clearly $V(i)$, and this curve is shown as $C''{:}V(i)$.

If one traces on the lower part, Fig. 23(b), the slope curve C''' of C'' (after multiplying its ordinates by kr) and subtracts the constant quantity $R + r$, one obtains two points $\pm i_1$ which are the roots of the square bracket in (6); these points are thus the *critical points*, and are also plotted on Fig. 23(a). If the representative point ρ (we use this notation to avoid confusion with R used here for a different purpose) on C''

finds itself in the interval ($\pm i_1$), the square bracket in (6) is positive, which shows that the motion is unstable (arrows on C''). On arrival to B, point ρ leaves the trajectory C'', as B is a critical point, and the condition of Mandelstam governs the discontinuous transition. As the

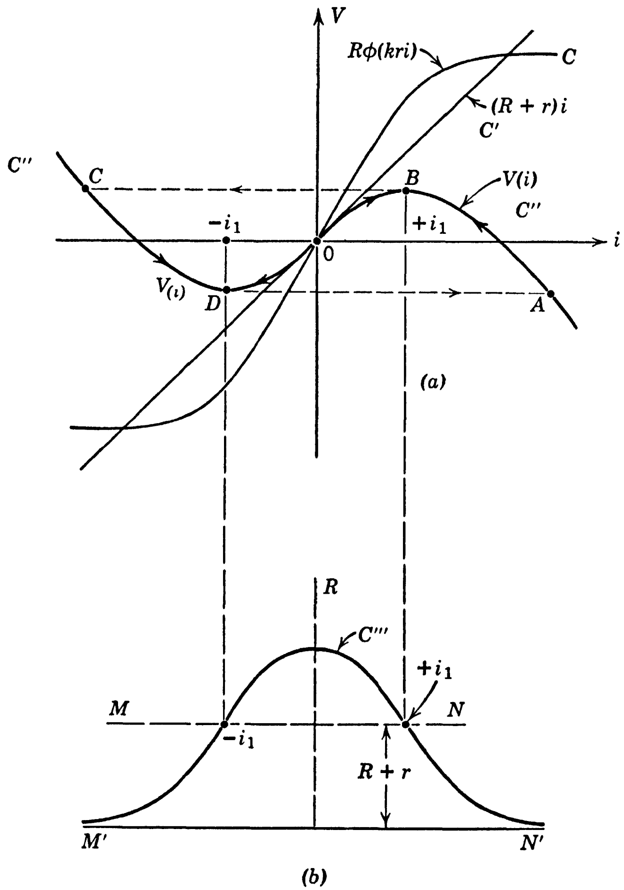


Fig. 23. Graphical determination of critical points.

energy here is purely electrostatic (since we operate with degenerate equation $L \equiv 0$), the invariance of V should be preserved during the jump which thus occurs along the parallel to the i-axis (dotted line BC). At the point C, ρ is again on C'' and the d.e. continues again to govern the phenomenon on the analytic arc CD. At the point D appears another critical point and another discontinuity DA takes place, thus completing the piecewise analytic cycle $ABCDA$.

An interesting case arises if resistor R (Fig. 23) is replaced by an inductance L. An application of Kirkhoff's laws results in the following system of d.e.

$$\dot{x} = y/\psi(x), \qquad \dot{y} = x/b + cy/\psi(x)$$

where $\psi(x) = \phi'(x) - d$; $x = kri$. In these notations $\phi(x)$ is the transfer characteristic of the electron tube and b, c, and d are certain constants. It is seen that in this case one has two d.e. of the first order *which cannot be reduced to the d.e. of van der Pol.* Moreover, the critical point occurs when $\phi'(x) = d\phi/dx$ becomes equal to d. Hence, whenever this occurs, the d.e. cease to govern the oscillatory phenomenon, and the subsequent discontinuous stretch is determined by the condition of Mandelstam. We refer to the Andronov and Chaikin book in which a detailed treatment of this and a number of other cases of relaxation oscillations is made on the basis of the discontinuous theory. It is to be noted also that all these conclusions have been verified by means of a cathode-ray oscillograph, at least qualitatively.

7. Asymptotic theory

In the asymptotic theory no use is made of the degenerate equations but, instead, the full d.e. is studied with a view to carrying out the passage to the limit $\mu \to \infty$ in the equation of integral curves. The conclusions are pretty much the same as in the discontinuous theory.

The advantage of this approach is in that the somewhat delicate argument of the theory of degeneration is avoided but the disadvantage (as compared to the discontinuous theory) is in that the phase-plane representation becomes impossible when the system reduces to two degrees of freedom.

As an example we shall consider the case investigated by Flanders and Stoker [71] of a d.e.

$$\ddot{x} + \mu F(\dot{x}) + x = 0 \tag{7}$$

The d.e. of integral curves here is

$$\frac{dy}{dx} = -[\mu F(y) + x]/y \tag{8}$$

with the variable $\xi = x/\mu$ it becomes

$$\frac{dy}{d\xi} = -\frac{\mu^2 [F(y) + \xi]}{y} \tag{9}$$

The equivalent system of two d.e. of the first order is

$$\text{(10)} \qquad \frac{d\xi}{dt} = y/\mu, \qquad \frac{dy}{dt} = -\mu\,[F(y) + \xi]$$

We shall consider in this example the case where μ is large. If one assumes for $F(y)$ the form: $F(y) = -y + \frac{1}{3}y^3$, the d.e. becomes

$$\text{(11)} \qquad \frac{dy}{d\xi} = \mu^2\left(y - \frac{1}{3}y^3 - \xi\right) = \mu^2\, G(\xi, y)$$

If μ is very large, $dy/d\xi$ is very large everywhere except on the curve $G = 0$ where it is zero. Fig. 24 shows the curve $G(\xi, y) = 0$ consisting of three branches 1, 2, and 3 separated by the points A and B of coordinates $(\frac{2}{3}, 1)$ and $(-\frac{2}{3}, -1)$ respectively at which the tangent to $G(\xi, y)$ is vertical. The direction field $dy/d\xi$ is obtained by investigating the sign of $G(\xi, y)$ slightly above and below each of the three branches. One finds that on branches 1 and 3 the direction field is *toward* these branches and on 2 is away from it, as is shown by arrows. Thus 1 and 3 may be regarded as stable branches whereas the branch 2 is unstable. On the curves 1, 2, and 3 themselves, the direction field is horizontal. Now if μ is very large, the direction field is practically vertical everywhere except within a very narrow band surrounding the curve.

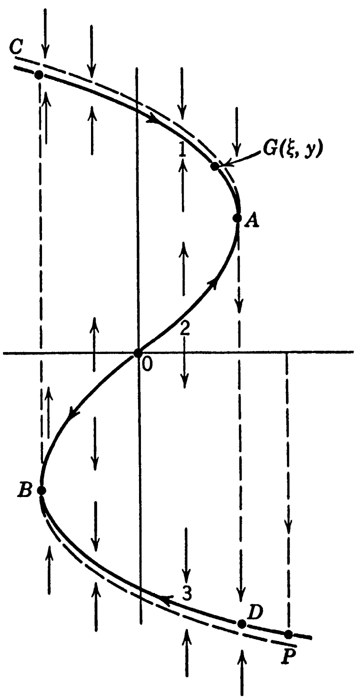


Fig. 24. Piecewise analytic circuit on the basis of the asymptotic theory.

To investigate the *motion* of R on G, one has to consider the system (10). One observes that R cannot follow exactly the curve G, as on this curve $dy/d\xi = 0$. It can, however, follow a curve slightly above 1 and slightly below 3 (broken line) on which the contribution of $dy/d\xi \neq 0$ renders this following possible. These *actual* trajectories are the nearer to the curve $G(\xi, y) = 0$, the larger is μ. The rest of the discussion is similar to that which we used on several occasions previously. One begins the discussion by establishing the positive directions on trajectories by considering the signs of $d\xi/dt$ and dy/dt in the various portions of the curve $G(\xi, y) = 0$. In this manner R follows the branch 3 from some point P toward B. But at B appears a kind of "analytical deadlock." In fact, B is not a point of equilibrium; on the other hand, R cannot turn on the branch 2, as the positive direction on that branch is from 0 to B. Hence, the only issue from this situation for R is to jump on branch 1 inasmuch as the direction field leads in that direction (point C). At the point C the point R is again on the integral curve (branch 1) which is followed to the point A, at which another jump AD takes place, thus completing the piecewise analytic cycle $ADBCA$.

It is useful to compare these results obtained on the basis of the asymptotic theory of this kind with a similar result obtained by the discontinuous theory.

In the first place, the piecewise analytic nature of the resultant motion remains exactly the same. The cycle consists of two analytic branches on which the motion is governed by the d.e., and of two discontinuities traversed in no time. On the discontinuous stretches the d.e. now governs the motion. On the basis of the discontinuous theory these stretches are traversed in accordance with the condition of Mandelstam, whereas, according to the asymptotic theory, this happens merely because there exists an appropriate direction field capable of effecting this jump. It might seem, therefore, that the asymptotic theory leads more logically to the conclusion.

In reality the matter is particularly simple here (for the asymptotic theory) only because the determination of the direction field for the van der Pol equation is simple. This is generally more complicated for an arbitrary relaxation circuit, particularly that involving two degrees of freedom where the asymptotic theory leads inevitably to a four-dimensional phase space, whereas the discontinuous theory operating with the degenerate equations is still capable of yielding simple conclusions in a phase plane.

It must be admitted that, in spite of its being more attractive in the initial stages of the argument, the asymptotic theory has not progressed so far as to yield a uniform representation of the relaxation phenomena as is accomplished by the discontinuous theory.

8. Concluding remarks

From this brief review it is seen that the present status of the theory of relaxation oscillations is less satisfactory than the theory concerning the nearly linear domain where all known phenomena are logically connected with the analytic theory.

The very existence of two different trends in these studies—the discontinuous and the analytic—reflects the difficulty of this problem.

If one judges a theory on the basis of its "convenience" (*commodité*, in the sense of Poincaré) [78], one cannot deny that the discontinuous theory is more *convenient* inasmuch as it is nearer to the real quasi-discontinuous character of the problem. If one can ascertain that the d.e. are of the form (1), with a possibility of setting $T(x_0, y_0) = 0$, one is certain that the discontinuities will appear in the cycle. Once this is established, the "direction" of the jumps follows from the physical properties of its parameters. If the system has either L or m degeneration (negligible inductance in an electric system; negligible mass in a mechanical system), the energy stored in the system is either purely electrostatic (for an electric system) or purely potential (for a mechanical system). This immediately gives an idea as to the "direction" of possible discontinuities in the phase plane.

In this case, and for an electric system, the jumps can occur only in the direction which corresponds to the constant voltage across the capacitor and in a mechanical system to that in which the velocity changes much at a practically constant coordinate. Opposite effects occur in the case of a C or k degeneration (infinite capacity for an electric system or negligible restoring force for the mechanical one). Combined with the nonlinear characteristics of the system, this procedure gives a simple means of determining the piecewise analytic cycle, at least qualitatively.

It is to be admitted, however, that this method may appear somewhat disappointing as contrasted with purely analytic methods used in the "nearly linear" domain. The work of Cartwright and Littlewood has shown, however, that even here an analytic approach is possible but only at the expense of a much greater amount of labor and under the present limitation of the analysis to the van der Pol equation which is not, apparently, the most general form of the d.e. capable of describing the relaxation processes. It is possible that a purely analytical approach can be extended also to the d.e. more closely connected with the relaxation phenomena, but no such attempts were made so far. Even if one succeeds in doing this, one can question as to whether an analytic description of an oscillatory phenomenon which by its very nature exhibits essentially nonanalytic features, at least at some points of its cycle, will be a simple matter.

It is recalled that at a certain time Hertz tried to explain the mechanism of shocks on the basis of a continuous theory by considering two different d.e.—one governing the motion before and after the separation of colliding bodies, the other during the (short) time when these bodies are in contact with each other. It is sufficient to assume the continuity of the solutions at the cost of the loss of analyticity at points where one d.e. replaces the other. In spite of the possibility of accomplishing this result, this theory was ultimately given up in favor of the present discontinuous idealization which is now classical in the theoretical mechanics. It was thus "*commodité*," in the sense of Poincaré, which gave the preference to the ultimate *discontinuous* theory of shocks.

It is quite probable that similar considerations may eventually be a deciding factor in the formation of the ultimate theory of relaxation oscillations, but one has to admit that the last word in this difficult field has not been said as yet.

BIBLIOGRAPHY

A complete bibliography on the subject of nonlinear oscillations can be found in the following lists of references. We indicate below only such references which relate to the various topics mentioned in this review.

Lists of general references

[a] N. Minorsky, *Introduction to Non-Linear Mechanics*, J. W. Edwards, Ann Arbor (1946).

[b] M. L. Cartwright, *Report to the British Association for Advancement of Science*, London (1949).

[c] J. J. Stoker, *Non-Linear Vibrations*, Interscience Publishers, New York (1950).

[d] K. Klotter, Neuere Methoden und Ergebnisse auf dem Gebiet nichtlinearer Schwingungen, *V.D.I. Berichte* (1955).

[e] G. Sansone and R. Conti, Equazioni differenziali non-lineari, *Cons. Naz. delle Ricerche*, Rome (1956).

[f] S. Lefschetz, *Lectures on D.E.*, Princeton University Press, Princeton (new edition to appear shortly).

References relative to topics outlined in this review

[1] B. van der Pol, *Radio Rev.*, **1** (1920); *Proc. I.R.E.*, **22** (1934).
E. V. Appleton, B. van der Pol, *Phil. Mag.*, **43** (1922).

[2] E. and H. Cartan, *Ann. Post. Télégr. Téléph.*, Dec. (1925).

[3] A. Liénard, *Rev. Gen. Élect.*, **23** (1928).

[4] A. Andronov, *C.R. Acad. Sci. Paris*, **189** (1929).

[5] H. Poincaré, Sur les courbes définies par une equation différentielle, *J. Math Pures Appl.*, (1885); also *Oeuvres*, **1**, Gauthiers-Villars, Paris.

[6] L. Mandelstam, N. Papalexi, etc., *J. Tech. Phys., Moscow*, **2** (1935); also reference [a].

[7] L. Mandelstam, N. Papalexi, *Z. Phys.*, **73** (1932).

[8] J. B. Kobsarev, *J. Tech. Phys., Moscow*, **3** (1933).

[9] B. Van der Pol, reference [1]; also *Phil. Mag.*, **2**, November (1929).
[10] J. Shohat, *J. Appl. Phys.*, **14** (1943); **15** (1944).
N. G. de Bruijn, *Phillips Res. Rep.*, **1** (1946).
[11] S. Chaikin, L. Lochakov, *J. Tech. Phys., Moscow*, **5** (1935); also reference [12].
[12] A. Andronov, S. Chaikin, *Theory of Oscillations*, Moscow, (1937); English translation by S. Lefschetz, Princeton University Press, Princeton (1949).
[13] M. L. Cartwright, J. E. Littlewood, *J. Lond. Math. Soc.*, **20** (1945); *Proc. Camb. Phil. Soc.*, **45** (1949); *Ass. Math.*, **48** (1947).
[14] R. Bellman, Project Rand, *Report R-256*, March (1954); also reference [15].
[15] N. Minorsky, *VIIth Intern. Congr. Appl. Mech.*, London (1948).
[16] K. Theodorchik, *Self-Oscillatory Systems* (in Russian), Moscow (1948); also N. Minorsky, *Symposium on Non-Linear Circuit Analysis*, Polytechnic Institute of Brooklyn, New York (1953).
[17] See reference 12.
[18] L. Bieberbach, *Differenzialgleichungen*, Julius Springer, Berlin (1930).
F. Tricomi, *Equazioni differenziali*, chapter 2, Einaudi, Torino (1948).
E. Picard, *Traite d'analyse*, T. III, chapter IX, Gauthier-Villars, Paris (1928).
[19] See reference [18], Bieberbach, p. 109.
[20] M. Frommer, *Math. Ann.*, **99** (1928).
B. Nemitzky, B. Stepanov, *Theory of Differential Equations* (in Russian), Moscow (1949); English edition by S. Lefschetz in preparation.
G. Sansone, *Equazioni differenziali nel campo reale*, Zanichelli, Bologna (1948).
[21] See reference [5]; also references [*a*] and [*c*].
[22] See reference [5]; also reference [*a*].
[23] See reference [5]; also reference (*e*).
I. Bendixson, *Acta Math.*, **24** (1901).
[24] See reference [5]; also references [*a*], [*c*], [*e*], and [*f*].
[25] J. La Salle, see appendix in reference [12], English edition.
[26] H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*, **1**, chapter IV, Gauthiers-Villars, Paris (1892).
A. Liapounov, Problème général de la stabilité du mouvement, *Ann. Fac. Sci. Toulouse*, **9** (1907).
I. G. Malkin, *Theory of Stability of Motion* (in Russian), Moscow (1949).
[27] H. Poincaré, reference [26], chapters III and IV.
[28] H. Poincaré, reference [26], p. 180.
[29] N. Minorsky, reference [*a*], chapter 4.
J. J. Stoker, reference [*c*], appendix 5; also reference [*e*].
[30] H. Poincaré, reference [26], p. 186.
[31] A. Liapounov, reference [26]; also reference [12].
[32] N. Minorsky, *C.R. Acad. Sci. Paris*, **237** (1953).
[33] A. Liapounov, reference [26].
I. G. Malkin, *Stability of Motion* (in Russian), Gostechisdat (1952).
N. G. Tshetajev, *ibid.* (1946).
Summary of these publications is given in J. T. Massera's lectures, University of Rome (1954).
[34] A. Andronov, reference [12].
[35] B. van der Pol, reference [9].
[36] N. Kryloff, N. Bogoliuboff, *Introduction to Non-Linear Mechanics* (in Russian), Kieff (1937); English version by S. Lefschetz, Princeton University Press, Princeton (1943); also see reference [*a*], chapter 10.

[37] A. Lindstedt, *Mém. Acad. Sci. St. Petersbourg*, **31** (1883); also reference [*a*], chapter 11.
[38] N. Kryloff, N. Bogoliuboff, reference 36; also reference [*a*], chapter XI.
[39] N. Minorsky, *C.R. Acad. Sci. Paris*, **228** (1949); also *Bull. Soc. Franc. Mec.* (1954).
[40] H. Poincaré, reference [26], **1.**
[41] B. van der Pol, *Phil. Mag.*, **2** (1926); **3** (1927).
[42] See reference [36]; also reference [*a*], chapter X.
[43] See reference [7].
[44] See reference [39].
[45] B. van der Pol, *Phil. Mag.*, **3,** January (1922).
[46] A. Andronov-Witt, *Arch. Elektrotech.*, **24** (1930).
[47] See reference [7].
[48] Communicated by Prof. B. van der Pol.
[49] See reference [45].
[50] K. Theodorchik and S. Chaikin, *J. Tech. Phys., Moscow*, **2** (1932).
[51] J. Haag, *Les mouvements vibratoires*, **1,** Presses Universitaires, Paris (1952).
[52] Lord Rayleigh, *Phil. Mag.*, **15,** April (1883).
[53] See reference [6].
[54] N. Minorsky, *J. Appl. Phys.*, **22** (1951).
[55] N. Minorsky, *J. Franklin Inst.*, **254,** July (1952).
[56] N. Minorsky, *Colloq. Int. Porquerolles*, no. 281, Ministère de l'air, Paris (1951).
[57] H. Helmholtz, *Sensations of Tone*, Longmans, Green & Company, Ltd., London (1895).
[58] See reference [7].
[59] G. Duffing, Sammlung Vieweg 41/42, Braunschweig (1918).
[60] See reference [*c*].
[61] See reference [36]; also reference [*a*], chapter XV.
[62] See reference [8].
[63] See reference [14].
[64] L. Pontriagin, see reference in [63].
[65] N. Minorsky, *VIIth Intern. Congr. Appl. Mech.*, London (1948); also *C.R. Acad. Sci. Paris*, **226** (1948).
[66] N. Minorsky, *J. A.S.M.E.*, December (1946).
[67] See references [68] and [69].
[68] S. Chaikin, *J. Tech. Phys., Moscow*, **7** (1930).
S. Chaikin, L. Lochakov, *ibid.*, **11** (1935); also references [*a*] and [12].
[69] T. Vogel, *R.C. Semin. Mat. Padova* (1951); *Colloq. Int. Porquerolles* (1951); *Ann. Télécomm.*, **6** (1951).
[70] K. O. Friedrichs, W. R. Wasow, *Duke Math. J.*, **13** (1946).
N. Levinson, *Acta Math.*, **82.**
[71] D. A. Flanders, J. J. Stoker, *Studies in Non-Linear Vibration Theory*, New York University, New York (1946).
[72] J. Haag, *C.R. Acad. Sci. Paris*, **292** (1936); *Ann. Sci. Ecole Norm. Sup.*, **60** (1943); **61** (1944).
[73] M. L. Cartwright, J. E. Littlewood, *J. Lond. Math. Soc.*, **20** (1945); *Ann. Math.*, **48** (1949).
M. L. Cartwright, *Proc. Camb. Phil. Soc.*, **45** (1949).
[74] J. Boussinesq, Applications des potentiels, Paris (1885).
[75] See reference [3].

[76] Ph. Le Corbeiller, *J. Inst. Elec. Eng.*, **79**, London (1936).
[77] M. L. Cartwright, *J. Res. Nat. Bur. Stand.*, **45** (1950); *J. Inst. Elec. Eng.* **95**, London (1948).
[78] H. Poincaré, *La science et l'hypothese*, Flammarion, Paris (1902).
[79] I. G. Malkin, *Certain Problems in the Theory of Non-Linear Oscillations* (in Russian), Moscow (1956).
[80] N. Bogoliuboff, J. Mitropolsky, *Asymptotic Methods in the Theory of Non-Linear Oscillations* (in Russian), Moscow (1955).
[81] J. A. Mitropolsky, *Non-Stationary Processes in Non-Linear Oscillatory Systems* (in Russian), Moscow (1955).
[82] A. Andronov, L. S. Pontriagin, *C.R. Acad. Sci., Moscow*, **14** (1937).
H. F. De Baggis, *Contributions to the Theory of Non-Linear Oscillations*, **2**, Princeton University Press, Princeton (1952).
[83] See reference [82], other papers in this collection.

AUTHOR INDEX
SOME RECENT ADVANCES IN THE DYNAMICS OF RIGID BODIES AND CELESTIAL MECHANICS

Subject Index