CHAPTER V

ORDINARY DIFFERENTIAL EQUATIONS

§1. Introduction

Examples of differential equations. The equations that we have encountered up to now have been for the most part concerned with finding the numerical value of one magnitude or another. When, for example, in the search for maxima and minima of functions, we solved an equation and found those points for which the rate of change of a function vanishes, or when in Chapter IV we considered the problem of finding the roots of polynomials, we were in each case looking for isolated numbers. But in the applications of mathematics there often arise problems of a qualitatively different sort, in which the unknown is itself a function, a law expressing the dependence of certain variables on others. For example, in investigating the process of the cooling of a body, our task is to determine how its temperature will change in the course of time; to describe the motion of a planet or a star we must determine the dependence of their coordinates on time, and so forth.

We can quite often construct an equation for finding the required unknown functions, such equations being called functional equations. The nature of these may, generally speaking, be extremely varied; in fact, it may be said that we have already met the simplest and most primitive functional equations when we were considering implicit functions.

The problem of finding unknown functions will concern us in Chapters V, VI, and VII. In the present chapter, and in the following one, we will consider the most important class of equations serving to determine such functions, namely *differential equations*; that is, equations in which not only the unknown function occurs, but also its derivatives of various orders.

The following equations may serve as examples:

$$\frac{dx}{dt} + P(t)\,x = Q(t), \frac{d^2x}{dt^2} + m^2x = A \sin \omega t, \frac{d^2x}{dt^2} = tx,$$

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \tag{1}$$

In the first three of these, the unknown function is denoted by the letter x and the independent variable by t; in the last three, the unknown function is denoted by the letter u and it depends on two arguments, x and t, or x and y.

The great importance of differential equations in mathematics, and especially in its applications, is due chiefly to the fact that the investigation of many problems in physics and technology may be reduced to the solution of such equations.

Calculations involved in the construction of electrical machinery or of radiotechnical devices, computation of the trajectory of projectiles, investigation of the stability of an aircraft in flight, or of the course of a chemical reaction, all depend on the solution of differential equations.

It often happens that the physical laws governing a phenomenon are written in the form of differential equations, so that the differential equations themselves provide an exact quantitative (numerical) expression of these laws. The reader will see in the following chapters how the laws of conservation of mass and of heat energy are written in the form of differential equations. The laws of mechanics discovered by Newton allow one to investigate the behavior of any mechanical system by means of differential equations.

Let us illustrate by a simple example. Consider a material particle of mass m moving along an axis Ox, and let x denote its coordinate at the instant of time t. The coordinate x will vary with the time, and knowledge of the entire motion of the particle is equivalent to knowledge of the functional dependence of x on the time t. Let us assume that the motion is caused by some force F, the value of which depends on the position of the particle (as defined by the coordinate x), on the velocity of motion $v = dx/dt$ and on the time t, i.e., $F = F(x, dx/dt, t)$. According to the laws of mechanics, the action of the force F on the particle necessarily produces an acceleration $w = d^2x/dt^2$ such that the product of w and the mass m of the particle is equal to the force, and so at every instant of the motion we have the equation

$$m\frac{d^2x}{dt^2} = F\left(x, \frac{dx}{dt}, t\right). \tag{2}$$

This is the differential equation that must be satisfied by the function $x(t)$ describing the behavior of the moving particle. It is simply a representation of laws of mechanics. Its significance lies in the fact that it enables us to reduce the mechanical problem of determining the motion of a particle to the mathematical problem of the solution of a differential equation.

Later in this chapter, the reader will find other examples showing how the study of various physical processes can be reduced to the investigation of differential equations.

The theory of differential equations began to develop at the end of the 17th century, almost simultaneously with the appearance of the differential and integral calculus. At the present time, differential equations have become a powerful tool in the investigation of natural phenomena. In mechanics, astronomy, physics, and technology they have been the means of immense progress. From his study of the differential equations of the motion of heavenly bodies, Newton deduced the laws of planetary motion discovered empirically by Kepler. In 1846 Leverrier predicted the existence of the planet Neptune and determined its position in the sky on the basis of a numerical analysis of the same equations.

To describe in general terms the problems in the theory of differential equations, we first remark that every differential equation has in general not one but infinitely many solutions; that is, there exists an infinite set of functions that satisfy it. For example, the equation of motion for a particle must be satisfied by any motion induced by the given force $F(x, dx/dt, t)$, independently of the starting point or the initial velocity. To each separate motion of the particle there will correspond a particular dependence of x on time t. Since under a given force F there may be infinitely many motions the differential equation (2) will have an infinite set of solutions.

Every differential equation defines, in general, a whole class of functions that satisfy it. The basic problem of the theory is to investigate the functions that satisfy the differential equation. The theory of these equations must enable us to form a sufficiently broad notion of the properties of all functions satisfying the equation, a requirement which is particularly important in applying these equations to the natural sciences. Moreover, our theory must guarantee the means of finding numerical values of the functions, if these are needed in the course of a computation. We will speak later about how these numerical values may be found.

If the unknown function depends on a single argument, the differential equation is called an *ordinary differential equation*. If the unknown function depends on several arguments and the equation contains derivatives with respect to some or all of these arguments, the differential equation is called a *partial differential equation*. The first three of the equations in (1) are ordinary and the last three are partial.

The theory of partial differential equations has many peculiar features which make them essentially different from ordinary differential equations. The basic ideas involved in such equations will be presented in the next chapter; here we will examine only ordinary differential equations.

Let us consider some examples.

Example 1. The law of decay of radium says that the rate of decay is proportional to the initial amount of radium present. Suppose we know that a certain time $t = t_0$ we had R_0 grams of radium. We want to know the amount of radium present at any subsequent time t.

Let $R(t)$ be the amount of undecayed radium at time t. The rate of decay is given by the value of $-(dR/dt)$. Since this is proportional to R, we have

$$-\frac{dR}{dt} = kR, \tag{3}$$

where k is a constant

In order to solve our problem, it is necessary to determine a function from the differential equation (3). For this purpose we note that the function inverse to $R(t)$ satisfies the equation

$$-\frac{dt}{dR} = \frac{1}{kR}, \tag{4}$$

since $dt/dR = (1/dR)/dt$. From the integral calculus it is known that equation (4) is satisfied by any function of the form

$$t = -\frac{1}{k}\ln R + C,$$

where C is an arbitrary constant. From this relation we determine R as a function of t. We have

$$R = e^{-kt+kC} = C_1 e^{-kt}. \tag{5}$$

From the whole set of solutions (5) of equation (3) we must select one which for $t = t_0$ has the value R_0. This solution is obtained by setting $C_1 = R_0 e^{kt_0}$.

From the mathematical point of view, equation (3) is the statement of a very simple law for the change with time of the function R; it says that the rate of decrease $-(dR/dt)$ of the function is proportional to the value of the function R itself. Such a law for the rate of change of a function is

satisfied not only by the phenomena of radioactive decay but also by many other physical phenomena.

We find exactly the same law for the rate of change of a function, for example, in the study of the cooling of a body, where the rate of decrease in the amount of heat in the body is proportional to the difference between the temperature of the body and the temperature of the surrounding medium, and the same law occurs in many other physical processes. Thus the range of application of equation (3) is vastly wider than the particular problem of the radioactive decay from which we obtained the equation.

Example 2. Let a material point of a mass m be moving along the horizontal axis Ox in a resisting medium, for example in a liquid or a gas, under the influence of the elastic force of two springs, acting under Hooke's law (figure 1), which states that the elastic force acts toward the

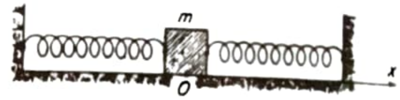

FIG. 1.

position of equilibrium and is proportional to the deviation from the equilibrium position. Let the equilibrium position occur at the point $x = 0$. Then the elastic force is equal to $-bx$ where $b > 0$.

We will assume that the resistance of the medium is proportional to the velocity of motion, i.e., equal to $-a(dx/dt)$, where $a > 0$ and the minus sign indicates that the resisting medium acts against the motion. Such an assumption about the resistance of the medium is confirmed by experiment.

From Newton's basic law that the product of the mass of a material point and its acceleration is equal to the sum of the forces acting on it, we have

$$m\frac{d^2x}{dt^2} = -bx - a\frac{dx}{dt}. \tag{6}$$

Thus the function $x(t)$, which describes the position of the moving point at any instant of time t, satisfies the differential equation (6). We will investigate the solutions of this equation in one of the later sections.

If, in addition to the forces mentioned, the .naterial point is acted upon by still another force, F outside of the system, then the equation of motion (6) takes the form

$$m\frac{d^2x}{dt^2} = -bx - a\frac{dx}{dt} + F \tag{6'}$$

Example 3. A mathematical pendulum is a material point of mass m, suspended on a string whose length will be denoted by l. We will assume that at all stages the pendulum stays in one plane, the plane of the drawing (figure 2). The force tending to restore the pendulum to the vertical position OA is the force of gravity mg, acting on the material point. The position of the pendulum at any time t is given by the angle ϕ by which it differs from the vertical OA. We take the positive direction of ϕ to be counterclockwise. The arc $AA' = l\phi$ is the distance moved by the material point from the position of equilibrium A. The velocity of motion v will be directed along the tangent to the circle and will have the following numerical value:

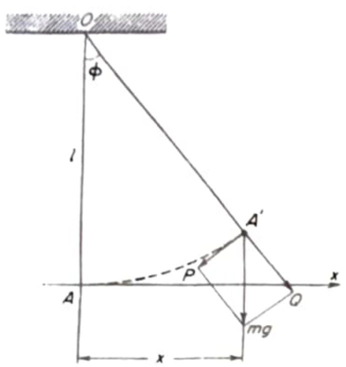

FIG. 2.

$$v = l\frac{d\phi}{dt}.$$

To establish the equation of motion, we decompose the force of gravity mg into two components Q and P, the first of which is directed along the radius OA' and the second along the tangent to the circle. The component Q cannot affect the numerical value of the rate v, since clearly it is balanced by the resistance of the suspension OA'. Only the component P can affect the value of the velocity v. This component always acts toward the equilibrium position A, i.e., toward a decrease in ϕ, if the angle ϕ is positive, and toward an increase in ϕ, if ϕ is negative. The numerical value of P is equal to $-mg \sin\phi$, so that the equation of motion of the pendulum is

$$m\frac{dv}{dt} = -mg\sin\phi$$

or

$$\frac{d^2\phi}{dt^2} = -\frac{g}{l}\sin\phi. \tag{7}$$

It is interesting to note that the solutions of this equation cannot be expressed by a finite combination of elementary functions. The set of

elementary functions is too small to give an exact description of even such a simple physical process as the oscillation of a mathematical pendulum. Later we will see that the differential equations that are solvable by elementary functions are not very numerous, so that it very frequently happens that investigation of a differential equation encountered in physics or mechanics leads us to introduce new classes of functions, to subject them to investigation, and thus to widen our arsenal of functions that may be used for the solution of applied problems.

Let us now restrict ourselves to small oscillations of the pendulum for which, with small error, we may assume that the arc AA' is equal to its projection x on the horizontal axis Ox and $\sin\phi$ is equal to ϕ. Then $\phi \approx \sin\phi = x/l$ and the equation of motion of the pendulum will take on the simpler form

$$\frac{d^2x}{dt^2} = -\frac{g}{l}x. \tag{8}$$

Later we will see that this equation is solvable by trigonometric functions and that by using them we may describe with sufficient exactness the "small oscillations" of a pendulum.

Example 4. Helmholtz' acoustic resonator (figure 3) consists of an air-filled vessel V, the volume of which is equal to v, with a cylindrical neck F. Approximately, we may consider the air in the neck of the container as cork of mass

$$m = \rho sl, \tag{9}$$

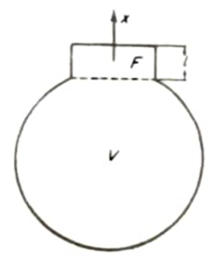

FIG. 3.

where ρ is the density of the air, s is the area of the cross section of the neck, and l is its length. If we assume that this mass of air is displaced from a position of equilibrium by an amount x, then the pressure of the air in the container with volume v is changed from the initial value p by some amount which we will call Δp.

We will assume that the pressure p and the volume v satisfy the adiabatic law $pv^k = C$. Then, neglecting magnitudes of higher order, we have

$$\Delta p \cdot v^k + pkv^{k-1} \cdot \Delta v = 0$$

and

$$\Delta p = -kp\frac{\Delta v}{v} = -\frac{kps}{v}x. \tag{10}$$

(In our case, $\Delta v = sx$.) The equation of motion of the mass of air in the neck may be written as:

$$m \frac{d^2x}{dt^2} = \Delta p \cdot s. \tag{11}$$

Here $\Delta p \cdot s$ is the force exerted by the gas within the container on the column of air in the neck. From (10) and (11) we get

$$\rho l \frac{d^2x}{dt^2} = -\frac{kps}{v} x, \tag{12}$$

where ρ, p, v, l, k, and s are constants.

Example 5. An equation of the form (6) also arises in the study of electric oscillations in a simple oscillator circuit. The circuit diagram is given in (figure 4). Here on the left we have a condenser of capacity C, in series with a coil of inductance L, and a resistance R. At some instant let the condenser have a voltage across its terminals. In the absence of inductance from the circuit, the current would flow until such time as the terminals of the condenser were at the same potential. The presence of an inductance alters the situation, since the circuit will now generate electric oscillations. To find a law for these oscillations, we denote by $v(t)$, or simply by v, the voltage across the condenser at the instant t, by $I(t)$ the current at the instant t, and by R the resistance. From well-known laws of physics, $I(t)R$ remains constantly equal to the total electromotive force, which is the sum of the voltage across the condenser and the inductance $-L(dI/dt)$. Thus,

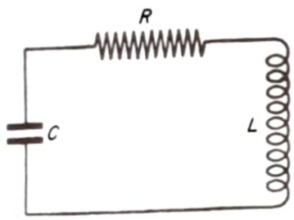

FIG. 4.

$$IR = -v - L\frac{dI}{dt}. \tag{13}$$

We denote by $Q(t)$ the charge on the condenser at time t. Then the current in the circuit will, at each instant, be equal to dQ/dt. The potential difference $v(t)$ across the condenser is equal to $Q(t)/C$. Thus $I = dQ/dt = C\,(dv/dt)$ and equation (13) may be transformed into

$$LC\frac{d^2v}{dt^2} + RC\frac{dv}{dt} + v = 0. \tag{14}$$

Example 6. The circuit diagram of an electron-tube generator of electromagnetic oscillations is shown in figure 5. The oscillator circuit consisting of a capacitance C, across a resistance R and an inductance L, represents the basic oscillator system. The coil L' and the tube shown in the center of figure 5 form a so-called "feedback." They connect a source of energy, namely the battery B, with the L-R-C circuit; K is the cathode of the tube, A the plate, and S the grid. In such an L-R-C circuit "self-oscillations" will arise. For any actual system in an oscillatory state the energy is transformed into heat or is dissipated in some other form to the surrounding bodies, so that to maintain a stationary state of oscillation it is necessary to have an outside source of energy. Self-oscillations differ from other oscillatory processes in that to maintain a stationary oscillatory state of the system the outside source does not have to be periodic. A self-oscillatory system is constructed in such a way that a constant source of energy, in our case the battery B, will maintain a stationary oscillatory state. Examples of self-oscillatory systems are a clock, an electric bell, a string and bow moved by the hand of the musician, the human voice, and so forth.

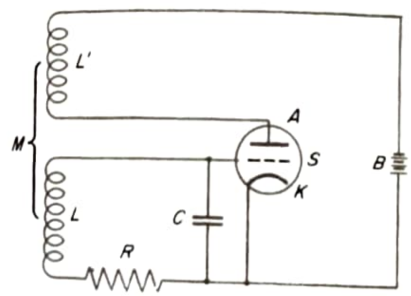

FIG. 5.

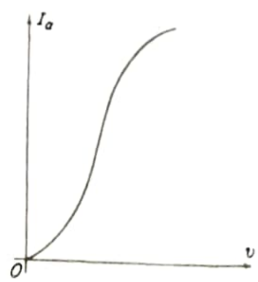

FIG. 6.

The current $I(t)$ in the oscillatory L-R-C circuit satisfies the equation

$$L\frac{dI}{dt} + RI + v = M\frac{dI_a}{dt}. \tag{15}$$

Here $v = v(t)$ is the voltage across the condenser at the instant t, $I_a(t)$ is the plate current through the coil L'; M is the coupling coefficient between the coils L and L'. In comparison with equation (13), equation (15) contains the extra term $M(dI_a/dt)$.

We will assume that the plate current $I_a(t)$ depends only on the voltage between the grid S and the cathode of the tube (i.e., we will neglect the reactance of the anode), so that this voltage is equal to the voltage $v(t)$ across the condenser C. The character of the functional dependence of I_a on v is given in figure 6. The curve as sketched is usually taken to be a cubical parabola and we write an approximate equation for it by:

$$I_a = a_1 v + a_2 v^2 + a_3 v^3.$$

Substituting this into the right side of equation (15), and using the fact that

$$\frac{dv}{dt} = I,$$

we get for v the equation

$$L\frac{d^2v}{dt^2} + [R - M(a_1 + 2a_2 v + 3a_3 v^2)]\frac{dv}{dt} + v = 0. \tag{16}$$

In the examples considered, the search for certain physical quantities characteristic of a given physical process is reduced to the search for solutions of ordinary differential equations.

Problems in the theory of differential equations. We now give exact definitions. *An ordinary differential equation of order n in one unknown function y is a relation of the form*

$$F[x, y(x), y'(x), y''(x), \cdots, y^{(n)}(x)] = 0 \tag{17}$$

between the independent variable x and the quantities

$$y(x), y'(x) = \frac{dy}{dx}, y''(x) = \frac{d^2y}{dx^2}, \cdots, y^{(n)}(x) = \frac{d^ny}{dx^n}.$$

The order of a differential equation is the order of the highest derivative of the unknown function appearing in the differential equation. Thus the equation in example 1 is of the first order, and those in examples 2, 3, 4, 5, and 6, are of the second order.

A function $\phi(x)$ is called a *solution of the differential equation* (17) if substitution of $\phi(x)$ for y, $\phi'(x)$ for y', $\cdots$, $\phi^{(n)}(x)$ for $y^{(n)}$ produces an identity.

Problems in physics and technology often lead to a system of ordinary differential equations with several unknown functions, all depending on the same argument and on their derivatives with respect to that argument.

For greater concreteness, the explanations that follow will deal chiefly with one ordinary differential equation of order not higher than the second and with one unknown function. With this example one may explain the essential properties of all ordinary differential equations and of systems of such equations in which the number of unknown functions is equal to the number of equations.

We have spoken earlier of the fact that, as a rule, every differential equation has not one but an infinite set of solutions. Let us illustrate this first of all by intuitive considerations based on the examples given in equations (2–6). In each of these, the corresponding differential equation is already fully defined by the physical arrangement of the system. But in each of these systems there can be many different motions. For example, it is perfectly clear that the pendulum described by equation (8) may oscillate with many different amplitudes. To each of these different oscillations of the pendulum there corresponds a different solution of equation (8), so that infinitely many such solutions must exist. It may be shown that equation (8) is satisfied by any function of the form

$$x = C_1 \cos\sqrt{\frac{g}{l}}\,t + C_2 \sin\sqrt{\frac{g}{l}}\,t, \tag{18}$$

where C_1 and C_2 are arbitrary constants.

It is also physically clear that the motion of the pendulum will be completely determined only in case we are given, at some instant t_0, the (initial) value x_0 of x (the initial displacement of the material point from the equilibrium position) and the initial rate of motion $x_0' = (dx/dt)|_{t=0}$. These intial conditions determine the constants C_1 and C_2 in formula (18).

In exactly the same way, the differential equations we have found in other examples will have infinitely many solutions.

In general, it can be proved, under very broad assumptions concerning the given differential equation (17) of order n in one unknown function that it has infinitely many solutions. More precisely: If for some "initial value" of the argument, we assign an "initial value" to the unknown function and to all of its derivatives through order $n - 1$, then one can find a solution of equation (17) which takes on these preassigned initial values. It may also be shown that such initial conditions completely determine the solution, so that there exists only one solution satisfying the initial conditions given earlier. We will discuss this question later in more detail. For our present aims, it is essential to note that the initial values of the function and the first $n - 1$ derivatives may be given arbitrarily. We have the right to make any choice of n values which define an "initial state" for the desired solution.

If we wish to construct a formula that will if possible include all solutions of a differential equation of order n, then such a formula must contain n

independent arbitrary constants, which will allow us to impose n initial conditions. Such solutions of a differential equation of order n, containing n independent arbitrary constants, are usually called *general solutions* of the equation. For example, a general solution of (8) is given by formula (18) containing two arbitrary constants; a general solution of equation (3) given by formula (5).

We will now try to formulate in very general outline the problems confronting the theory of differential equations. These are many and varied, and we will indicate only the most important ones.

If the differential equation is given together with its initial conditions, then its solution is completely determined. The construction of formulas giving the solution in explicit form is one of the first problems of the theory. Such formulas may be constructed only in simple cases, but if they are found, they are of great help in the computation and investigation of the solution.

The theory should provide a way to obtain some notion of the behavior of a solution: whether it is monotonic or oscillatory, whether it is periodic or approaches a periodic function, and so forth.

Suppose we change the initial values for the unknown function and its derivatives; that is, we change the initial state of the physical system. Then we will also change the solution, since the whole physical process will now run differently. The theory should provide the possibility of judging what this change will be. In particular, for small changes in the initial values will the solution also change by a small amount and will it therefore be stable in this respect, or may it be that small changes in the initial conditions will give rise to large changes in the solution so that the latter will be unstable?

We must also be able to set up a qualitative, and where possible, quantitative picture of the behavior not only of the separate solutions of an equation, but also of all of the solutions taken together.

In machine construction there often arises the question of making a choice of parameters characterizing an apparatus or machine that will guarantee satisfactory operation. The parameters of an apparatus appear in the form of certain magnitudes in the corresponding differential equation. The theory must help us make clear what will happen to the solutions of the equation (to the working of the apparatus) if we change the differential equation (change the parameters of the apparatus).

Finally, when it is necessary to carry out a computation, we will need to find the solution of an equation numerically, and here the theory will be obliged to provide the engineer and the physicist with the most rapid and economical methods for calculating the solutions.

§2. Linear Differential Equations with Constant Coefficients

For certain important classes of ordinary differential equations the general solution may be expressed in terms of simple well-known functions. One of these classes consists of those differential equations with constant coefficients that are linear with respect to the unknown function and its derivatives (in short, linear). The differential equations (3), (6), (8), and (14) are examples of such equations. A linear equation is called homogeneous if it has no term which does not contain the unknown variable, and nonhomogeneous if there is such a term.

Homogeneous linear equations of the second order with constant coefficients. Such equations have the form

$$m\frac{d^2x}{dt^2} + a\frac{dx}{dt} + bx = 0, \tag{6}$$

where m, a, and b are constants. We will assume that m is positive; this does not restrict the generality, since we can always ensure this situation if need be by changing the sign of all coefficients, provided that $m \neq 0$, which we will assume.

We will look for a solution of this equation in the form of an exponential function $e^{\lambda t}$ and ask how the constant λ should be chosen so that the function $x = e^{\lambda t}$ satisfies the equation. Putting $x = e^{\lambda t}$, $dx/dt = \lambda e^{\lambda t}$ and $d^2x/dt^2 = \lambda^2 e^{\lambda t}$ in the left side of equation (6), we get

$$e^{\lambda t}(m\lambda^2 + a\lambda + b).$$

Thus, in order that $x(t) = e^{\lambda t}$ be a solution of equation (6) it is necessary and sufficient that

$$m\lambda^2 + a\lambda + b = 0. \tag{19}$$

If λ_1 and λ_2 are two real roots of equation (19), then it is easy to prove that a solution of equation (6) is given by every function of the form

$$x = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}, \tag{20}$$

where C_1 and C_2 are arbitrary constants.

Below we will show that formula (20) gives all solutions of equation (6) in the case that equation (19) has distinct real roots.

We note the following important properties of the solution of equation (6):

1. The sum of two solutions is also a solution.
2. A solution multiplied by a constant is also a solution.

In case λ_1 is a multiple root of equation (19), i.e., $m\lambda_1^2 + a\lambda_1 + b = 0$ and $2m\lambda_1 + a = 0$,* then a solution of equation (6) will also be given by the function $te^{\lambda_1 t}$, since if we substitute this function and its derivatives into the left side of equation (6) we get

$$te^{\lambda_1 t}(m\lambda_1^2 + a\lambda_1 + b) + e^{\lambda_1 t}(2m\lambda_1 + a),$$

which is seen from the previous equations to be identically zero.

The general solution of equation (6) in this case has the form

$$x = C_1 e^{\lambda_1 t} + C_2 t e^{\lambda_1 t}. \tag{21}$$

Now let equation (19) have complex roots. These roots will be complex conjugates of each other since m, a, and b are real numbers. Let $\lambda = \alpha \pm i\beta$ The equation

$$m(\alpha + i\beta)^2 + a(\alpha + i\beta) + b = 0$$

is equivalent to the two equations

$$m\alpha^2 - m\beta^2 + a\alpha + b = 0 \text{ and } 2m\alpha\beta + a\beta = 0. \tag{22}$$

It is easy to show that in this case the functions $x = e^{\alpha t}\cos\beta t$ and $x = e^{\alpha t}\sin\beta t$ are solutions of equation (6). Thus, for example, putting the function $x(t) = e^{\alpha t}\cos\beta t$ and its derivatives in the left side of equation (6), we get

$$e^{\alpha t}\cos\beta t(m\alpha^2 - m\beta^2 + a\alpha + b) - e^{\alpha t}\sin\beta t(2m\alpha\beta + a\beta).$$

By equation (22) this expression is identically equal to zero.

The general solution of equation (6), if equation (19) has complex roots, has the form

$$x = C_1 e^{\alpha t}\sin\beta t + C_2 e^{\alpha t}\cos\beta t, \tag{23}$$

where C_1 and C_2 are arbitrary constants.

In this way, if we know the roots of equation (19), called the *characteristic equation*, we can write down the general solution of equation (6).

We note that the general solution of a linear homogeneous equation of order n with constant coefficients

$$a_n \frac{d^n x}{dt^n} + a_{n-1}\frac{d^{n-1}x}{dt^{n-1}} + \cdots + a_1\frac{dx}{dt} + a_0 x = 0$$

may be written in a similar manner as a polynomial in exponential and

* The sum of the roots λ_1 and λ_2 of the quadratic equation (19) is $\lambda_1 + \lambda_2 = -a/m$, and if the roots are the same, that is $\lambda_1 = \lambda_2$, then the second of the previous equations is true.

trigonometric functions, provided we know the roots of the algebraic equation

$$a_n\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0 = 0,$$

which again is called the *characteristic equation*. Thus, the problem of integrating a linear ordinary differential equation with constant coefficients is reduced to an algebraic problem.

We now show that formulas (20), (21), and (23) give all the solutions of equation (6). We note that C_1 and C_2 in these formulas may always be so chosen that the function $x(t)$ satisfies arbitrary initial conditions $x(t_0) = x_0$, $x'(t_0) = x_0'$. For this C_1 and C_2 need only to be determined from the system of equations

$$x_0 = C_1 e^{\lambda_1 t_0} + C_2 e^{\lambda_2 t_0},$$
$$x_0' = \lambda_1 C_1 e^{\lambda_1 t_0} + \lambda_2 C_2 e^{\lambda_2 t_0}.$$

in the case of formula (20), or by two similar equations in the case of formulas (21) and (23). Clearly, if there existed a solution of equation (6) not contained among the solutions we have constructed, then there would exist two distinct solutions of equation (6) satisfying the same initial conditions. Their difference $x_1(t)$ would not be identically zero and would satisfy the zero initial conditions $x_1(t_0) = 0$, $x_1'(t_0) = 0$. We will show that a solution of equation (6) which satisfies the zero initial conditions can only be $x_1(t) = 0$. Let us first show this under the assumption that $m > 0$, $a > 0$, and $b > 0$. We multiply the two sides of the equation

$$m\frac{d^2x_1}{dt^2} + a\frac{dx_1}{dt} + bx_1 = 0 \tag{24}$$

by $2(dx_1/dt)$. Since

$$2\frac{dx_1}{dt}\cdot\frac{d^2x_1}{dt^2} = \frac{d}{dt}\left(\frac{dx_1}{dt}\right)^2 \text{ and } 2x_1(t)\frac{dx_1}{dt} = \frac{d}{dt}(x_1^2),$$

equation (24) may be put in the form

$$\frac{d}{dt}\left[m\left(\frac{dx_1}{dt}\right)^2\right] + 2a\left(\frac{dx_1}{dt}\right)^2 + b\frac{d}{dt}(x_1^2) = 0.$$

Integrating this identity between t_0 and t, we get

$$m\left(\frac{dx_1}{dt}\right)^2 + 2a\int_{t_0}^{t}\left(\frac{dx_1}{dt}\right)^2 dt + bx_1^2(t) = 0.$$

This equation is possible only if $x_1(t) \equiv 0$. Otherwise, for $t = t_0$, we would clearly have a positive quantity on the left and zero on the right, with a similar situation for $t < t_0$.

In order to establish our proposition for all constant coefficients m, a, and b, we consider the function $y_1(t) = x_1(t)e^{-\alpha t}$ which, as it is easy to show, also satisfies the zero boundary condtions. If the value of $\alpha > 0$ is chosen sufficiently large, then the function $y_1(t)$ will satisfy some equation of the form (6) for $a > 0$, $b > 0$, and $m > 0$. This equation is easily derived by substituting the function $x_1(t) = y_1(t)e^{\alpha t}$ and its derivatives into equation (6). Then, as was shown earlier, we have $y_1(t) \equiv 0$, which means that $x_1(t) = y_1(t)e^{\alpha t} \equiv 0$.

Thus we have shown that formulas (20), (21), and (23) give all the solutions of equation (6).

Let us see what information these formulas give about the character of the solutions of equations (6). To this end we note the formulas

$$\lambda_{1,2} = -\frac{a}{2m} \pm \sqrt{\frac{a^2}{4m^2} - \frac{b}{m}} \tag{25}$$

for the roots of equation (19). In accordance with the physical applications which led us to equation (6), we will assume $m > 0$, $a \geqslant 0$, and $b > 0$.

Case 1. $a^2 > 4bm$. The two roots of the characteristic equation (19) are real, negative, and distinct. In this case the function $x(t)$ given by by formula (20) is a general solution of equation (6). All the functions given by this formula together with their first derivatives tend to zero for $t \to +\infty$, and there is no more than one value of t for which they vanish. It follows that the function $x(t)$ has no more than one maximum or minimum. Physically, this means that the resistance of the medium is sufficiently large to prevent oscillations. The moving point cannot pass through the equilibrium position $x = 0$ more than once. From then on, after attaining a maximum distance from the point $x = 0$, it will begin a slow approach to the point but will never pass through it again.

Case 2. $a^2 = 4bm$. The two roots of equation (19) are equal to each other and the general solution of equation (6) given by formula (21). In this case again all solutions $x(t)$ and their first derivatives tend to zero for $t \to +\infty$. Here $x(t)$ and $x'(t)$ cannot vanish more than once. The character of the motion of the material point with abscissa $x(t)$ is the same as in the first case.

Case 3. $a^2 < 4bm$. The roots of the characteristic equation (19) have nonzero imaginary parts. The general solution of equation (6) is given by formula (23). The point x performs oscillations along the x-axis with a constant period $2\pi/\beta$, which is the same for all solutions of (6), and with amplitude $Ce^{\alpha t}$, where $\alpha = -(a/2m)$.

The oscillations of a physical system which take place without the action of an exterior force are called *characteristic oscillations* (eigenvibrations) of the system. From the previous discussion, it follows that the period of such oscillations for the systems discussed in examples 2, 3, 4 and 5, depends only on the structure of the system and will be the same for all oscillations which could possibly arise in it. In example 2 this period is equal to $2\pi\sqrt{b/m - a^2/4m^2}$; in example 4 to $2\pi\sqrt{kps/v_0 l}$; and example 5 to $2\pi\sqrt{1/LC - R^2/4L^2}$.

If $a = 0$, i.e., if the medium offers no resistance to the motion, then the amplitude of the oscillations is constant: the point oscillates harmonically. But if $a > 0$, i.e., if the medium offers resistance to the motion, although this resistance is small ($a^2 < 4bm$), then the amplitude of the oscillations tends to zero and the oscillations die out.

Finally, the solution $x(t) \equiv 0$ of equation (6) in all cases indicates a state of rest for the point x at the position $x = 0$, which is called the position of equilibrium.

If the real parts of both roots of equation (19) are negative, then it can be seen from formulas (20), (21), and (23), that all the solutions of equation (6), together with their derivatives, tend to zero for $t \to +\infty$; that is, the oscillations die out with the passage of time.

However, if the real part of even one of the roots of equation (19) is positive, then there are solutions of equation (6) not tending to zero for $t \to +\infty$, so that some of the solutions of (6) would not even be bounded for $t \to +\infty$. Such a case can occur only for negative b or negative a, if $m > 0$. Physically, this would correspond to the case in which the elastic force does not attract the point x to the equilibrium position but repels it or else that the resistance of the medium is negative. Such cases cannot be realized in the physical examples considered at the beginning of this chapter, but they are entirely realizable in other physical models.

If the real part of the roots λ_1 and λ_2 of equation (19) is equal to zero, which is possible only if the coefficient a in equation (19) is zero, then for $\alpha = 0$ the point $x(t)$, as can be seen from formula (23), carries out harmonic oscillations with bounded amplitude and bounded velocity.

Nonhomogeneous linear equations with constant coefficients. Let us consider in detail the equation

$$m\frac{d^2x}{dt^2} + a\frac{dx}{dt} + bx = A\cos\omega t. \tag{26}$$

This is the equation of linear oscillations of a material point under the action of an elastic force, of the resistance of a medium and of an external periodic force $A \cos \omega t$ (see equation (6') in §1).

Equation (26) is a nonhomogeneous linear equation and (6) is the corresponding homogeneous equation.

We will now look for the general solution to equation (26).

We note that the sum of a solution of a nonhomogeneous equation and a solution of the corresponding homogeneous equation is also a solution of the nonhomogeneous linear equation. Thus, in order to find a general solution of equation (26), it is sufficient to find any one particular solution. The general solution of equation (26) will then be given in the form of the sum of this particular solution and a general solution of the corresponding homogeneous equation.

It is natural to expect that the motion will follow the rhythm of the external periodic force and to look for a particular solution of equation (26) in the form $x = B \cos(\omega t + \delta)$, where B and δ are as yet undetermined constants. We will attempt to determine B and δ in such a way that the function $x = B\cos(\omega t + \delta)$ will satisfy equation (26). Calculating the derivatives $dx/dt = -B\omega \sin(\omega t + \delta)$ and $d^2x/dt^2 = -B\omega^2 \cos(\omega t + \delta)$ and substituting them into equation (26), we get

$$m[-B\omega^2 \cos(\omega t + \delta)] + a[-B\omega \sin(\omega t + \delta)] + bB\cos(\omega t + \delta) = A\cos \omega t.$$

Applying well-known formulas, we have

$$B[(b - m\omega^2)\cos(\omega t + \delta) - a\omega \sin(\omega t + \delta)] = B\sqrt{(b - m\omega^2)^2 + a^2\omega^2}\cos(\omega t + \delta') = A\cos\omega t,$$

where $\delta' = \delta + \gamma$ and $\gamma = \arctan a\omega/(b - m\omega^2)$. Obviously, if we set

$$\delta = -\arctan\frac{a\omega}{b - m\omega^2} \quad \text{and} \quad B = \frac{A}{\sqrt{(b - m\omega^2)^2 + a^2\omega^2}},$$

the function $x = B\cos(\omega t + \delta)$ will satisfy equation (26).

A solution of the form $B\cos(\omega t + \delta)$ will always exist if $(b - m\omega^2)^2 + a^2\omega^2 \neq 0$. In case $(b - m\omega^2)^2 + a^2\omega^2 = 0$, i.e., when $a = 0$ and $b = m\omega^2$, equation (26) has the form

$$m\frac{d^2x}{dt^2} + m\omega^2 x = A\cos\omega t.$$

A particular solution in this case, as is easily established, is $x = (At/2\sqrt{mb})\sin\omega t$.

Solutions of the nonhomogeneous equation (26) are called forced oscillations. The multiplier $\phi(\omega) = 1/\sqrt{(b - m\omega^2)^2 + a^2\omega^2}$ characterizes the relation of the amplitude B of the forced oscillation to the amplitude A of the disturbing force. The graph of the function $\phi(\omega)$ is called the resonance curve. The frequency ω for which $\phi(\omega)$ attains its maximum is called the resonant frequency. Let us calculate it. If $\phi(\omega)$ attains the maximum at $\omega_1 \neq 0$, then for this value of ω the derivative $\phi'(\omega)$ vanishes, i.e., $-4(b - m\omega_1^2)m\omega_1 + 2a^2\omega_1 = 0$, so that $\omega_1 = \sqrt{b/m - a^2/2m^2}$. For this value of ω_1

$$\phi(\omega_1) = \frac{1}{a\sqrt{b/m - a^2/4m^2}}.$$

Hence it can be seen that the amplitude of the forced oscillation for $\omega = \omega_1$ is greater for smaller values of a. For very small a, the frequency ω_1 is very close to the value $\sqrt{b/m}$, i.e., to the frequency the free oscillations. For $a = 0$ and $b = m\omega^2$, as we saw, the forced oscillation has the form

$$x = \frac{At}{2\sqrt{mb}}\sin\omega t,$$

i.e., the amplitude of this oscillation increases beyond all bounds as $t \to +\infty$, a situation which represents the mathematical meaning of resonance. Resonance will occur if the period of the external force is the same as the period of one of the characteristic oscillations of the system. In the practical world, in cases where the period of the external force and the period of the characteristic oscillations are close together, the displacements of the system may become extremly large.

The possibility of large oscillations is often made use of in the construction of various kinds of amplifiers, for example in radio technology. But large oscillations may also lead to the breaking up of structures such as bridges or the framework of machines. Thus it is very important to foresee the possibility of resonance or of oscillations close to it.

From the remarks made earlier, any solution of equation (26) can be written as a sum of the forced oscillation we have found and of one of the solutions of the homogeneous equation given in formulas (20), (21), and (23). For $a > 0$ and $b > 0$ the solution of the homogeneous equation tends to zero for $t \to +\infty$, i.e., any motion eventually approximates the forced oscillations. If $a = 0$ and $b > 0$, the forced oscillation is superposed on a nondecaying characteristic oscillation of the system. For $b = m\omega^2$ and $a = 0$, we have resonance.

If a periodic external force $f(t)$ is imposed on the sytem, the forced oscillations of the system may be found in the following manner. We may represent $f(t)$ with sufficient exactness as a segment of a trigonometric series*

$$\sum_{i=1}^{n} (a_i \cos \omega_i t + b_i \sin \omega_i t). \tag{27}$$

Let us find the forced oscillations corresponding to each term of this sum. Then the oscillation corresponding to the force $f(t)$ will be found by adding together the oscillations corresponding to the various terms of the sum (27). If any of these frequencies is identical with the frequency of a characteristic oscillation of the system, we will have resonance.

§3. Some General Remarks on the Formation and Solution of Differential Equations

There are not many differential equations with the property that all their solutions can be expressed explicitly in terms of simple functions, as is the case for linear equations with constant coefficients. It is possible to give simple examples of differential equations whose general solution cannot be expressed by a finite number of integral of known functions, or as one says, in quadratures.

As Liouville showed in 1841, the solution of the Riccati equation of the form $dy/dx + ay^2 = x^2$, for $a > 0$, cannot be expressed as a finite combination of integrals of elementary functions. So it becomes important to develop methods of approximation to the solutions of differential equations, which will be applicable to wide classes of equations.

The fact that in such cases we find not exact solutions but only approximations should not bother us. First of all, these approximate solutions may be calculated, at least in principle, to any desired degree of accuracy. Second, it must be emphasized that in most cases the differential equations describing a physical process are themselves not altogether exact, as can be seen in all the examples discussed in §1.

An especially good example is provided by the equation (12) for the acoustic resonator. In deriving this equation, we ignored the compressibility of the air in the neck of the container and the motion of the air in the container itself. As a matter of fact, the motion of the air in the neck sets into motion the mass of the air in the vessel, but these two motions have different velocities and displacements. In the neck the displacement of the particles of air is considerably greater than in the container. Thus we ignored the motion of the air in the container, and

* Cf. Chapter XII, §7.

took account only of its compression. For the air in the neck, however, we ignored the energy of its compression and took account only of the kinetic energy of its motion.

To derive the differential equation for a physical pendulum, we ignored the mass of the string on which it hangs. To derive equation (14) for electric oscillations in a circuit, we ignored the self-inductance of the wiring and the resistance of the coils. In general, to obtain a differential equation for any physical process, we must always ignore certain factors and idealize others. In view of this, A. A. Andronov drew especial attention to the fact that for physical investigations we are especially interested in those differential equations whose solutions do not change much for arbitrary small changes, in some sense or another, in the equations themselves. Such differential equations are called "intensitive." These equations deserve particularly complete study. structural stability

It should be stated that in physical investigations not only are the differential equations that describe the laws of change of the physical quantities themselves inexactly defined but even the number of these quantities is defined only approximately. Strictly speaking, there are no such things as rigid bodies. So to study the oscillations of a pendulum, we ought to take into account the deformation of the string from which it hangs and the deformation of the rigid body itself, which we approximated by taking it as a material point. In exactly the same way, to study the oscillations of a load attached to springs, we ought to consider the masses of the separate coils of the springs. But in these examples it is easy to show that the character of the motion of the different particles, which make up the pendulum and its load together with the springs, has little influence on the character of the oscillation. If we wished to take this influence into account, the problem would become so complicated that we would be unable to solve it to any suitable approximation. Our solution would then bear no closer relation to physical reality than the solution given in §1 without consideration of these influences. Intelligent idealization of a problem is always unavoidable. To describe a process, it is necessary to take into account the essential features of the process but by no means to consider every feature without exception. This would not only complicate the problem a great deal but in most cases would result in the impossibility of calculating a solution. The fundamental problem of physics or mechanics, in the investigation of any phenomenon, is to find the smallest number of quantities, which with sufficient exactness describe the state of the phenomenon at any given moment, and then to set up the simplest differential equations that are good descriptions of the laws governing the changes in these quantities. This problem is often very difficult. Which features are the essential ones and which are nonessential is a question that in the final analysis can be decided only by long experience. Only by comparing the answers provided by an idealized argument with the results of experiment can we judge whether the idealization was a valid one.

The mathematical problem of the possibility of decreasing the number of quantities may be formulated in one of the simplest and most characteristic cases, as follows.

Suppose that to begin with we characterize the state of a physical system at time t by the two magnitudes $x_1(t)$ and $x_2(t)$. Let the differential equations expressing their rates of change have the form

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(t, x_1, x_2), \\ \epsilon\frac{dx_2}{dt} &= f_2(t, x_1, x_2), \end{aligned} \tag{28}$$

In the second equation the coefficient of the derivative is a small constant parameter ϵ. If we put $\epsilon = 0$, the second of equations (28) will cease to be a differential equation. It then takes the form

$$f_2(t, x_1, x_2) = 0.$$

From this equation, we define x_2 as a function of t and x_1 and we substitute it into the first of the equations (28). We then have the differential equation

$$\frac{dx_1}{dt} = F(t, x_1)$$

for the single variable x_1. In this way the number of parameters entering into the situation is reduced to one. We now ask, under what conditions will the error introduced by taking $\epsilon = 0$ be small. Of course, it may happen that as $\epsilon \to 0$ the value dx_2/dt grows beyond all bounds, so that the right side of the second of equations (28) does not tend to zero as $\epsilon \to 0$.

§4. Geometric Interpretation of the Problem of Integrating Differential Equations; Generalization of the Problem

For simplicity we will consider initially only one differential equation of the first order with one unknown function

$$\frac{dy}{dx} = f(x, y), \tag{29}$$

where the function $f(x, y)$ is defined on some domain G in the (x, y) plane.

This equation determines at each point of the domain the slope of the tangent to the graph of a solution of equation (29) at that point. If at each point (x, y) of the domain G we indicate by means of a line segment the the direction of the tangent (either of the two directions may be used) as determined by the value of $f(x, y)$ at this point, we obtain a field of directions. Then the problem of finding a solution of the differential equation (29) for the initial conditon $y(x_0) = y_0$ may be formulated thus: In the domain G we have to find a curve $y = \phi(x)$, passing through the point $M_0(x_0, y_0)$, which at each of its points has a tangent whose slope is given by equation (29), or briefly, which has at each of its points a preassigned direction.

From the geometric point of view this statement of the problem has two unnatural features:

1. By requiring that the slope of the tangent at any given point (x, y) of the domain G be equal to $f(x, y)$, we automatically exclude tangents parallel to Oy, since we generally consider only finite magnitudes; in particular, it is assumed that the function $f(x, y)$ on the right side of equation (29) assumes only finite values.

2. By considering only curves which are graphs of functions of x, we also exclude those curves which are intersected more than once by a line perpendicular to the axis Ox, since we consider only single-valued functions; in particular, every solution of a differential equation is assumed to be a single-valued function of x.

So let us generalize to some extent the preceding statement of the problem of finding a solution to the differential equation (29). Namely, we will now allow the tangent at some points to be parallel to the axis Oy. At these points, where the slope of the tangent with respect to the axis Ox has no meaning, we will take the slope with respect to the axis Oy. In other words, we consider, together with the differential equation (29), the equation

$$\frac{dx}{dy} = f_1(x, y), \tag{29'}$$

where $f_1(x, y) = 1/f(x, y)$, if $f(x, y) \neq 0$, using the second equation when the first is meaningless. The problem of integrating the differential equations (29) and (29') then becomes: In the domain G to find all curves having at each point the tangent defined by these equations. These curves will be called integral curves (integral lines) of the equations (29) and (29') or of the tangent field given by these equations. In place of the plural "equations (29), (29')", we will often use the singular "equation (29), (29')". It is clear that the graph of any solution of equation (29) will also be an integral curve of equation (29), (29'). But not every integral curve of equation (29), (29′) will be the graph of a solution of equation (29). This case will occur, for example, if some perpendicular to the axis Ox intersects this curve at more than one point.

In what follows, if it can be clearly shown that

$$f(x, y) = \frac{M(x, y)}{N(x, y)},$$

then we will write only the equation

$$\frac{dy}{dx} = \frac{M(x, y)}{N(x, y)},$$

and omit writing

$$\frac{dx}{dy} = \frac{N(x, y)}{M(x, y)}.$$

Sometimes in place of these equations we introduce a parameter t, and write the system of equations

$$\frac{dx}{dt} = N(x, y), \frac{dy}{dt} = M(x, y),$$

where x and y are considered as functions of t.

Example 1. The equation

$$\frac{dy}{dx} = \frac{y}{x} \tag{30}$$

defines a tangent field everywhere except at the origin. This tangent field is sketched in figure 7. All the tangents given by equation (30) pass through the origin.

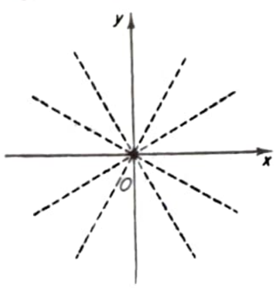

FIG. 7.

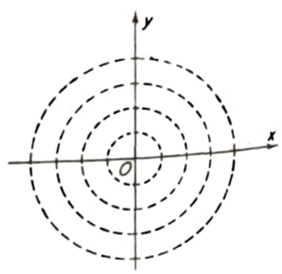

FIG. 8.

It is clear that for every k the function

$$y = kx \tag{31}$$

is a solution of equation (30). The collection of all integral curves of this equation is then defined by the relation

$$ax + by = 0, \tag{32}$$

where a and b are arbitrary constants, not both zero. The axis Oy is an integral curve of equation (30), but it is not the graph of a solution of it.

Since equation (30) does not define a tangent field at the origin, the curves (31) and (32) are, strictly speaking, integral curves everywhere except at the origin. Thus it is more correct to say that the integral curves of equation (30) are not straight lines passing through the origin but half lines issuing from it.

Example 2. The equation

$$\frac{dy}{dx} = -\frac{x}{y} \tag{33}$$

defines a field of tangents everywhere except at the origin, as sketched in figure 8. The tangents defined at a given point (x, y) by equations (30) and (33) are perpendicular to each other. It is clear that all circles centered at the origin will be integral curves of equation (33). However the solutions of this equation will be the functions

$$y = +\sqrt{R^2 - x^2}, y = -\sqrt{R^2 - x^2}, -R \leqslant x \leqslant R.$$

For brevity in what follows we will sometimes say "a solution passes through the point (x, y)" in place of the more exact statement "the graph of a solution passes through the point (x, y)."

§5. Existence and Uniqueness of the Solution of a Differential Equation; Approximate Solution of Equations

The question of existence and uniqueness of the solution of a differential equation. We return to the differential equation (17) of arbitrary order n. Generally, it has infinitely many solutions and in order that we may pick from all the possible solutions some one specific one, it is necessary to attach to the equation some supplementary conditions, the number of which should be equal to the order n of the equation. Such conditions may be of extremely varied character, depending on the physical, mechanical, or other significance of the original problem. For example, if we have to investigate the motion of a mechanical system beginning with some specific initial state, the supplementary conditions will refer to a specific (initial) value of the independent variable and will be called initial conditions of the problem. But if we want to define the curve of a cable in a suspension bridge, or of a loaded beam resting on supports at each end, we encounter conditions corresponding to different values of the independent variable, at the ends of the cable or at the points of support of the beam. We could give many other examples showing the variety of conditions to be fulfilled in connection with differential equations.

We will assume that the supplementary conditions have been defined and that we are required to find a solution of equation (17) that satisfies them. The first question we must consider is whether any such solution exists at all. It often happens that we cannot be sure of this in advance. Assume, say, that equation (17) is a description of the operation of some physical apparatus and suppose we want to determine whether periodic motion occurs in this apparatus. The supplementary conditions will then be conditions for the periodic repetition of the initial state in the apparatus, and we cannot say ahead of time whether or not there will exist a solution which satisfies them.

In any case the investigation of problems of existence and uniqueness of a solution makes clear just which conditions can be fulfilled for a given differential equation and which of these conditions will define the solution in a unique manner. But the determination of such conditions and the proof of existence and uniqueness of the solution for a differential equation corresponding to some physical problem also has great value for the physical theory itself. It shows that the assumptions adopted in setting up the mathematical description of the physical event are on the one hand mutually consistent and on the other constitute a complete description of the event.

The methods of investigating the existence problem are manifold, but among them an especially important role is played by what are called direct methods. The proof of the existence of the required solution is provided by the construction of approximate solutions, which are proved to converge to the exact solution of the problem. These methods not only establish the existence of an exact solution, but also provide a way, in fact the principal one, of approximating it to any desired degree of accuracy.

For the rest of this section we will consider, for the sake of definiteness, a problem with initial data, for which we will illustrate the ideas of Euler's method and the method of successive approximations.

Euler's method of broken lines. Consider in some domain G of the (x, y) plane the differential equation

$$\frac{dy}{dx} = f(x, y). \tag{34}$$

As we have already noted, equation (34) defines in G a field of tangents. We choose any point (x_0, y_0) of G. Through it there will pass a straight line L_0 with slope $f(x_0, y_0)$. On the straight line L_0 we choose a point (x_1, y_1), sufficiently close to (x_0, y_0); in figure 9 this point is indicated by the number 1. We draw the straight line L_1 through the point (x_1, y_1) with slope $f(x_1, y_1)$ and on it mark the point (x_2, y_2); in the figure this point is denoted by the number 2. Then on the straight line L_2 corresponding to the point (x_2, y_2) we mark the point (x_3, y_3), and continue in the same manner with $x_0 < x_1 < x_2 < x_3 < \cdots$. It is assumed, of course, that all the points (x_0, y_0), (x_1, y_1), (x_2, y_2), $\cdots$ are in the domain G. The broken line joining these points is called an Euler broken line. One may also construct an Euler broken line in the direction of decreasing x; the corresponding vertices on our figure are denoted by -1, -2, -3.

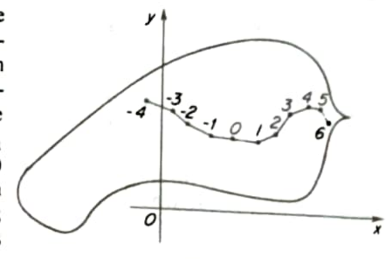

FIG. 9.

It is reasonable to expect that every Euler broken line through the point (x_0, y_0) with sufficiently short segments gives a representation of an integral curve l passing through the point (x_0, y_0), and that with decrease in the length of the links, i.e., when the length of the longest link tends to zero, the Euler broken line will approximate this integral curve.

Here, of course, it is assumed that the integral curve exists. In fact it is not hard to prove that if the function $f(x, y)$ is continuous in the domain G, one may find an infinite sequence of Euler broken lines, the length of the largest links tending to zero, which converges to an integral curve l. However, one usually cannot prove uniqueness: there may exist different sequences of Euler broken lines that converge to different integral curves passing through one and the same point (x_0, y_0). M. A. Lavrent'ev has constructed an example of a differential equation of the form (29) with a continuous function $f(x, y)$, such that in any neighborhood of any point P of the domain G there passes not one but at least two integral curves.

In order that through every point of the domain G there pass only one integral curve, it is necessary to impose on the function $f(x, y)$ certain conditions beyond that of continuity. It is sufficient, for example, to assume that the function $f(x, y)$ is continuous and has a bounded derivative with respect to y on the whole domain G. In this case it may be proved that through each point of G there passes one and only one integral curve and that every sequence of Euler broken lines passing through the point (x_0, y_0) converges uniformly to this unique integral curve, as the length of the longest link of the broken lines tends to zero. Thus for sufficiently small links the Euler broken line may be taken as an approximation to the integral curve of equation (34).

From the preceding it can be seen that the Euler broken lines are so constituted that small pieces of the integral curves are replaced by line segments tangent to these integral curves. In practice, many approximations to integral curves of the differential equation (34) consist not of straight-line segments tangent to the integral curves, but of parabolic segments that have a higher order of tangency with the integral curve. In this way it is possible to find an approximate solution with the same degree of accuracy in a smaller number of steps (with a smaller number of links in the approximating curve). The coefficients of the equation for the (higher order) parabola

$$y = a_0 + a_1(x - x_k) + a_2(x - x_k)^2 + \cdots + a_n(x - x_k)^n, \quad (35)$$

which at the point (x_k, y_k) has nth-order tangency with the integral curves of equation (34) through this point, are given by the following formulas:

$$a_0 = y_k, \quad (36)$$

$$a_1 = \left(\frac{dy}{dx}\right)_{x=x_k} = f(x_k, y_k), \quad (36')$$

$$2a_2 = \left(\frac{d^2y}{dx^2}\right)_{x=x_k} = \left[\frac{df(x, y)}{dx}\right]_{x=x_k} = f'_x(x_k, y_k) + f'_y(x_k, y_k)\left(\frac{dy}{dx}\right)_{x=x_k}$$

$$= f'_x(x_k, y_k) + f'_y(x_k, y_k) f(x_k, y_k), \quad (36'')$$

$$6a_3 = \left(\frac{d^3y}{dx^3}\right)_{x=x_k} = \left\{\frac{d}{dx}[f'_x(x, y(x)) + f'_y(x, y(x)) f(x, y(x))]\right\}_{x=x_k}$$

$$= f''_{xx}(x_k, y_k) + 2f''_{xy}(x_k, y_k) f(x_k, y_k)$$

$$+ f''_{yy}(x_k, y_k) f^2(x_k, y_k) + f'^2_y(x_k, y_k) f(x_k, y_k)$$

$$+ f'_y(x_k, y_k) f'_x(x_k, y_k). \quad (36''')$$

The polynomial (35) is needed only in order to compute its value for $x = x_{k+1}$. The actual values of the coefficients $a_0, a_1, a_2, \cdots, a_n$ themselves are not needed. There are many ways of computing the value for $x = x_{k+1}$ of the polynomial (35) whose coefficients are given by formula (36), without computing the coefficients $a_0, a_1, \cdots, a_n$ themselves.

Other approximation methods exist for finding the solution of the differential equation (34), which are based on other ideas. One convenient method was developed by A. N. Krylov (1863–1945).

The method of successive approximations. We now describe another method of successive approximation, which is as widely used as the method of the Euler broken lines. We assume again that we are required to find a solution $y(x)$ of the differential equation (34) satisfying the initial condition

$$y(x_0) = y_0.$$

For the initial approximation to the function $y(x)$, we take an arbitrary function $y_0(x)$. For simplicity we will assume that it also satisfies the initial condition, although this is not necessary. We substitute it into the right side $f(x, y)$ of the equation for the unknown function y and construct a first approximation y_1 to the solution y from the following requirements:

$$\frac{dy_1}{dx} = f[x, y_0(x)], \; y_1(x_0) = y_0.$$

Since there is a known function on the right side of the first of these equations the function $y_1(x)$ may be found by integration:

$$y_1(x) = y_0 + \int_{x_0}^{x} f[t, y_0(t)]\, dt.$$

It may be expected that $y_1(x)$ will differ from the solution $y(x)$ by less than $y_0(x)$ does, since in the construction of $y_1(x)$ we made use of the differential equation itself, which should probably introduce a correction into the original approximation. One would also think that if we improve the first approximation $y_1(x)$ in the same way, then the second approximation

$$y_2(x) = y_0 + \int_{x_0}^{x} f[t, y_1(t)]\, dt$$

will be still closer to the desired solution.

Let us assume that this process of improvement has been continued indefinitely and that we have constructed the sequence of approximations

$$y_0(x),\; y_1(x),\; \cdots,\; y_n(x),\; \cdots.$$

Will this sequence converge to the solution $y(x)$?

More detailed investigations show that if $f(x, y)$ is continuous and f_y' is bounded in the domain G, the functions $y_n(x)$ will in fact converge to the exact solution $y(x)$ at least for all x sufficiently close to x_0 and that if we break off the computation after a sufficient number of steps, we will be able to find the solution $y(x)$ to any desired degree of accuracy.

Exactly in the same way as for the integral curves of equation (34), we may also find approximations to integral curves of a system of two or more differential equations of the first order. Essentially the necessary condition here is to be able to solve these equations for the derivatives of the unknown functions. For example, suppose we are given the system

$$\frac{dy}{dx} = f_1(x, y, z), \quad \frac{dz}{dx} = f_2(x, y, z). \tag{37}$$

Asuming that the right sides of these equations are continuous and have bounded derivatives with respect to y and z in some domain G in space, it may be shown under these conditions that through each point (x_0, y_0, z_0) of the domain G, in which the right sides of the equations in (37) are defined, there passes one and only one integral curve

$$y = \phi(x), \quad z = \psi(x)$$

of the system (37). The functions $f_1(x, y, z)$ and $f_2(x, y, z)$ give the direction numbers at the point (x, y, z), of the tangent to the integral curve passing through this point. To find the functions $\phi(x)$ and $\psi(x)$ approximately, we may apply the Euler broken line method or other methods similar to the ones applied to the equation (34).

The process of approximate computation of the solution of ordinary differential equations with initial conditions may be carried out on computing machines. There are electronic machines that work so rapidly that if, for example, the machine is programmed to compute the trajectory of a projectile, this trajectory can be found in a shorter space time than it takes for the projectile to hit its target (cf. Chapter XIV).

The connection between differential equations of various orders and a system of a large number of equations of first order. A system of ordinary differential equations, when solved for the derivative of highest order of each of the unknown functions, may in general be reduced, by the introduction of new unknown functions, to a system of equations of the first order, which is solved for all the derivatives. For example, consider the differential equation

$$\frac{d^2y}{dx^2} = f\left(x, y, \frac{dy}{dx}\right). \tag{38}$$

We set

$$\frac{dy}{dx} = z. \tag{39}$$

Then equation (38) may be written in the form

$$\frac{dz}{dx} = f(x, y, z). \tag{40}$$

Hence, to every solution of equation (38) there corresponds a solution of the system consisting of equations (39) and (40). It is easy to show that to every solution of the system of equations (39) and (40) there corresponds a solution of equation (38).

Equations not explicitly containing the independent variable. The problems of the pendulum, of the Helmholtz acoustic resonator, of a simple electric circuit, or of an electron-tube generator considered in §1 lead to differential equations in which the independent variable (time) does not explicitly appear. We mention equations of this type here, because the corresponding differential equations of the second order may be reduced in each case to a single differential equation of the first order rather than to a system of first-order equations as in the paragraph above for the general equation of the second order. This reduction greatly simplifies their study.

Let us then consider a differential equation of the second order, not containing the argument t in explicit form

$$F\left(x, \frac{dx}{dt}, \frac{d^2x}{dt^2}\right) = 0. \tag{41}$$

We set

$$\frac{dx}{dt} = y \tag{42}$$

and consider y as a function of x, so that

$$\frac{d^2x}{dt^2} = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{dy}{dt} = \frac{dy}{dx}\cdot\frac{dx}{dt} = y\frac{dy}{dx}.$$

Then equation (41) may be rewritten in the form

$$F\left(x, y, y\frac{dy}{dx}\right) = 0. \tag{43}$$

In this manner, to every solution of equation (41) there corresponds a unique solution of equation (43). Also to each of the solutions $y = \phi(x)$

of equation (43) there correspond infinitely many solutions of equation (41). These solutions may be found by integrating the equation

$$\frac{dx}{dt} = \phi(x), \tag{44}$$

where x is considered as a function of t.

It is clear that if this equation is satisfied by a function $x = x(t)$, then it will also be satisfied by any function of the form $x(t + t_0)$, where t_0 is an arbitrary constant.

It may happen that not every integral curve of equation (43) is the graph of a single function of x. This will happen, for example, if the curve is closed. In this case the integral curve of equation (43) must be split up into a number of pieces, each of which is the graph of a function of x. For every one of these pieces, we have to find an integral of equation (44).

The values of x and dx/dt which at each instant characterize the state of the physical system corresponding to equation (41) are called the *phases* of the system, and the (x, y) plane is correspondingly called the *phase plane* for equation (41). To every solution $x = x(t)$ of this equation there corresponds the curve

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

in the (x, y) plane; t here is considered as a parameter. Conversely, to every integral curve $y = \phi(x)$ of equation (43) in the (x, y) plane there corresponds an infinite set of solutions of the form $x = x(t + t_0)$ for equation (41); here t_0 is an arbitrary constant. Information about the behavior of the integral curves of equation (43) in the plane is easily transformed into information about the character of the possible solutions of equation (41). Every closed integral curve of equation (43) corresponds, for example, to a periodic solution of equation (41).

If we subject equation (6) to the transformation (42), we obtain

$$\frac{dy}{dx} = \frac{-ay - bx}{my}. \tag{45}$$

Setting $v = x$ and $dv/dt = y$ in equation (16), in like manner we get

$$L\frac{dy}{dx} = \frac{-[R - M(a_1 + 2a_2x + 3a_3x^2)]\,y - x}{y}. \tag{46}$$

Just as the state at every instant of the physical system corresponding to the second-order equation (41) is characterized by the two magnitudes* (phases) x and $y = dx/dt$, the state of a physical system described by equations of higher order or by a system of differential equations is characterized by a larger number of magnitudes (phases). Instead of a phase plane, we then speak of a phase space.

* The values of d^2x/dt^2, d^3x/dt^3, $\cdots$ at the same instant of time are defined by the values of x and dx/dt from equation (41) and from the equations obtained from (45) by differentiation (cf. formula (36)).

§6. Singular Points

Let the point $P(x, y)$ be in the interior of the domain G in which we consider the differential equation

$$\frac{dy}{dx} = \frac{M(x, y)}{N(x, y)}. \tag{47}$$

If there exists a neighborhood R of the point P through each point of which passes one and only one integral curve (47), then the point P is called an *ordinary point* of equation (47). But if such a neighborhood does not exist, then the point P is called a *singular point* of this equation. The study of singular points is very important in the qualitative theory of differential equations, which we will consider in the next section.

Particularly important are the so-called *isolated singular points*, i.e., singular points in some neighborhood of each of which there are no other singular points. In applications one often encounters them in investigating equations of the form (47), where $M(x, y)$ and $N(x, y)$ are functions with continuous derivatives of high orders with respect to x and y. For such equations, all the interior points of the domain at which $M(x, y) \neq 0$ or $N(x, y) \neq 0$ are ordinary points. Let us now consider any interior point (x_0, y_0) where $M(x, y) = N(x, y) = 0$. To simplify the notation we will assume that $x_0 = 0$ and $y_0 = 0$. This can always be arranged by translating the original origin of coordinates to the point (x_0, y_0). Expanding $M(x, y)$ and $N(x, y)$ by Taylor's formula into powers of x and y and restricting ourselves to terms of the first order, we have, in a neighborhood of the point (0, 0),

$$\frac{dy}{dx} = \frac{M'_x(0, 0)\,x + M'_y(0, 0)\,y + \phi_1(x, y)}{N'_x(0, 0)\,x + N'_y(0, 0)\,y + \phi_2(x, y)}, \tag{48}$$

where $\phi_1(x, y)$ and $\phi_2(x, y)$ are functions of x and y for which

$$\lim_{\substack{x\to 0\\ y\to 0}} \frac{\phi_1(x, y)}{\sqrt{x^2 + y^2}} = 0 \quad \text{and} \quad \lim_{\substack{x\to 0\\ y\to 0}} \frac{\phi_2(x, y)}{\sqrt{x^2 + y^2}} = 0.$$

Equations (45) and (46) are of this form. Equation (45) does not define either dy/dx or dx/dy for $x = 0$ and $y = 0$. If the determinant

$$\begin{vmatrix} M'_x(0,0) & M'_y(0,0) \\ N'_x(0,0) & N'_y(0,0) \end{vmatrix} \neq 0,$$

then, whatever value we assign to dy/dx at the origin, the origin will be a point of discontinuity for the values dy/dx and dx/dy, since they tend to different limits depending on the manner of approach to the origin. The origin is a singular point for our differential equation.

It has been shown that the character of the behavior of the integral curves near an isolated singular point (here the origin) is not influenced by the behavior of the terms $\phi_1(x, y)$ and $\phi_2(x, y)$ in the numerator and denominator, provided only that the real part of both roots of the equation

$$\begin{vmatrix} \lambda - M'_y(0,0) & -M'_x(0,0) \\ -N'_y(0,0) & \lambda - N'_x(0,0) \end{vmatrix} = 0 \tag{49}$$

is different from zero. Thus, in order to form some idea of this behavior, we study the behavior near the origin of the integral curves of the equation

$$\frac{dy}{dx} = \frac{ax + by}{cx + dy} \tag{50}$$

for which the determinant

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0.$$

We note that the arrangement of the integral curves in the neighborhood of a singular point of a differential equation has great interest for many problems of mechanics, for example in the investigation of the trajectories of motions near the equilibrium position.

It has been shown that everywhere in the plane it is possible to choose coordinates ξ, η, connected with x, y by the equations

$$\begin{aligned} x &= k_{11}\xi + k_{12}\eta, \\ y &= k_{12}\xi + k_{22}\eta, \end{aligned} \tag{51}$$

where the k_{ij} are real numbers such that equation (50) is tranformed into one of the following three types:

$$1)\ \frac{d\eta}{d\xi} = k\frac{\eta}{\xi}, \quad \text{where} \quad k = \frac{\lambda_2}{\lambda_1}. \tag{52}$$

$$2)\ \frac{d\eta}{d\xi} = \frac{\xi + \eta}{\xi}. \tag{53}$$

$$3)\ \frac{d\eta}{d\xi} = \frac{\beta\xi + \alpha\eta}{\alpha\xi - \beta\eta}. \tag{54}$$

Here λ_1 and λ_2 are the roots of the equation

$$\begin{vmatrix} c - \lambda & d \\ a & b - \lambda \end{vmatrix} = 0. \tag{55}$$

If these roots are real and different, then equation (50) is transformed into the form (52). If these roots are equal, then equation (50) is transformed either into the form (52) or into the form (53), depending on whether $a^2 + d^2 = 0$ or $a^2 + d^2 \neq 0$. If the roots of equation (55) are complex, $\lambda = \alpha \pm \beta i$, then equation (51) is transformed into the form (54).

We will consider each of the equations (52), (53), (54). To begin with, we note the following.

Even though the axes Ox and Oy were mutually perpendicular, the axes $O\xi$ and $O\eta$ need not, in general, be so. But to simplify the diagrams, we will assume they are perpendicular. Further, in the transformation (51) the scales on the $O\xi$ and $O\eta$ axes may be changed; they may not be the same as the ones originally chosen on the axes Ox and Oy. But again, for the sake of simplicity, we assume that the scales are not changed. Thus, for example, in place of the concentric circles, as in figure 8, there could in general occur a family of similar and similarly placed ellipses with common center at the origin.

All integral curves of equation (52) are given by a relation of the form

$$a\eta + b\,|\,\xi\,|^k = 0,$$

where a and b are arbitrary constants.

The integral curves of equation (52) are graphed in figure 10; here we we have assumed that $k > 1$. In this case all integral curves except one, the axis $O\eta$, are tangent at the origin to the axis $O\xi$. The case $0 < k < 1$ is the same as the case $k > 1$ with interchange of ξ and η, i.e., we have only to interchange the roles of the axes ξ and η. For $k = 1$, equation (52) becomes equation (30), whose integral curves were illustrated in figure 7.

An illustration of the integral curves of equation (52) for $k < 0$ is given in figure 11. In this case we have only two integral curves that pass through the point O: these are the axis $O\xi$ and the axis $O\eta$. All other integral

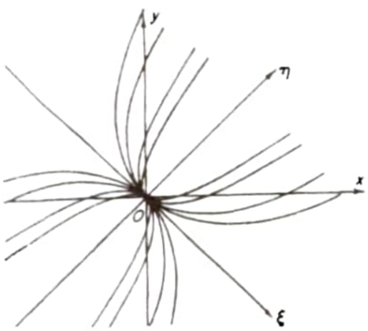

FIG. 10.

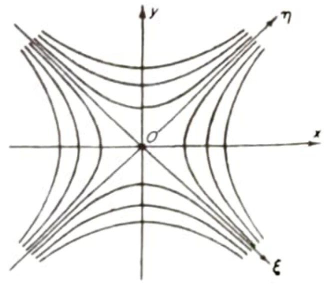

FIG. 11.

curves, after approaching the origin no closer than to some minimal distance, recede again from the origin. In this case we say that the point O is a *saddle point* because the integral curves are similar to the contours on a map representing the summit of a mountain pass (saddle).

All integral curves of equation (53) are given by the equation

$$b\eta = \xi(a + b \ln |\xi|),$$

where a and b are arbitrary constants. These are illustrated schematically in figure 12; all of them are tangent to the axis $O\eta$ at the origin.

If every integral curve entering some neighborhood of the singular point O passes through this point and has a definite direction there, i.e., has a definite tangent at the origin, as is illustrated in figures 10 and 12, then we say that the point O is a *node*.

Equation (54) is most easily integrated, if we change to polar coordinates ρ and ϕ, putting

$$\xi = \rho \cos \phi, \quad \eta = \rho \sin \phi.$$

Then this equation changes into the equation

$$\frac{d\rho}{d\phi} = k\rho, \quad \text{where} \quad k = \frac{\alpha}{\beta},$$

and hence,

$$\rho = Ce^{k\phi}. \tag{56}$$

If $k > 0$ then all the integral curves approach the point O, winding infinitely often around this point as $\phi \to -\infty$ (figure 13). If $k < 0$,

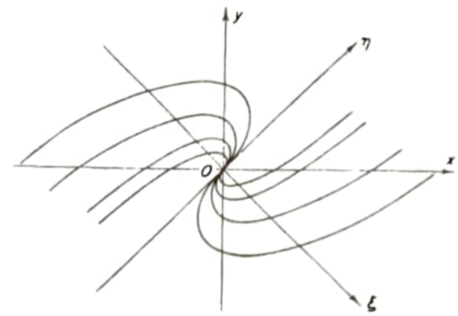

FIG. 12.

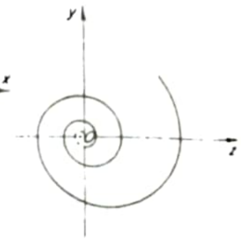

FIG. 13.

then this happens for $\phi \to +\infty$. In these cases, the point O is called a *focus*. If, however, $k = 0$, then the collection of integral curves of (56) consists of curves with center at the point O. Generally, if some neighborhood of the point O is completely filled by closed integral curves, surrounding the point O itself, then such a point is called a *center*.

A center may easily be transformed into a focus, if in the numerator and the denominator of the right side of equation (54) we add a term of arbitrarily high order; consequently, in this case the behavior of integral curves near a singular point is not given by terms of the first order.

Equation (55), corresponding to equation (45), is identical with the characteristic equation (19). Thus figures 10 and 12 schematically represent the behavior in the phase plane (x, y) of the curves

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

corresponding to the solutions of equation (6) for real λ_1 and λ_2 of the same sign; Figure 11 corresponds to real λ_1 and λ_2 of opposite signs, and figures 13 and 8 (the case of a center) correspond to complex λ_1 and λ_2. If the real parts of λ_1 and λ_2 are negative, then the point $(x(t), y(t))$ approaches 0 for $t \to +\infty$; in this case the point $x = 0$, $y = 0$ corresponds to stable equilibrium. If, however, the real part of either of the numbers

λ_1 and λ_2 is positive, then at the point $x = 0$, $y = 0$, there is no stable equilibrium.

§7. Qualitative Theory of Ordinary Differential Equations

An important part of the general theory of ordinary differential equations is the qualitative theory of differential equations. It arose at the end of the last century from the requirements of mechanics and astronomy.

In many practical problems, it is necessary to establish the character of the solution of a differential equation describing some physical process and to describe the properties of its solutions as the independent variable ranges over a finite or infinite interval. For example, in celestial mechanics, which studies the motion of heavenly bodies, it is important to have information about the behavior of the solutions of differential equations describing the motion of the planets or other heavenly bodies for unbounded periods of time.

As we said earlier, for only a few particularly simple equations can a general solution be expressed in terms of integrals of known functions. So there arose the problem of investigating the properties of the solutions of a differential equation from the equation itself. Since the solution of a differential equation is given in the form of a curve in a plane or in space, the problem consisted of investigating the properties of integral curves, their distribution and their behavior in the neighborhood of singular points. For example, do they lie in a bounded part of the plane or do they have branches tending to infinity, are some of them closed curves, and so forth? The investigation of such questions constitutes the qualitative theory of differential equations.

The founders of the qualitative theory of differential equations are the Russian mathematician A M. Ljapunov and the French mathematician H. Poincaré.

In the preceding section, we considered in detail one of the important questions of the qualitative theory, namely the distribution of integral curves in a neighborhood of a singular point. We turn now to some other basic questions in qualitative theory.

Stability. In the examples considered at the beginning of the chapter, the question of stability or instability of the equilibrium of a system was easily answered from physical considerations, without investigating the differential equations. Thus in example 3 it is obvious that if the pendulum, in its equilibrium position OA, is moved by some external force to a nearby position OA', i.e., if a small change is made in the initial conditions, then the subsequent motion of the pendulum cannot carry it very far from the equilibrium position, and this deviation will be smaller for smaller original deviations OA', i.e., in this case the equilibrium position will be stable.

For other more complicated cases, the question of stability of the equilibrium position is considerably more complicated and can be dealt with only by investigating the corresponding differential equations. The problem of the stability of equilibrium is closely connected with the question of the stability of motion. Fundamental results in this field were established by A. M. Ljapunov.

Let some physical process be described by the system of equations

$$\frac{dx}{dt} = f_1(x, y, t),$$
$$\frac{dy}{dt} = f_2(x, y, t). \tag{57}$$

For simplicity, we consider only a system of two differential equations, although our conclusions remain valid for a system with a larger number of equations. Each particular solution of the system (57), consisting of two functions $x(t)$ and $y(t)$, will sometimes be called a motion, following the usage of Ljapunov. We will assume that $f_1(x, y, t)$ and $f_2(x, y, t)$ have continuous partial derivatives. It has been shown that, in this case, the solution of the system of differential equations (57) is uniquely defined if at any instant of time $t = t_0$ the initial values $x(t_0) = x_0$ and $y(t_0) = y_0$ are given.

We will denote by $x(t, x_0, y_0)$ and $y(t, x_0, y_0)$ the solution of the system of equations (57) satisfying the initial conditions

$$x = x_0 \text{ and } y = y_0 \text{ for } t = t_0.$$

A solution $x(t, x_0, y_0)$, $y(t, x_0, y_0)$ is called *stable in the sense of Ljapunov* if for all $t > t_0$ the functions $x(t, x_0, y_0)$ and $y(t, x_0, y_0)$ have arbitrarily small changes for sufficiently small changes in the initial values x_0 and y_0.

More exactly, for a solution to be stable in the sense of Ljapunov, the differences

$$\begin{aligned} &| x(t, x_0 + \delta_1, y_0 + \delta_2) - x(t, x_0, y_0) |, \\ &| y(t, x_0 + \delta_1, y_0 + \delta_2) - y(t, x_0, y_0) | \end{aligned} \tag{58}$$

may be made less than any previously given number ϵ for all $t > t_0$, if the numbers δ_1 and δ_2 are taken sufficiently small in absolute value.

Every motion that is not stable in the sense of Ljapunov is called *unstable*.

In his investigation, the motion $x(t, x_0, y_0)$ and $y(t, x_0, y_0)$ was called by Ljapunov unperturbed, and the motion $x(t, x_0 + \delta_1, y_0 + d_2)$, $y(t, x_0 + \delta_1, y_0 + \delta_2)$ with nearby initial conditions was called perturbed. In this way stability in the sense of Ljapunov for an unperturbed motion means that for all $t > t_0$ the perturbed motion must differ only a little from the unperturbed.

The stability of equilibrium is a special case of stability of motion, corresponding to the case in which the unperturbed motion is

$$x(t, x_0, y_0) \equiv 0 \text{ and } y(t, x_0, y_0) \equiv 0.$$

Conversely, the question of the stability of any motion $x = \phi_1(t)$ and $y = \phi_2(t)$ of the system (57) may be reduced to the question of the stability of equilibrium for some system of differential equations. To this end we replace the unknown functions $x(t)$ and $y(t)$ in the system (57) by the new unknown functions

$$\xi = x - \phi_1(t) \text{ and } \eta = y - \phi_2(t). \tag{59}$$

In the system (57) transformed in this way, the motion $x = \phi_1(t)$ and $y = \phi_2(t)$ will correspond to the motion $\xi \equiv 0$ and $\eta \equiv 0$, i.e., the position of equilibrium. In what follows we will everywhere assume that the transformation (59) has been made, so that we may consider stability in the sense of Ljapunov only for the solution $x = 0$, $y = 0$.

The condition of stability in the sense of Ljapunov now means that, for δ_1 and δ_2 sufficiently small and $t > t_0$, the trajectory in the (x, y) plane of a perturbed motion does not pass outside of the square with sides of length 2 parallel to the coordinate axes and with center at the point $x = 0, y = 0$.

We will be interested in those cases in which, without knowing an integral of the system (57), we can nevertheless arrive at conclusions about the stability or instability of a motion. Stability is a very important practical question in the motion of projectiles, or of aircraft; and the stability of orbits is important in celestial mechanics, where the motion of planets and other heavenly bodies leads to this kind of investigation.

We assume that the functions $f_1(x, y, t)$ and $f_2(x, y, t)$ may be represented in the form

$$\begin{aligned} f_1(x, y, t) &= a_{11}x + a_{12}y + R_1(x, y, t), \\ f_2(x, y, t) &= a_{21}x + a_{22}y + R_2(x, y, t), \end{aligned} \tag{60}$$

where the a_{ij} are constants, and $R_1(x, y, t)$ and $R_2(x, y, t)$ are functions of x, y, and t such that

$$| R_1(x, y, t) | \leqslant M(x^2 + y^2) \text{ and } | R_2(x, y, t) | \leqslant M(x^2 + y^2), \tag{61}$$

where M is a positive constant.

If in the system (57) we substitute equations (60), neglecting $R_1(x, y, t)$ and $R_2(x, y, t)$, we get a system of differential equations with constant coefficients

$$\begin{aligned} \frac{dx}{dt} &= a_{11}x + a_{12}y, \\ \frac{dy}{dt} &= a_{21}x + a_{22}y, \end{aligned} \tag{62}$$

which is called the *system of first approximation to the nonlinear system* (57).

Before the time of Ljapunov, researchers confined themselves to investigating stability of the first approximation, believing that the results obtained would carry over to the question of stability for the basic nonlinear system (57). Ljapunov was the first to show that in the general case this conclusion is false. On the other hand, he gave a series of very wide conditions under which the question of stability for the nonlinear system is completely solved by the first approximation. One of these conditions is the following. If the real parts of both the roots of the equation

$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = 0$$

are negative and the functions $R_1(x, y, t)$ and $R_2(x, y, t)$ fulfill condition (61), then the solution $x(t) \equiv 0$, $y(t) \equiv 0$ is stable in the sense of Ljapunov. If the real part of either of the roots is positive, then the solution $x(t) \equiv 0$, $y(t) \equiv 0$ of an equation satisfying the conditions (61) is unstable. Ljapunov also gave a series of other sufficient conditions for stability and instability of a motion.*

If the right sides of equations (57) do not depend on t, then dividing the first equation of the system (57) by the second we get

$$\frac{dy}{dx} = \frac{f_1(x, y)}{f_2(x, y)}. \tag{63}$$

The origin will be a singular point for this equation. In the case of stability of equilibrium, this point may be a focus, a node, or a center, but cannot be a saddle point.

* A. M. Ljapunov, *The general problem of stability of motion.*

Thus the character of a singular point may be determined from the stability or instability of the equilibrium position.

The behavior of integral curves in the large. It is sometimes important to construct a schematized representation of the behavior of the integral curves "in the large"; that is, in the entire domain of the given system of differential equations, without attempting to preserve the scale. We will consider a space in which this system defines a field of directions as the phase space of some physical process. Then the general scheme of the integral curves, corresponding to the system of differential equations, will give us an idea of the character of all processes (motions) which can possibly occur in this system. In figures 10–13 we have constructed approximate schematized representations of the behavior of the integral curves in the neighborhood of an isolated singular point.

One of the most fundamental problems in the theory of differential equations is the problem of finding as simple a method as possible for constructing such a scheme for the behavior of the family of integral curves of a given system of differential equations in the entire domain of definition, in order to study the behavior of the integral curves of this system of differential equations "in the large." This problem remains almost untouched for spaces of dimension higher than 2. It is still very far from being solved for the single equation of the form

$$\frac{dy}{dx} = \frac{M(x, y)}{N(x, y)} \tag{64}$$

even when $M(x, y)$ and $N(x, y)$ are polynomials.

In what follows, we will assume that the functions $M(x, y)$ and $N(x, y)$ have continuous partial derivatives of the first order.

If all the points of a simply connected domain G, in which the right side of the differential equation (64) is defined, are ordinary points, then the family of integral curves may be represented schematically as a family of segments of parallel straight lines; since in this case one integral curve will pass through each point, and no two integral curves can intersect. For an equation (64) of more general form, which may have singular points, the structure of the integral curves may be much more complicated. The case in which equation (64) has an infinite set of singular points (i.e., points where the numerator and the denominator both vanish) may be excluded, at least when $M(x, y)$ and $N(x, y)$ are polynomials. Thus we restrict our consideration to those cases in which equation (64) has a finite number of isolated singular points. The behavior of the integral curves that are near to one of these singular points forms the essential

element in setting up a schematized representation of the behavior of all the integral curves of the equation.

A very typical element in such a scheme for the behavior of all the integral curves of equation (64) is formed by the so-called *limit cycles*. Let us consider the equation

$$\frac{d\rho}{d\phi} = \rho - 1, \tag{65}$$

where ρ and ϕ are polar coordinates in the (x, y) plane.

The collection of all integral curves of equation (65) is given by the formula

$$\rho = 1 + Ce^{\phi}, \tag{66}$$

where C is an arbitrary constant, different for different integral curves. In order that ρ be nonnegative, it is necessary that ϕ have values no larger than $-\ln|C|$, $C < 0$. The family of integral curves will consist of

1. the circle $\rho = 1$ $(C = 0)$;
2. the spirals issuing from the origin, which approach this circle from the inside as $\phi \to -\infty$ $(C < 0)$;
3. the spirals, which approach the circle $\rho = 1$ from the outside as $\phi \to -\infty$ $(C > 0)$ (figure 14).

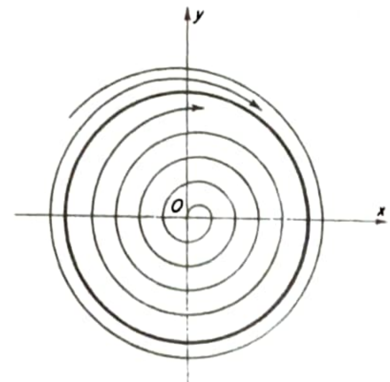

FIG. 14.

The circle $\rho = 1$ is called a limit cycle for equation (65). In general a closed integral curve l is called a *limit cycle*, if it can be enclosed in a disc all points of which are ordinary for equation (64) and which is entirely filled by nonclosed integral curves.

From equation (65) it can be seen that all points of the circle are ordinary. This means that a small piece of a limit cycle is not different from a small piece of any other integral curve.

Every closed integral curve in the (x, y) plane gives a periodic solution $[x(t), y(t)]$ of the system

$$\frac{dx}{dt} = N(x, y), \quad \frac{dy}{dt} = M(x, y), \tag{67}$$

describing the law of change of some physical system. Those integral curves in the phase plane that as $t \to +\infty$ approximate a limit cycle are motions that as $t \to +\infty$ approximate periodic motions.

Let us suppose that for every point (x_0, y_0) sufficiently close to a limit cycle l, we have the following situation: If (x_0, y_0) is taken as initial point (i.e., for $t = t_0$) for the solution of the system (67), then the corresponding integral curve traced out by the point $[x(t), y(t)]$, as $t \to +\infty$ approximates the limit cycle l in the (x, y) plane. (This means that the motion in question is approximately periodic.) In this case the corresponding limit cycle is called *stable*. Oscillations that act in this way with respect to a limit cycle correspond physically to self-oscillations. In some self-oscillatory systems, there may exist several stable oscillatory processes with different amplitudes, one or another of which will be established by the initial conditions. In the phase plane for such "self-oscillatory systems," there will exist corresponding limit cycles if the processes occuring in these systems are described by an equation of the form (67).

The problem of finding, even if only approximately, the limit cycles of a given differential equation has not yet been satisfactorily solved. The most widely used method for solving this problem is the one suggested by Poincaré of constructing "cycles without contact." It is based on the following theorem. We assume that on the (x, y) plane we can find two closed curves L_1 and L_2 (cycles) which have the following properties:

1. The curve L_2 lies in the region enclosed by L_1.

2. In the annulus Ω, between L_1 and L_2, there are no singular points of equation (64).

3. L_1 and L_2 have tangents everywhere, and the directions of these tangents are nowhere identical with the direction of the field of directions for the given equation (64).

4. For all points of L_1 and L_2 the cosine of the angle between the interior normals to the boundary of the domain Ω and the vector with components $[N(x, y), M(x, y)]$ never changes sign.

Then between L_1 and L_2, there is at least one limit cycle of equation (64).

Poincaré called the curves L_1 and L_2 *cycles without contact*.

The proof of this theorem is based on the following rather obvious fact. We assume that for decreasing t (or for increasing t) all the integral curves

$$x = x(t), \quad y = y(t)$$

of equation (64) (or, what amounts to the same thing, of equations (67), where t is a parameter), which intersect L_1 or L_2, enter the annulus Ω between L_1 and L_2. Then they must necessarily tend to some closed curve l lying between L_1 and L_2, since none of the integral curves lying in the annulus can leave it, and there are no singular points there.

But the problem of finding cycles without contact is also a complicated one and no general methods are known for solving it. For particular examples it has been possible to find cycles without contact, thereby proving the existence of limit cycles.

In radio technology it is important to find limit cycles (self-oscillatory processes) for equation (16) for the electron-tube generator. For equations of the type of (16), N. M. Krylov and N. N. Bogoljubov gave a method, about twenty years ago, for approximate computation of a certain limit cycle that exists for this equation. At about the same time the Soviet physicists L. I. Mandel'stam, N. D. Papaleksi, and A. A. Andronov gave a proof of the possibility of applying what is called the method of the small parameter, a method that to some extent had been used earlier in practice, though without any rigorous justification. Andronov was also the first to make systematic practical use, in the analysis of self-oscillatory systems, of the theoretical methods already developed by Ljapunov and Poincaré. In this manner he obtained a whole series of important results.

As was mentioned earlier, an important role is played in physics by "insensitive" systems (cf. §3). Andronov, together with L. S. Pontrjagin, set up a catalogue of the elements from which one could construct a complete chart of the behavior of the integral curves in the (x, y) plane for an insensitive differential equation of the form (64). It had been long known, for example, that a center near a singular point is easily destroyed by small changes in the equations (64). Thus in the construction of a chart of the behavior of the integral curves of equation (64), we cannot have a center, i.e., a family of closed integral curves surrounding a singular point, if the equation is "insensitive."

The question of the behavior of the integral curves in the large is still far from its final solution. We note that the analogous and probably simpler question of the form of real algebraic curves in the plane, i.e., curves defined by the equation

$$P(x, y) = 0,$$

where $P(x, y)$ is a polynomial of degree n, is also far from a complete solution. The form of these curves is completely known only for $n < 6$.

The solutions of the system (64) define motions in the plane. If we replace each point (x_0, y_0) in the plane by the corresponding point $[x(t, x_0, y_0), y(t, x_0, y_0)]$, where $x(t, x_0, y_0)$ and $y(t, x_0, y_0)$ are the solution of the system (64) with initial conditions $x = x_0$ and $y = y_0$ for $t = t_0$, we obtain a transformation of the points of the plane depending on the parameter t. Similar transformations depending on a parameter, together with the motions they generate, may be considered on a sphere, a torus, or other manifolds. The properties of these motions are studied in the theory of dynamical systems. In a neighborhood of every point these motions are the solutions of some system of differential equations. In the past decade the theory of dynamical systems has been developed on a broad basis in the works of V. V. Stepanov, A. Ja. Hinčin, N. N. Bogoljubov, N. M. Krylov, A. A. Markov, V. V. Nemyckiĭ and others, and also in the works of G. D. Birkhoff and other mathematicians.

In this chapter we have given a brief outline of the present state of the theory of ordinary differential equations and have attempted to describe the problems that are considered in this theory. Our study in no sense pretends to be complete. We have had to omit consideration of many branches of the theory that arise in the study of more special problems or that require broader mathematical knowledge than the reader of this book is assumed to possess. For example, we have nowhere touched upon the general and important area in which the theory of differential equations with complex arguments is considered. We have had no opportunity to examine the theory of boundary-value problems and in particular, of eigenfunctions, which is of great importance in the applications.

We have also been able to pay very little attention to approximative methods for the numerical or analytical solution of differential equations. For these questions, we recommend that the reader consult the specialized literature.

Suggested Reading

R. P. Agnew, *Differential equations*, 2nd ed., McGraw-Hill, New York, 1960.

E. A. Coddington, *An introduction to ordinary differential equations*, Prentice-Hall, Englewood Cliffs, N. J., 1961.

E. A. Coddington and N. Levinson, *Theory of ordinary differential equations*, McGraw-Hill, New York, 1955.

W. Hurewicz, *Lectures on ordinary differential equations*, Technology Press and Wiley, New York, 1958.

S. Lefschetz, *Differential equations: geometric theory*, Interscience, New York, 1957.