Graduate Texts in Mathematics

Readings in Mathematics

Wolfgang Walter

Ordinary Differential Equations

Springer

Graduate Texts in Mathematics 182

Readings in Mathematics

Springer Science+Business Media, LLC

Graduate Texts in Mathematics

Readings in Mathematics

Ebbinghaus/Hermes/Hirzebruch/Koecher/Mainzer/Neukirch/Prestel/Remmert: *Numbers*
Fulton/Harris: *Representation Theory: A First Course*
Remmert: *Theory of Complex Functions*

Undergraduate Texts in Mathematics

Readings in Mathematics

Anglin: *Mathematics: A Concise History and Philosophy*
Anglin/Lambek: *The Heritage of Thales*
Bressoud: *Second Year Calculus*
Hairer/Wanner: *Analysis by Its History*
Hämmerlin/Hoffmann: *Numerical Mathematics*
Isaac: *The Pleasures of Probability*
Samuel: *Projective Geometry*
Stillwell: *Numbers and Geometry*
Toth: *Glimpses of Algebra and Geometry*

Wolfgang Walter

Ordinary Differential Equations

Translated by Russell Thompson

Springer

Wolfgang Walter
Mathematisches Institut I
Universität Karlsruhe
D-76128 Karlsruhe
Germany

Russell Thompson
Utah State University
College of Science
Department of Mathematics and Statistics
Logan, UT 84322-3900

Mathematics Subject Classification (1991): 34-01

Library of Congress Cataloging-in-Publication Data
Walter, Wolfgang, 1927–
Ordinary differential equations / Wolfgang Walter.
p. cm. — (Graduate texts in mathematics ; 182. Readings in mathematics)
Includes bibliographical references and index.
ISBN 978-1-4612-6834-5 ISBN 978-1-4612-0601-9 (eBook)
DOI 10.1007/978-1-4612-0601-9
1. Differential equations. I. Title. II. Series: Graduate texts in mathematics ; 182. III. Series: Graduate texts in mathematics. Readings in mathematics.
QA372.W224 1998
515′.352—dc21 98-4754

Printed on acid-free paper.

Production managed by Allan Abrams; manufacturing supervised by Jacqui Ashri.
Photocomposed copy prepared from the translator's TeX files.

9 8 7 6 5 4 3 2 1

ISBN 978-1-4612-6834-5

Preface

The author's book on *Gewöhnliche Differentialgleichungen* (Ordinary Differential Equations) was published in 1972. The present book is based on a translation of the latest, 6th, edition, which appeared in 1996, but it also treats some important subjects that are not found there. The German book is widely used as a textbook for a first course in ordinary differential equations. This is a rigorous course, and it contains some material that is more difficult than that usually found in a first course textbook; such as, for example, Peano's existence theorem. It is addressed to students of mathematics, physics, and computer science and is usually taken in the third semester. Let me remark here that in the German system the student learns calculus of one variable at the gymnasium[1] and begins at the university with a two-semester course on real analysis which is usually followed by ordinary differential equations.

Prerequisites. In order to understand the main text, it suffices that the reader have a sound knowledge of calculus and be familiar with basic notions from linear algebra. For complex differential equations, some facts about holomorphic functions and their integrals are required. These are summarized at the beginning of § 8 and more fully described and partly proved in part C of the Appendix. Functional analysis is developed in the text when needed. In several places there are sections denoted as Supplements, where more special subjects are treated or the theory is extended. More advanced tools such as Lebesgue's theory of integration or Schauder's fixed point theorem are occasionally used in those sections. The supplements and also § 13 can be omitted in a first reading.

Outline of contents. The book treats significantly more topics than can be covered in a one-semester course. It also contains material that is seldom found in textbooks and—what is perhaps more important—it uses new proofs for basic theorems. This aspect of the book calls for a closer look at contents and methods with emphasis on those places where we depart from the mainstream.

The first chapter treats classical cases of first order equations that can be solved explicitly. By means of a number of examples the student encounters the essential features of the initial value problem such as uniqueness and nonuniqueness, maximal solutions in the case of nonuniqueness, and continuous dependence on initial values in the small, but not in the large; see 1.VI–VIII. The

[1] In the German school system, the gymnasium is an academic high school that prepares students for study at the university.

phase plane and phase portraits are explained in 3.VI–VIII.

The theory proper starts with Chapter II. In this and the following chapter the initial value problem is treated first for one equation and then for systems of equations. The repetition caused by this separation of cases is minimal since all proofs carry over, while the student has the benefit that the reasoning is not burdened by technicalities about vector functions. The complex case, where the solutions are holomorphic functions, is treated in § 8; the proofs follow the pattern set in § 6 for the real case. The theory of differential inequalities in § 9 is one-dimensional by its very nature. An extension to n dimensions leads to new phenomena that are treated in Supplement I of § 10.

Chapter IV is devoted to linear systems and linear differential equations of higher order. In a Supplement to § 18 the Floquet theory for systems with periodic coefficients is presented.

Linear systems in the complex domain is the topic of Chapter V. The main properties of systems with isolated singularities are developed in a novel way (see below). Equations of mathematical physics are discussed in § 25.

The main subject of Chapter VI is the Sturm–Liouville theory of boundary value and eigenvalue problems. Nonlinear boundary value problems and corresponding existence, uniqueness, and comparison theorems are also treated. In § 28 the eigenvalue theory for compact self-adjoint operators in Hilbert space is developed and applied to the Sturm–Liouville eigenvalue problem.

The last chapter deals with stability and asymptotic behavior of solutions. The linearization theorem of Grobman–Hartman is given without proof (the author is still looking for a really good proof). The method of Lyapunov is developed and applied in § 30.

An appendix consisting of four parts A (topology), B (real analysis), C (complex analysis), and D (functional analysis) contains notions and theorems that are used in the text or can lead to a deeper understanding of the subject. The fixed point theorems of Brouwer and Schauder are proved in B.V and D.XII.

In closing this overview, we point out that applications, mostly from mechanics and mathematical biology, are found in many places. Exercises, which range from routine to demanding, are dispersed throughout the text, some with an outline of the solution. Solutions of selected exercises are found at the end of the book.

Special Features. Two general themes exercise a profound influence throughout the book: functional analysis and differential inequalities.

Functional Analysis. The *contraction principle*, that is, the fixed point theorem for contractive mappings in a Banach space, is at the center. This theorem has all necessary properties to make it a fundamental principle of analysis: It is elementary, widely applicable, and far-reaching.[2] Its flexibility in connection with our subject comes to light when appropriate weighted maximum norms

---

[2] A remarkable theorem of Bessaga (1959) sheds light on the versatility of the contraction principle. Consider a map $T : S \to S$, where S is an arbitrary set, and assume that T has a unique fixed point which is also the only fixed point of $T^2, T^3, \ldots$. Then there is a metric on S that makes S a complete metric space and T a contraction. One can even find metrics for which the Lipschitz constant of T is arbitrarily small.

are used. A first example is found in the dissertation of Morgenstern (1952); references to later authors in the literature are historically unjustified. In linear complex systems, the weighted maximum norm in 21.II leads to global existence without using analytic continuation and the monodromy theorem. Moreover, this proof gives the growth properties of solutions that are needed in the treatment of singular points. The theorems on continuous dependence on initial values and parameters and on holomorphy with regard to complex parameters follow directly from the contraction principle, a fact which is still little known. Differentiability with respect to real parameters requires Ostrowski's theorem on approximate iteration 13.IV.

In the treatment of linear systems with weakly singular points, the crucial convergence proofs are also reduced to the contraction principle in a suitable Banach space.[3] For holomorphic solutions, i.e., power series expansions, this method was discovered by Harris, Sibuya, and Weinberg (1969). The logarithmic case can also be treated along these lines. This approach leads also to theorems of Lettenmeyer and others, which are beyond the scope of this book; cf. the original work cited above.

A theorem in Appendix D.VII, which is partly due to Holmes (1968), establishes a relation between the norm of a linear operator and its spectral radius. As explained in Section D.IX, this result gives a better insight into the role of weighted maximum norms.

Differential Inequalities. The author, who also wrote the first monograph on differential inequalities (1964, 1970), has encountered many instances where authors are unaware of basic theorems on differential inequalities that would have made their reasoning much simpler and stronger. The distinction between weak and strong inequalities is a matter of fundamental importance. In partial differential equations this is common knowledge: weak maximum or comparison principles versus strong principles of this type. Not so in ordinary differential equations. Theorem 9.IX is a strong comparison principle that prescribes precisely the occurrence of strict inequalities, while most (all?) textbooks are content with the weak "less than or equal" statement. This principle is essential for our treatment of the Sturm–Liouville theory via Prüfer transformation. Its usefulness in nonlinear Sturm theory can be seen from a recent paper, Walter (1997).

Supplement I in § 10 brings the two basic theorems on systems of differential inequalities, (i) the comparison theorem for quasimonotone systems, and (ii) Max Müller's theorem for the general case. Both were found in the mid twenties. *Quasimonotonicity* is a necessary and sufficient condition for extending the classical theory (including maximal and minimal solutions) from one equation to systems of equations. More recently, both theorems (i) and (ii) have been applied to population dynamics, but it is not generally known that results on

---

[3]The Banach space H_δ of 24.I, which is indeed a Banach algebra, can be used for a short and elegant proof of two fundamental theorems for functions of several complex variables, the preparation theorem and the division theorem of Weierstrass. This proof has been propagated by Grauert and Remmert since the sixties and can be found, e.g., in their book *Coherent Analytic Sheaves* (Grundlehren 265, Springer 1984); cf. Walter (1992) for other applications.

invariant rectangles are special cases of Müller's theorem. Theorem 10.XII is the strong version of (i); it contains M. Hirsch's theorem on strongly monotone flows, cf. Hirsch (1985) and Walter (1997).

A Supplement to § 26 describes a new approach to minimum principles for boundary value problems of Sturmian type that applies also to nonlinear differential operators; cf. Walter (1995). The strong minimum principle is generalized in 26.XIX, so that it includes now the first eigenvalue case.

In Supplement II of § 26 on nonlinear boundary value problems the method of upper and lower solutions for existence and Serrin's sweeping principle for uniqueness are presented.

Miscellaneous Topics. *Differential equations in the sense of Carathéodory.* The initial value problem is treated in Supplement II of § 10 and a Sturm–Liouville theory under Carathéodory assumptions in 26.XXIV and 27.XXI. As a rule, the earlier proofs for the classical case carry over. This applies in particular to the strong comparison theorem 10.XV and the strong minimum principle in 26.XXV.

Radial solutions of elliptic equations. This subject plays an active role in recent research on nonlinear elliptic problems. The radial Δ-operator is an operator of Sturm–Liouville type with a singularity at 0. The corresponding initial value problem is treated in a supplement of § 6, and the eigenvalue problem and nonlinear boundary value problems for the unit ball in $\mathbb{R}^n$ (for radial solutions) in a Supplement to § 27.

Separatrices is the theme of a Supplement in § 9. Differential inequalites are essential for proving existence and uniqueness.

Special Applications. We mention the generalized logistic equation in a supplement to § 2, general predator–prey models in 3.VII, delay-differential equations in 7.XIV-XV, invariant sets in 10.XVI and the rubber band as a model for nonlinear oscillations in a nonsymmetric mechanical system in 11.X.

Exact Numerics. We give examples in which a combination of a numerical procedure and a sup-superfunction technique allows a mathematically exact computation of special values. The numerical part is based on an algorithm, developed by Rudolf Lohner (1987, 1988), that computes exact enclosures for the solutions of an initial value problem. In blow-up problems one obtains rather sharp enclosures for the location of the asymptote of the solutions; cf. 9.V. A different kind of sub- and supersolutions is used to compute a separatrix; in general, a separatrix is an unstable solution.

Acknowledgments. It is a pleasure to thank all those who have contributed to the making of this volume. The translator, Professor Russell Thompson, worked with expertise and patience in the face of changes and additions during the translation and furnished beautiful figures. He also suggested an improved division into chapters. Irene Redheffer acted as a mediator between author and translator with exceptional care and insight and translated the Solutions section. Her help and advice and that of Professor Ray Redheffer were indispensable. My sincere thanks go to all of them and also to other helping hands and minds.

Karlsruhe, August 1997 Wolfgang Walter

Table of Contents

Note to the Reader

In references to another paragraph, the number of the paragraph is given before the number of the formula, theorem, lemma,.... For example, formula (7) in § 15 is denoted as (15.7), and theorem 15.III or corollary 15.III refers to the theorem or corollary in section III of § 15. But when citing within § 15, we write simply formula (7), Theorem III, and Corollary III. A reference to B.V refers to Section V in Part B of the Appendix.

When the name of an author is followed by the year of publication, as in Perron (1926), the source is found in the bibliography at the end of the book. My two books on analysis are cited as Walter 1 and Walter 2. A compilation of general notions and a list of symbols are found at the end of the book.

The German word *Ansatz* is used repeatedly; a footnote in Part II of the introduction gives an explanation.

Introduction

A differential equation is an equation containing independent variables, functions, and derivatives of functions. The equation

$$y' + 2xy = 0 \tag{1}$$

is a differential equation. Here x is the independent variable and y is the unknown function. A *solution* is a function $y = \phi(x)$ that satisfies (1) identically in x, that is, $\phi'(x) + 2x \cdot \phi(x) \equiv 0$. It is easy to check that the function $y = \mathrm{e}^{-x^2}$ is a solution of (1):

$$\frac{d}{dx}(\mathrm{e}^{-x^2}) + 2x\mathrm{e}^{-x^2} \equiv 0 \quad \text{for} \quad -\infty < x < \infty.$$

We will see later that the collection of all solutions of (1) can be written in the form $y = C \cdot \mathrm{e}^{-x^2}$, where C runs through the set of real numbers.

Equation (1) is a *differential equation of first order*. The general differential equation of first order has the form

$$F(x, y, y') = 0. \tag{2}$$

A function $y = y(x)$ is called a *solution* of (2) in an interval J if $y(x)$ is differentiable in J and

$$F(x, y(x), y'(x)) \equiv 0 \quad \text{holds for all} \quad x \in J.$$

If a differential equation contains higher order derivatives, say up to nth order, then the equation is called an *nth order differential equation*. Such an equation can always be written in the general form

$$F(x, y, y', \ldots, y^{(n)}) = 0. \tag{3}$$

Here a *solution* is defined to be an n-times differentiable function such that equation (3) is satisfied identically when $y(x)$ and its derivatives are substituted into F. A differential equation of nth order is called *explicit* if it has the form

$$y^{(n)} = f(x, y, y', \ldots, y^{(n-1)}); \tag{4}$$

otherwise it is called *implicit*. For a first order ordinary differential equation the explicit form is

$$y' = f(x, y). \tag{5}$$

The above comments apply to *ordinary differential equations*, that is, to differential equations for functions $y(x)$ of a *single* independent variable x. If several independent variables and hence also partial derivatives are present, then the equation is called a *partial differential equation*. For example,

$$u_x + u_y = x + y$$

is a partial differential equation of first order for an unknown function $u(x, y)$. The function $u(x, y) = xy$ is a particular solution to this equation. An important example of a second order partial differential equation is the *potential equation* in three-space

$$\Delta u \equiv u_{xx} + u_{yy} + u_{zz} = 0,$$

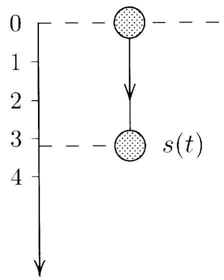
where $u = u(x, y, z)$.

In this book we will be concerned only with ordinary differential equations. The primary emphasis will be on differential equations in the real domain where the independent variable x is a real variable and $y(x)$ is a real function. However, the fundamental facts about differential equations in the complex domain will also be treated.

The expression *integral of a differential equation* is another term used for a *solution*, and the terms *solution curve* and *integral curve* are used to emphasize the geometric interpretation of a solution as a curve. A family of functions $y(x; C_1, \ldots, C_n)$, depending on x and n parameters $C_1, \ldots, C_n$ (which vary in a point set $M \subset \mathbb{R}^n$), is called a *complete integral* or a *general solution* of the nth order differential equation (4) if it satisfies the following two requirements: first, each function $y(x; C_1, \ldots, C_n)$ is a solution to the differential equation (4) for an arbitrary choice of the parameters $(C_1, \ldots, C_n) \in M$, and second, all solutions can be obtained in this manner. The notion of a general solution does not play a major role in the theory of differential equations. It is used here in connection with simple examples, where it is actually possible to give all solutions explicitly in a form depending on n parameters.

Differential equations play a cardinal role in the natural sciences and technology, especially in physics, for the simple reason that many physical laws take the form of a differential equation. Differential equations also appear in other scientific domains where mathematical models and theories are used. The three examples that follow are intended to give a first impression of the type of problems that arise. They all deal with the motion of a body in a gravitational field.

I. Free Fall. When a body at rest is suddenly released, it falls downward under the influence of gravity. This motion can be described mathematically by a function $s = s(t)$ which gives the distance that the body (or more exactly, its center of mass) has traveled up to time t. Other quantities of interest that can be derived from s include the instantaneous velocity $v(t) = \dfrac{d}{dt}s(t) = \dot{s}(t)$ and the acceleration $a(t) = \dfrac{d}{dt}v(t) = \ddot{s}(t)$. (When describing processes in which



the independent variable represents time, it is customary to denote the independent variable by t instead of x, and derivatives by dots instead of primes.) We learn in elementary mechanics that the acceleration of such bodies may be assumed to be constant, in fact, equal to the acceleration g due to gravity at the earth's surface. Thus the distance-time function $s(t)$ satisfies the second order differential equation

$$\ddot{s} = g. \tag{6}$$

It is easy to find all of the solutions here. Indeed, it follows from integrating the equation $\dot{v}(t) = g$ that $v(t) = gt + C_1$, and likewise from $\dot{s}(t) = gt + C_1$ that

$$s(t) = \frac{1}{2}gt^2 + C_1 t + C_2 \qquad (C_1,\ C_2 \text{ constant}).$$

We have thus found the complete integral of the differential equation (6).

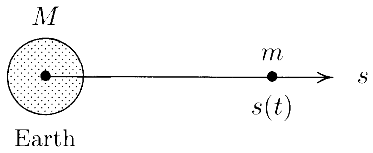
To go from this family of $\ddot{s} = g$ to the solution that corresponds to a particular physical process requires some additional information, the so-called *initial conditions*. Let us assume, for instance, that in the example above the body is at rest and is then released at time $t = 0$. Corresponding initial conditions are given by $s(0) = 0$ and $\dot{s}(0) = v(0) = 0$. From the first of these conditions it follows that $C_2 = 0$, from the second that $C_1 = 0$, and in this manner one obtains the solution

$$s(t) = \frac{1}{2}gt^2.$$

Other initial conditions lead in a like manner to other solutions.

II. Free Fall from a Large Distance. Now suppose that the body is at a large distance from the earth. The assumption of constant gravitational acceleration made in I is valid only near the surface of the earth. According to Newton's law of gravitation, two bodies a distance s apart with masses M (earth) and m (test body) attract each other with a force equal to $K = \gamma \frac{Mm}{s^2}$, where γ is the gravitational constant. By Newton's second law the acceleration now satisfies the equation

$$\ddot{s} = -\gamma M \cdot \frac{1}{s^2}. \tag{7}$$



The minus sign on the right-hand side indicates that the direction of the force is opposite to the positive s-direction. This differential equation of second order is significantly more difficult to integrate than equation (6). Nonetheless, the solutions can be given explicitly; we will return to this later in §11.XII. Suppose that at time $t = 0$ a test body is located a distance R from the earth's center and released at rest. Then one has for initial conditions $s(0) = R$, $\dot{s}(0) = 0$.

A simple and sometimes successful method of finding solutions to a differential equation is to look for a likely "ansatz"[4] (possibly containing parameters) and to investigate whether it leads to a solution. We will try this approach in the case of equation (7) using the ansatz

$$s(t) = a \cdot t^b.$$

When this function is substituted into equation (7), the result is

$$ab(b-1)t^{b-2} = -\gamma M a^{-2} t^{-2b}.$$

Equating exponents and coefficients leads to $b - 2 = -2b$, that is, $b = \frac{2}{3}$, and $a \cdot \frac{2}{3}(-\frac{1}{3}) = -\gamma M a^{-2}$, from which follows $a = (9\gamma M/2)^{1/3}$. Thus $s(t) = a \cdot t^{2/3}$ is a solution. It is easy to check that any function of the form

$$s(t) = a(c \pm t)^{2/3} \quad \text{with} \quad a = (9\gamma M/2)^{1/3}, \quad c \text{ arbitrary}, \tag{8}$$

is a solution to the differential equation (7) as long as $c \pm t > 0$. Note that none of the solutions from this collection satisfies the initial conditions mentioned above. The solution

$$s(t) = a\left(\frac{R\sqrt{2R}}{\sqrt{9\gamma M}} - t\right)^{2/3},$$

for example, satisfies $s(0) = R$, but $v(0) = \dot{s}(0) = -\sqrt{2\gamma M/R}$. This describes an object falling to earth from the position $s = R$ with initial speed at time $t = 0$ equal to $\sqrt{2\gamma M/R}$.

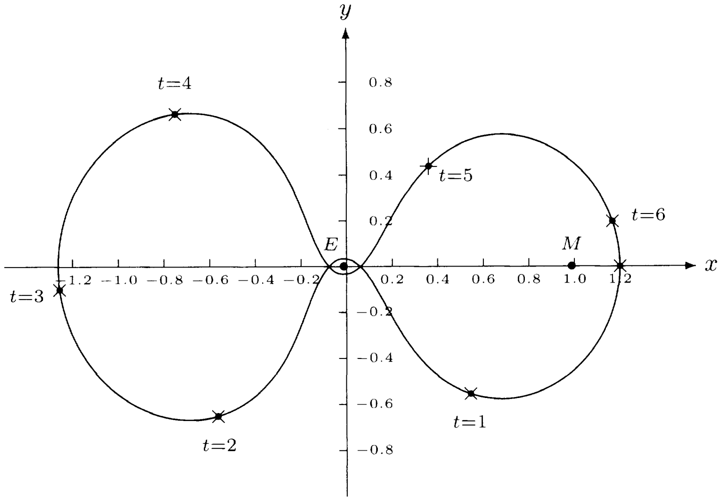
One of the solutions of (8) with $c = 0$ is

$$\tilde{s}(t) = at^{2/3}. \tag{9}$$

An object on this trajectory does not return to earth, since $\tilde{s}(t) \to \infty$ as $t \to \infty$; however, the velocity $\tilde{v}(t) = \frac{2}{3}at^{-1/3}$ tends to 0 as $t \to \infty$. Since $\tilde{v}(t)^2 \cdot \tilde{s}(t) =$

---

[4] The word *ansatz* is a German word that has become part of modern mathematical language; it has no exact English counterpart. An ansatz is an "educated guess" at the probable form of a solution. The plural of *ansatz* is *ansätze*.



$\frac{4}{9}a^3 = 2\gamma M$, the velocity as a function of distance R from the earth's center can be expressed in the form

$$v_R = \sqrt{2\gamma M/R}.$$

Substituting $R = 6.370 \cdot 10^8$ cm and $M = 5.97 \cdot 10^{27}$g for the radius and mass of the earth and taking $\gamma = 6.685 \cdot 10^{-8}$ dyn $\cdot$ cm$^2 \cdot$ g^{-2}, one obtains

$$v_R = 1.12 \cdot 10^6 \,\text{cm/sec} = 11.2\,\text{km/sec}.$$

This is the well-known "escape velocity," the minimum velocity that a projectile fired from the surface of the earth must have in order to escape the effect of the earth's gravitational pull and never return. Compare this result with the exercise at the end of this introduction.

III. Motion in the Gravitational Field of Two Bodies (Satellite Orbits). The following equations (10) describe the motion of a small body (a satellite) in the force field of two larger bodies (earth and moon). It is assumed here that the motion of the three bodies takes place in a fixed plane and that the two larger bodies rotate with the same constant angular velocity about their common center of mass and maintain a constant distance to it. In particular, the effect of the small body on the motion of the two larger bodies will be ignored (this is the meaning of the adjectives 'small' and 'large'). In a corotating coordinate system with the center of mass at the origin, the two larger bodies appear to be at rest. The path of the small body can be described by a function pair $(x(t), y(t))$ that satisfies the following system of two second order differential equations:

$$\begin{aligned} \ddot{x} &= x + 2\dot{y} - \mu' \frac{x+\mu}{[(x+\mu)^2 + y^2]^{3/2}} - \mu \frac{x-\mu'}{[(x-\mu')^2 + y^2]^{3/2}}, \\ \ddot{y} &= x - 2\dot{x} - \mu' \frac{y}{[(x+\mu)^2 + y^2]^{3/2}} - \mu \frac{y}{[(x-\mu')^2 + y^2]^{3/2}}. \end{aligned} \tag{10}$$

Here the two larger bodies are assumed to lie on the x-axis, and the parameter μ, respectively μ', is the ratio of the mass of the body lying on the positive, respectively negative, x-axis to the combined mass of both bodies. Further, the unit of length is chosen such that the distance between the two bodies is equal to 1, and the unit of time such that the angular velocity of the rotation is also equal to 1 (i.e., a complete revolution lasts 2π time units). A closed orbit is reproduced in the figure. Here $\mu \approx 0.01213$, which corresponds to the mass ratio of the earth–moon system. The initial conditions are

$$\begin{aligned} x(0) &= 1.2, & y(0) &= 0, \\ \dot{x}(0) &= 0, & \dot{y}(0) &\approx -1.04936. \end{aligned}$$

The period T (duration of one complete revolution) is approximately equal to 6.19217.

These examples suggest a variety of problems. First we made use of elementary methods of solution and discovered in the process that for some differential equations all solutions can be given in closed form (Examples I, II). For differential equations in general, just as in the problem of finding the antiderivative of an elementary function in integral calculus, the adage holds: Explicit solutions are the exception! The theory of differential equations proper has as its goal a general theory of existence, uniqueness, and other related subjects (for example, continuous dependence of solutions on various kinds of data) together with qualitative statements about the behavior of solutions in the large such as boundedness, oscillation properties, stability, and asymptotic behavior. Theorems about inequalities are also important, as the exercise at the end of this introduction illustrates.

Several important topics can only be touched briefly in an introductory work like this one. These include, for instance, the investigation of *periodic solutions* to nonlinear differential equations. Periodic solutions have important applications in mechanics (oscillations) and celestial mechanics (closed orbits). However, their mathematical theory is often difficult. Some results in this direction will be presented in 3.VI–VII and 11.X–XI. For the earth–moon–satellite problem described in III, a special case of the "restricted three-body problem," it was suggested some time ago that a spaceship on a periodic orbit could be used as a kind of "bus line" between the earth and the moon. The ensuing investigation led to the discovery of a new class of periodic orbits; see Arenstorf (1963).

Also, the problem of solving differential equations numerically will not be treated here. We note that difficult numerical problems arise in connection with space flight (determining the trajectories of spacecraft). There are efficient numerical algorithms available today that allow the determination and correction of such trajectories with sufficient accuracy and a tolerable amount of computational effort; see, for instance, the work of Bulirsch and Stoer (1966), from which the algorithm that produced the above figure is taken.

IV. Exercise. Prove the assertion at the end of Example II. More precisely, show: If $s(t)$ is a positive solution of the differential equation (7) in the interval $0 < t_0 \le t < t_1$ (with $t_1 = \infty$ allowed), $\bar{s}(t)$ the solution given by (9), and $s(t_0) = \bar{s}(t_0)$, $0 < \dot{s}(t_0) = v(t_0) < \bar{v}(t_0)$, then $s(t) < \bar{s}(t)$ and $v(t) < \bar{v}(t)$ for $t_0 < t < t_1$. Further, $s(t)$ is bounded above, and $v(t)$ has one zero (thus, $s(t)$ describes a return trajectory).

Hint. Derive a differential equation for the difference $d(t) = \bar{s} - s$ and conclude from it that $\dot{d}$ is monotone increasing as long as d is positive. Note that $\ddot{s} < 0$ and $\bar{v}(t) \to 0$ as $t \to \infty$.

Chapter I
First Order Equations: Some Integrable Cases

§ 1. Explicit First Order Equations

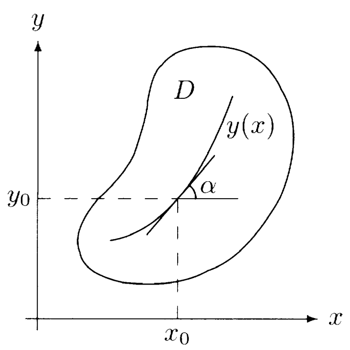
We consider the explicit first order differential equation

$$y' = f(x, y). \tag{1}$$

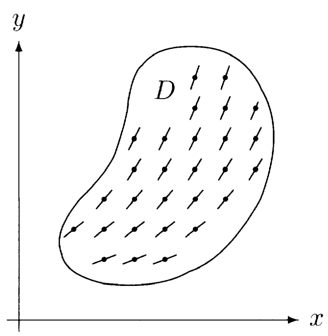
The right-hand side $f(x, y)$ of the equation is assumed to be defined as a real-valued function on a set D in the xy-plane.

I. Solution. Line Element. Direction Field. Let J be an interval. (In general, J can be open, closed, half-open, a half-line, or the whole real line; when special restrictions are necessary, they will be stated explicitly.) A function $y(x) : J \to \mathbb{R}$ is called a *solution* to the differential equation (1) (in J) if y is differentiable in J, the graph of y is a subset of D, and (1) holds, i.e., if

$$(x, y(x)) \in D \quad \text{and} \quad y'(x) = f(x, y(x)) \quad \text{for all} \quad x \in J.$$



Slope and line element



Direction field

The differential equation (1) has a simple geometric interpretation. If $y(x)$ is an integral curve of (1) that passes through a point (x_0, y_0) (i.e., $y(x_0) = y_0$), then the differential equation specifies the slope of the curve at that point: $y'(x_0) = f(x_0, y_0)$. This leads naturally to the notions of line element and direction field, which we will now define. We interpret a numerical triple of the form (x, y, p) geometrically in the following way: (x, y) gives a point in the plane, and the third component p gives the slope of a line through the point (x, y) (α, with $\tan\alpha = p$, is the angle of inclination of the line; see the figure). Such a triple (or its geometric equivalent) is called a *line element*. The collection of all line elements of the form $(x, y, f(x, y))$, i.e., those with $p = f(x, y)$, is called a *direction field*.

The connection between direction fields and the differential equation (1) can be expressed in geometric terms as follows: A solution $y(x)$ of a differential equation "fits" its direction field, i.e., the slope at each point on the solution curve agrees with the slope of the line element at that point. To put it another way, if $y(x)$ is a solution in J, then the set of line elements $(x, y(x), y'(x))$, with $x \in J$, is contained in the set of all line elements $(x, y, f(x, y))$, $(x, y) \in D$.

The strategy of sketching a few of the line elements in the direction field and then trying to draw curves that fit these line elements can be used to get a rough idea of the nature of the solutions to a differential equation. This procedure suggests quite naturally the view that for each point (ξ, η) in D there is exactly one solution curve $y(x)$ passing through that point. A precise formulation of this idea leads to the notion of

II. The Initial Value Problem. Let a function $f(x, y)$, defined on a set D in the (x, y)-plane, and a fixed point $(\xi, \eta) \in D$ be given. A function $y(x)$ is sought that is differentiable in an interval J with $\xi \in J$ such that

$$y'(x) = f(x, y(x)) \quad \text{in} \quad J, \tag{2}$$

$$y(\xi) = \eta. \tag{3}$$

Equation (3) is called the *initial condition*. Naturally, in (2) it is assumed that $\operatorname{graph} y = \{(x, y(x)) : x \in J\} \subset D$ (otherwise the right-hand side of (2) would not even be defined).

III. Remarks. (a) *Differential Equations and Families of Curves.* The geometric line of reasoning outlined above can be turned around. Given a family of curves that completely covers a set D in the plane (precise analytic formulation: a set M of differentiable functions whose graphs are pairwise disjoint and have D as their union), there is a differential equation that has these curves as its solutions. The right-hand side of this differential equation is determined as follows: For every $(x_0, y_0) \in D$, find the unique function ϕ belonging to M with $\phi(x_0) = y_0$ and set $f(x_0, y_0) = \phi'(x_0)$.

This relationship does not give us very much from the mathematical point of view. However, it does give an idea of some of the possibilities for the behavior of differential equations, and in addition, it can be used in the construction of examples.

We will now briefly discuss some mathematical shorthand that is frequently used.

(b) Sometimes (particularly in examples) when a differential equation is solved, the function found as the solution is only a solution to the equation on a subinterval of its domain of definition. When this happens, the expression "$\phi(x)$ is a solution to the differential equation in the interval J" means that ϕ is defined at least on J and that the restriction $\phi|_J$ is a solution in the sense of the definition in I.

(c) If $\phi : J \to \mathbb{R}$ is a solution of the differential equation (1) and J' is a subinterval of J, then, in a trivial way, the restriction $\psi = \phi|_{J'}$ is also a solution of (1). This is not regarded as a new solution. For instance, the statement "the differential equation has exactly one solution existing in the interval J" means: There exists a solution with J as its domain of definition, and every other solution is a restriction of this solution.

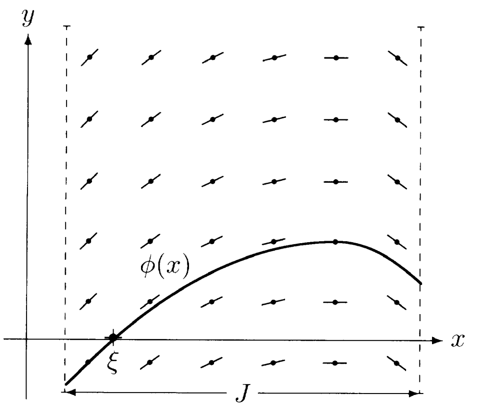
Before giving a detailed investigation of initial value problems, we will study some simple examples.

IV. $\boxed{y' = f(x)}$

Suppose the function $f(x)$ is continuous in an interval J. Then the set D is a strip $J \times \mathbb{R}$. The direction field is independent of y. This leads naturally to the conjecture that all of the solutions can be obtained by translating any one particular solution in the direction indicated by the y-axis. The analysis confirms that this guess is true. If $\xi \in J$ is fixed, then by the fundamental theorem of calculus, the function

$$\phi(x) := \int_\xi^x f(t)\, dt$$

is a solution of the differential equation satisfying the initial condition $y(\xi) = 0$;



Direction field in the case where the right side depends only on x

and the general solution can be written in the form

$$y = y(x; C) = \phi(x) + C,$$

where C is an arbitrary constant. It follows in particular that the initial value problem (2), (3) has exactly one solution in this case, namely

$$y(x) = \phi(x) + \eta. \tag{4}$$

The solution exists in all of J.

Note: If J is not compact, then f need not be bounded and as a result may not be integrable over J. However, since $[\xi, x]$ is a compact subinterval of J and f is continuous on $[\xi, x]$, the integral in the definition of $\phi(x)$ exists for each $x \in J$, and the equation $\phi' = f$ holds in all of J.

Example. The equation

$$y' = x^3 + \cos x$$

has as solutions

$$y(x; C) = \frac{1}{4}x^4 + \sin x + C.$$

If the initial condition is $y(1) = 1$, then the corresponding solution is

$$y = \frac{1}{4}x^4 + \sin x - \sin 1 + \frac{3}{4}.$$

Thus the problem of finding a solution to a differential equation of type (IV) is purely a problem in integral calculus—that of finding an antiderivative of a given function $f(x)$. This motivates a commonly used expression: "Integrating a differential equation" is synonymous with finding its solutions.

V. $\boxed{y' = g(y)}$

Let $g(y)$ be continuous in an interval J. The direction field here is similar to the one in IV but with the roles of x and y exchanged. This suggests interchanging x and y and writing the solution curves in the form $x = x(y)$.

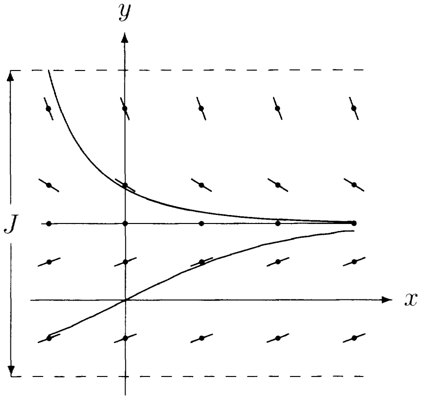
A formal calculation gives

$$\frac{dy}{dx} = g(y) \Leftrightarrow \frac{dy}{g(y)} = dx$$

and hence the solution

$$\int \frac{dy}{g(y)} = \int dx = x + C. \tag{5}$$

If $g \neq 0$, then (5) gives a function $x = x(y)$ whose inverse function $y(x)$ is a solution to the differential equation, as we will show in VII. Finding a solution



Direction field in the case where the right side depends only on y

that satisfies an initial condition $y(\xi) = \eta$ involves making a choice of the constant of integration such that $x(\eta) = \xi$, i.e.,

$$x(y) = \xi + \int_\eta^y \frac{dz}{g(z)}. \tag{6}$$

This is a special case of a "differential equation with separable variables." This type of equation is investigated in more detail in VII. A theorem that will be proved in that section shows that the initial value problem has exactly one solution $y(x)$ in a neighborhood of the point ξ if $g(\eta) \neq 0$ (if $g(\eta) \neq 0$, then by continuity $g \neq 0$ in a neighborhood of η). This solution can be obtained by first using formula (6) to get $x(y)$ and then finding the inverse function $y(x)$. If $g(\eta) = 0$, then $y(x) \equiv \eta$ is a solution. In this special case, it is quite possible that there are also other solutions that pass through the point (ξ, η), i.e., that the initial value problem has more than one solution. For more in this regard, see Example 2 and the discussion in VIII.

The nature of the direction field and the location of the constant of integration in formula (5) suggest that a translation of a solution in the x-direction will produce another solution: If $y(x)$ is a solution, then so is $\bar{y}(x) := y(x + C)$. Indeed, this follows from

$$\bar{y}'(x) = y'(x + C) = g(y(x + C)) = g(\bar{y}(x)).$$

Example 1.

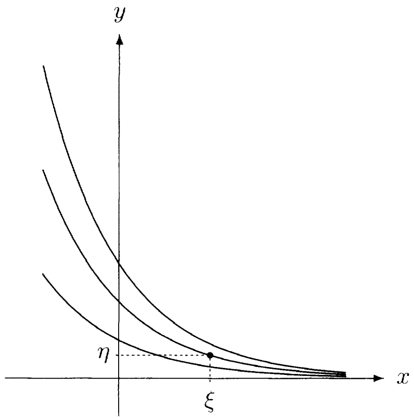
$$y' = -2y.$$

Here $D = \mathbb{R}^2$. Using the procedure in (5) one obtains

$$\frac{dy}{y} = -2\,dx \Leftrightarrow \ln|y| = -2x + C \Leftrightarrow |y| = \mathrm{e}^{C-2x}.$$

The general solution (with $\pm\mathrm{e}^C$ replaced with C) is

$$y(x; C) = C\mathrm{e}^{-2x} \quad (C \in \mathbb{R}).$$



Solution curves $y = Ce^{-2x}$ of the differential equation $y' = -2y$

The proof that every solution is of this form is elementary: If $\phi(x)$ is any solution of the differential equation, then

$$(\phi e^{2x})' = \phi' e^{2x} + 2\phi e^{2x} = 0,$$

i.e., ϕe^{2x} is a constant. (One could also appeal to the uniqueness statement proved in VII.) It follows that exactly one solution passes through each point (ξ, η), namely,

$$y(x; \eta e^{2\xi}) = \eta e^{2(\xi - x)}.$$

Thus we have shown that the initial value problem is uniquely solvable, with a solution that exists in $\mathbb{R}$.

Example 2.

$$y' = \sqrt{|y|}.$$

Again $D = \mathbb{R}^2$. Since the direction field is symmetric, it follows that if $y(x)$ is a solution, then $z(x) = -y(-x)$ is also a solution. Indeed, we have

$$z'(x) = y'(-x) = \sqrt{|y(-x)|} = \sqrt{|z(x)|}.$$

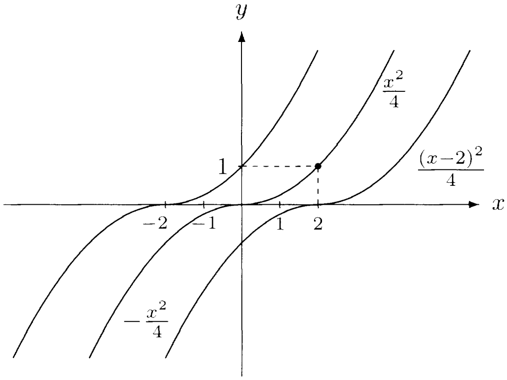
Thus it is sufficient to consider only positive solutions. From (5) it follows that

$$\int \frac{dy}{\sqrt{y}} = 2\sqrt{y} = x + C,$$

hence

$$y(x; C) = \frac{(x+C)^2}{4} \quad \text{in} \quad (-C, \infty) \quad (C \in \mathbb{R})$$

(note that $\sqrt{y}$ is positive, whence $x > -C$, and that for $x < -C$ this formula does not give a solution to the differential equation). This function gives all of the positive solutions (this also follows from the uniqueness statement in VII).



Solution curves of the differential equation $y' = \sqrt{|y|}$

Additionally, $y \equiv 0$ is also a solution, and the functions $-y(-x; C)$ give the negative solutions; they exist for $x < C$.

Solutions that exist in all of $\mathbb{R}$ can be constructed by piecing these functions together; for example,

$$\phi(x) = \begin{cases} x^2/4 & \text{for} \quad x > 0, \\ 0 & \text{for} \quad -2 \le x \le 0, \\ -(x+2)^2/4 & \text{for} \quad x < -2. \end{cases}$$

Note (and check for yourself!) that at the "splice points" the function ϕ is differentiable and satisfies the differential equation.

In this example we encounter for the first time the phenomenon of

VI. Nonuniqueness. It is easy to see that every initial value problem in Example 2 has infinitely many solutions. For instance, the set of solutions through the point $(2, 1)$ is given by the functions

$$\phi(x; a) = \begin{cases} x^2/4 & \text{for} \quad x > 0, \\ 0 & \text{for} \quad a \le x \le 0, \\ -(x-a)^2/4 & \text{for} \quad x < a, \end{cases}$$

where a is any nonpositive number, together with

$$\psi(x) = \begin{cases} x^2/4 & \text{for} \quad x > 0, \\ 0 & \text{for} \quad x \le 0 \end{cases}$$

(we recall the convention, introduced in III.(c), that restrictions of solutions will not be regarded as separate solutions).

Two types of nonuniqueness are illustrated in this example, depending on whether the initial value $\eta = y(\xi)$ is zero or different from zero. In the first case, the solutions all branch directly at the point (ξ, η), and in the second case, the solution begins as a unique solution which can then split at some distance from (ξ, η).

In the latter case, the initial value problem is said to be *locally uniquely solvable*. This means that there exists a neighborhood U of the point ξ such that exactly one solution of the initial value problem exists in U. Thus, in Example 2 of Section V, initial value problems with $\eta \neq 0$ are locally uniquely solvable, but those with $\eta = 0$ are not.

VII. $\boxed{y' = f(x)g(y)}$ **Equations with Separated Variables.** This class of equations, which includes the types discussed in IV and V as special cases, can also be solved by quadrature using the *method of separation of variables.* We will describe this method first in heuristic terms. One goes from

$$\frac{dy}{dx} = f(x)g(y) \quad \text{to the equation} \quad \frac{dy}{g(y)} = f(x)\,dx$$

and then by integration to the equation

$$\int \frac{dy}{g(y)} = \int f(x)\,dx, \tag{7}$$

from which a solution can be obtained by solving for y. In order to get the solution that passes through the point (ξ, η), it is necessary to choose the limits of integration such that equation (7) is satisfied when $x = \xi$, $y = \eta$. This is accomplished by setting

$$\int_\eta^y \frac{ds}{g(s)} = \int_\xi^x f(t)\,dt. \tag{8}$$

The following theorem gives conditions under which this procedure is permitted. It concerns the initial value problem

$$y' = f(x)g(y), \quad y(\xi) = \eta \tag{9}$$

under the following general hypothesis:

(H) $f(x)$ is continuous in an interval J_x; $g(y)$ is continuous in an interval J_y; and $\xi \in J_x$, $\eta \in J_y$.

Theorem. *Let η be an interior point of J_y with $g(\eta) \neq 0$ and let* (H) *hold. Then there exists a neighborhood of ξ (in the case where ξ is a boundary point of J_x, a one-sided neighborhood) in which the initial value problem* (9) *has a unique solution $y(x)$. It can be obtained from equation* (8) *by solving for y.*

Proof. We recall the following result from analysis: If ϕ is a differentiable function in an interval J and $\phi'(x) \neq 0$, then ϕ has a differentiable inverse function $\psi : J' \to J$, where $J' = \phi(J)$.

Denote the left-hand side of (8) by $G(y)$, the right-hand side by $F(x)$. In this notation (8) becomes $G(y) = F(x)$. The function $g(y) \neq 0$ in a neighborhood of η. Therefore, $G(y)$ exists in this neighborhood, and because $G' = 1/g \neq 0$, G has an inverse function H. Applying H to both sides of the equation $G(y) = F(x)$ and using $y \equiv H(G(y))$ gives

$$y(x) = H(F(x)). \tag{10}$$

We will show that $y(x)$ satisfies (9). Since H and F are differentiable, it follows that y is differentiable. Differentiation of the identity $G(y(x)) = F(x)$ yields

$$G'(y(x)) \cdot y'(x) = F'(x) = f(x).$$

Since $G' = 1/g$, it follows that y satisfies the differential equation

$$y'(x) = f(x)g(y(x)).$$

Furthermore, from the relations $F(\xi) = 0$, $G(\eta) = 0$, $H(0) = \eta$ we have that $y(\xi) = H(F(\xi)) = \eta$. This shows that $y(x)$ is a solution of the initial value problem (9).

We now show that there are no other solutions. Suppose $z(x)$ is another solution. Then as long as $g(z) \neq 0$ (this is certainly true in a neighborhood of ξ), the equation

$$\frac{z'(x)}{g(z(x))} = f(x)$$

holds. Integrating this identity between ξ and x and using the change of variables $s = z(x)$, one obtains

$$\int_\xi^x f(t)\,dt = \int_\xi^x \frac{z'(t)\,dt}{g(z(t))} = \int_\eta^{z(x)} \frac{ds}{g(s)}.$$

In the notation introduced earlier, this equation says that $F(x) = G(z(x))$, and therefore $z(x) = H(F(x)) = y(x)$. ∎

VIII. The Case $g(\eta) = 0$. If $g(\eta) = 0$ in (9), then one solution can be immediately given: $y(x) \equiv \eta$. However, it may be the case that there are other solutions, as we have already seen in Example 2 of V.

Theorem. *Let hypothesis* (H) *from* VII *hold, let $g(\eta) = 0$ and $g(y) \neq 0$ in an interval $\eta < y \leq \eta + \alpha$ or $\eta - \alpha \leq y < \eta$, ($\alpha > 0$), and let the improper integral*

$$\int_\eta^{\eta+\alpha} \frac{dz}{g(z)} \quad \text{or} \quad \int_{\eta-\alpha}^{\eta} \frac{dz}{g(z)},$$

respectively, be divergent. Then every solution that starts above or below the line $y = \eta$ remains (strictly) above or below this line (in both directions).

It follows from this theorem that a solution $y(x)$ which satisfies $y > \eta$ at one point remains greater than η for all x (a corresponding statement holds for $<$). In particular, if η is an interior point of J_y and if both integrals diverge, then the initial value problem (9) has only one solution $y(x) \equiv \eta$. This is the case, for example, if $g(y)$ has an isolated zero at the point η and satisfies a Lipschitz condition at η

$$|g(y) - g(\eta)| = |g(y)| \le K|y - \eta|,$$

hence, in particular, if $g'(\eta)$ exists and is different from 0.

Proof. Let us assume that there exists a solution $y(x)$ to the initial value problem that is not identically equal to η. To focus in on one of the four possible cases, suppose that there exists a point $\bar{\xi}$ to the right of ξ such that $\eta < y(\bar{\xi}) = \bar{\eta} < \eta + \alpha$. Then by (8) with $(\bar{\xi}, \bar{\eta})$ in place of (ξ, η), we have that

$$\int_{\bar{\eta}}^{y(x)} \frac{ds}{g(s)} = \int_{\bar{\xi}}^{x} f(t)\,dt$$

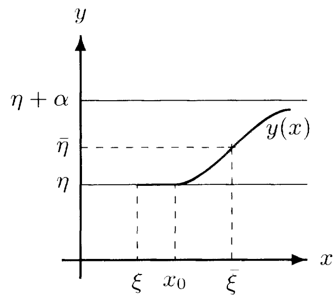
holds, at least for as long as $y(x)$ remains in the strip $\eta < y(x) \le \eta + \alpha$. Suppose that x_0 is the first point to the left of $\bar{\xi}$ with $y(x_0) = \eta$. Then the above formula leads immediately to a contradiction, since the integral on the right stays bounded as $x \to x_0$, while the one on the left goes to infinity. ∎

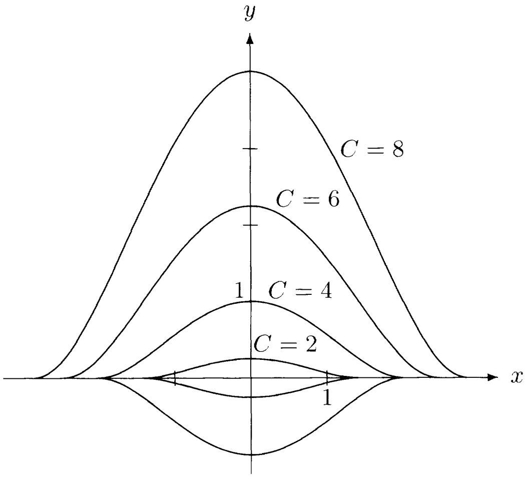
We return again to the two examples from V that correspond to the integrals $\displaystyle\int_0^{\pm\alpha} \frac{dy}{y}$ and $\displaystyle\left|\int_0^{\pm\alpha} \frac{dy}{\sqrt{|y|}}\right|$. In Example 1, the integrals on the left are divergent, and thus $y \equiv 0$ is the only solution through the origin. In Example 2, the integrals on the right are convergent, and there are several solutions.

However, it is entirely possible that the integral converges and the solution to the initial value problem is still unique. To illustrate this point, consider the following:

Example 1.

$$y' = -x(\operatorname{sgn} y)\sqrt{|y|} = \begin{cases} -x\sqrt{y} & \text{for} \quad y \ge 0, \\ x\sqrt{-y} & \text{for} \quad y < 0. \end{cases}$$





Solution curves of the differential equation $y' = -x(\operatorname{sgn} y)\sqrt{|y|}$

The direction field is symmetric to the x-axis; i.e., if $y(x)$ is a solution, then so is $-y(x)$. Thus it is sufficient to calculate the positive solutions. From

$$\int \frac{dy}{\sqrt{y}} = -2\sqrt{y} = -\int x\,dx = \frac{1}{2}(C - x^2)$$

it follows that

$$y(x;C) = \frac{1}{16}(C - x^2)^2 \quad \text{in} \quad (-\sqrt{C}, \sqrt{C}) \quad (C > 0)$$

(note that $\sqrt{y} > 0$). If this function is extended by setting $y(x;C) = 0$ for $|x| \geq \sqrt{C}$, then one clearly has a solution defined in $\mathbb{R}$. Thus we have the solutions $\pm y(x;C)$ for $C > 0$ and $y \equiv 0$. There are no other solutions. On the one hand, they (that is, their graphs) cover the whole plane; on the other hand, $g(y) = \sqrt{|y|}$ vanishes only for $y = 0$. Thus each initial value problem with $\eta \neq 0$ is locally uniquely solvable.

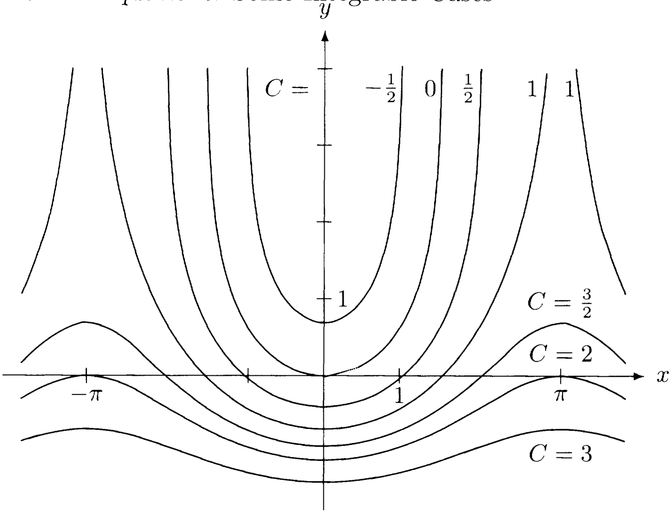
One can see from the figure that every initial value problem with $y(\xi) = \eta \neq 0$ is uniquely solvable, not only locally, but also globally. For the initial condition $y(\xi) = 0$, there are infinitely many solutions in the case where $\xi \neq 0$, but only one solution in the case where $\xi = 0$.

Example 2.

$$y' = e^y \sin x.$$

The direction field is symmetric with respect to the y-axis and periodic in x of period 2π, i.e., if $y(x)$ is a solution, then so are $u(x) = y(-x)$ and $v(x) = y(x + 2k\pi)$. By separation of variables (7) one obtains

$$\int e^{-y}\,dy = -e^{-y} = \int \sin x\,dx = -\cos x - C;$$



Solution curves of the differential equation $y' = \mathrm{e}^y \sin x$

i.e.,

$$y(x; C) = -\log(\cos x + C) \quad (C + \cos x > 0).$$

The reader should verify that all solutions have been found and that each initial value problem is uniquely solvable.

This example exhibits a new and important phenomenon. The solutions can have quite different behavior, depending on the value of the constant C. While in the case $C > 1$ solutions exist for all x and are bounded, in the case $-1 < C \leq 1$ the solutions exist only in finite intervals and increase without bound.

Consider, for purposes of illustration, the initial condition $y(0) = \eta$. The corresponding solution is

$$y(x; \mathrm{e}^{-\eta} - 1) = -\log(\cos x + \mathrm{e}^{-\eta} - 1).$$

In particular, if $\eta = -\log 2$, then y is given by

$$y(x; 1) = -\log(1 + \cos x).$$

This solution exists in $(-\pi, \pi)$ and cannot be extended beyond this interval. It tends to ∞ as $x \to \pm\pi$. Solutions with $\eta < -\log 2$ exist in all of $\mathbb{R}$ and are bounded. For $\eta > -\log 2$, the solutions exist only in the interval $|x| < \arccos(1 - \mathrm{e}^{-\eta})$; the length of this interval of existence converges to 0 as $\eta \to \infty$.

Existence and Behavior in the Large. This example shows, first of all, that the solution of an initial value problem does not necessarily exist in all of $\mathbb{R}$, but possibly only in a very small interval, and that this is true even if the right-hand side of the differential equation is defined and "smooth" on all of $\mathbb{R}^2$. Take another look at the formulation of Theorem VII in this regard. This raises the question whether it is possible to make any general statements about the domain of definition of a solution. We will prove in 6.VII that a solution can always be extended to the boundary of D (D is the domain of the right-hand side of the differential equation). Secondly, the example also shows that the behavior of solutions "in the large" can change dramatically with small changes in the initial conditions; this occurs for solutions with initial values $y(0) = \eta$, where η is close to $-\log 2$, i.e., $C = 1$.

The three types of differential equations that follow can be reduced by simple transformations to equations of the types already discussed. In all three cases the function $f(s)$ that appears is assumed to be continuous in an interval.

IX. $\boxed{y' = f(ax + by + c)}$

The structure of the differential equation suggests that we look for a solution of the form

$$u(x) = ax + by(x) + c \tag{11}$$

(the case $b \neq 0$ is the only interesting one). If $y(x)$ is a solution, then $u(x)$ satisfies

$$u' = a + by'(x) = a + bf(u), \tag{12}$$

which is a solvable equation of type V. Conversely, it is easy to see that a solution $y(x)$ of (11) can be obtained from a solution $u(x)$ of (12). All solutions can be obtained in this manner.

Example.

$$y' = (x + y)^2 + .$$

Using the ansatz $u(x) = x + y(x)$ we have

$$u' = u^2 + 1, \quad \text{and hence} \quad u = \tan(x + C)$$

(why does this give all the solutions?). The general solution can be written

$$y(x; C) = \tan(x + C) - x.$$

X. $\boxed{y' = f\left(\frac{y}{x}\right)}$ **Homogeneous Differential Equation.**

Using the ansatz $u(x) = \dfrac{y(x)}{x}$ $(x \neq 0)$ and calculating the derivative, one obtains the relation

$$y' = u + xu' = f(u),$$

and thus a differential equation for $u(x)$ with separated variables,

$$u' = \frac{f(u) - u}{x}. \tag{13}$$

One sees immediately that every solution $u(x)$ of (13) leads to a solution $y(x) = x \cdot u(x)$ of the given differential equation.

Example. The initial value problem

$$y' = \frac{y}{x} - \frac{x^2}{y^2}, \quad y(1) = 1$$

transforms into an initial value problem

$$u' = -\frac{1}{xu^2}, \quad u(1) = 1$$

with the solution

$$\int_1^u z^2\,dz = -\int_1^x \frac{dt}{t}; \quad \text{i.e.,} \quad \frac{u^3 - 1}{3} = -\log x.$$

Thus the solution to the original problem is given by

$$y = x\sqrt[3]{1 - 3\log x} \quad \text{for} \quad 0 < x < \sqrt[3]{\mathrm{e}} \approx 1.396.$$

XI. $$\boxed{y' = f\left(\frac{ax + by + c}{\alpha x + \beta y + \gamma}\right)}$$

In the case where the determinant $\begin{vmatrix} a & b \\ \alpha & \beta \end{vmatrix} = 0$, that is, where $a = \lambda\alpha$ and $b = \lambda\beta$, the equations can be reduced to one of the types we have already considered. If this determinant is not zero, then the linear system of equations

$$\begin{aligned} ax + by + c &= 0, \\ \alpha x + \beta y + \gamma &= 0 \end{aligned} \tag{14}$$

has a unique solution (x_0, y_0). If a new system of coordinates $(\bar{x}, \bar{y})$ is introduced by translating the origin to the point (x_0, y_0),

$$\bar{x} := x - x_0, \quad \bar{y} := y - y_0,$$

then in the new coordinate system a solution curve $y(x)$ is described by the function

$$\bar{y}(\bar{x}) := y(\bar{x} + x_0) - y_0.$$

The differential equation in the $(\bar{x}, \bar{y})$ coordinate system

$$\begin{aligned} \frac{d\bar{y}(x)}{d\bar{x}} = y'(\bar{x} + x_0) &= f\left(\frac{a(\bar{x} + x_0) + b(\bar{y}(\bar{x}) + y_0) + c}{\alpha(\bar{x} + x_0) + \beta(\bar{y}(\bar{x}) + y_0) + \gamma}\right) \\ &= f\left(\frac{a\bar{x} + b\bar{y}(\bar{x})}{\alpha\bar{x} + \beta\bar{y}(\bar{x})}\right) = f\left(\frac{a + b\bar{y}/\bar{x}}{\alpha + \beta\bar{y}/\bar{x}}\right) \end{aligned}$$

is just the special case $c = \gamma = 0$ of the original equation. It is homogeneous and can be solved using the techniques in X.

How to proceed. (i) Determine the point (x_0, y_0) that satisfies (14).

(ii) Solve the differential equation with $c = \gamma = 0$ using techniques from X (this equation is homogeneous).

(iii) A solution $\bar{y}(\bar{x})$ of this equation generates a solution to the original equation using the substitution $\bar{x} = x - x_0$, $\bar{y} = y - y_0$, that is, $y(x) := y_0 + \bar{y}(x - x_0)$.

We will illustrate these steps in the following

Example.

$$y' = \frac{y+1}{x+2} - \exp\left(\frac{y+1}{x+2}\right).$$

From (14) we obtain $x_0 = -2$, $y_0 = -1$. The differential equation for $\bar{y}$ is

$$\frac{d\bar{y}}{d\bar{x}} = \frac{\bar{y}}{\bar{x}} - \exp\left(\frac{\bar{y}}{\bar{x}}\right),$$

and for $u = \bar{y}/\bar{x}$ the differential equation is

$$\bar{x}u' = -e^u,$$

which gives

$$-\int e^{-u}\, du = \int \frac{1}{\bar{x}}\, d\bar{x} \quad \text{or} \quad e^{-u} = \log|\bar{x}| + C = \log c|\bar{x}|$$

(the constant of integration has been written as $C = \log c$ $(c > 0)$). One obtains $u = \log(\log c|\bar{x}|)$ as long as $c|\bar{x}| > 1$. The functions

$$y(x) = -1 - (x+2)\log(\log c|x+2|) \quad \text{for} \quad c|x+2| > 1$$

are the solutions of the original differential equation.

The value $c = \frac{1}{2}e^{e^{-1/2}}$ gives the solution that passes through the origin. It exists for $x > 1/c - 2 \approx -0.9095$.

XII. Exercises. (a) In the above example, determine a solution y with the property that $\lim_{x \to 0+} y(x) = \infty$. Is it uniquely determined?

Determine all of the solutions to the following differential equations and find the particular solution that passes through the origin.

(b) $y' = \dfrac{y+1}{x+2} + \exp\left(\dfrac{y+1}{x+2}\right)$.

(c) $y' = \dfrac{x+y+1}{x+2} - \exp\left(\dfrac{x+y+1}{x+2}\right)$.

(d) $y' = \dfrac{x+2y+1}{2x+y+2}$.

(e) $y' = \dfrac{2x+y+1}{x+2y+2}$.

XIII. Exercises. Determine all solutions of the differential equations and in each case sketch the solution curves and determine the set of all points (ξ, η) for which the initial value problem is not locally unique.

(a) $y' = 3|y|^{2/3}$ $(y \in \mathbb{R})$.

(b) $y' = 3(\operatorname{sgn} y)|y|^{2/3}$ $(y \in \mathbb{R})$.

(c) $y' = \sqrt{|y|(1-y)}$ $(y \le 1)$.

Solve the following initial value problems and in each case give the maximal interval of existence of the solution.

(d) $y' = \dfrac{e^{-y^2}}{y(2x+x^2)}$, $y(2) = 0$.

(e) $y' = \dfrac{y \ln y}{\sin x}$, $y(\pi/2) = e^e$.

(f) $y' = \dfrac{\cos x}{\cos^2 y}$, $y(\pi) = \dfrac{\pi}{4}$.

Determine all solutions of the differential equations

(g) $y' = (x - y + 3)^2$,

(h) $y' = \dfrac{2y(y-1)}{x(2-y)}$,

(i) $y = xy' - \sqrt{x^2 + y^2}$.

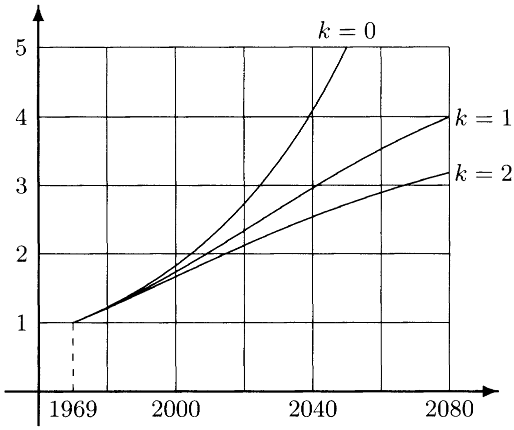
Give a differential equation of the first order for the following families of curves (parameter $c \in \mathbb{R}$).

(j) $y = cx^2$,

(k) $y = cx^2 + c$,

(l) $y = cx^2 + (\operatorname{sgn} c)c^2$.

XIV. Population Growth Models. In this section we investigate some simple ecological models for the growth of a population. Let $y(t)$ be the size of a population at time t. If the relative growth rate of the population per unit time is denoted by $c = c(t, y)$, then $\dfrac{y'}{y} = c$; i.e., $y' = cy$. In any ecological system, the resources available to support life are limited, and this in turn places a limit on the size of the population that can survive in the system. The number N denoting the size of the largest population that can be supported by the system is called the *carrying capacity* of the ecosystem. We consider a sequence of three single-population models, which incorporate the following assumptions: The relative population growth rate depends only on y (that is to



Population in multiples of y_0 (year 1969) under different assumptions ($\beta = 5$)

say, not explicitly on t) and goes to zero as the population approaches N. In particular, we assume that $c = c(y)$ is given by one of the following:

$$c(y) = \alpha(N - y)^k \quad \text{with} \quad k = 0, 1, 2.$$

To illustrate these ideas we will model the human population of the earth and choose the year 1969 as the starting point ($t = 0$, with t measured in years). Let y_0 denote the population of the earth in the year 1969 and c_0 the relative annual population growth rate for the year 1969. These are given by $y_0 = 3.55 \cdot 10^9$ and $c_0 = 0.02$. From the condition $c(0) = c_0$, it follows that $\alpha = c_0(N - y_0)^{-k}$.

If we measure $y(t)$ in multiples of y_0, i.e., set $y(t) = y_0 u(t)$, $N = \beta y_0$, where β gives the carrying capacity in multiples of y_0, then one obtains the initial value problem

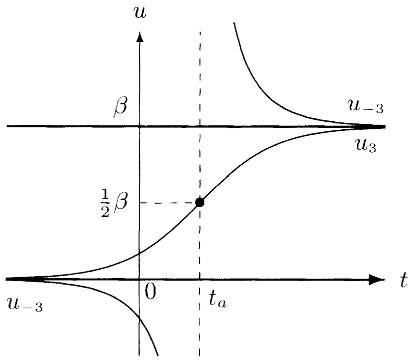
$$u' = c_0 \left(\frac{\beta - u}{\beta - 1}\right)^k \cdot u, \quad u(0) = 1 \qquad (k = 0, 1, 2). \tag{15}$$

If $k = 0$, equation (15) reduces to the equation $u' = c_0 u$, which produces the well-known exponential growth function $u(t) = \mathrm{e}^{c_0 t}$. For the other cases,

$$k = 1: \; c_0 t = (\beta - 1) \int_1^u \frac{ds}{s(\beta - s)} = \frac{\beta - 1}{\beta} \log\left(\frac{(\beta - 1)u}{\beta - u}\right),$$

$$\begin{aligned} k = 2: \; c_0 t &= (\beta - 1)^2 \int_1^u \frac{ds}{s(\beta - s)^2} \\ &= \left(\frac{\beta - 1}{\beta}\right)^2 \left\{\log\left(\frac{(\beta - 1)u}{\beta - u}\right) + \frac{\beta}{\beta - u} - \frac{\beta}{\beta - 1}\right\}. \end{aligned}$$

Solving these equations for u is easy when $k = 1$ but difficult when $k = 2$; however, for many questions, solving for u is not necessary For instance, we can calculate the doubling time by putting $u = 2$. For the case $k = 0$, the population of the year 1969 doubles in $50 \cdot \log 2 = 34.7$ years, and if $\beta = 5$ is



Logistic equation.
The solutions u_3 and u_{-3}
with $t_a = \dfrac{1}{b}\ln 3$

used, it doubles in $50 \cdot \frac{4}{5} \cdot \frac{8}{3} = 39.2$ years in the case $k = 1$ and in 44.8 years for $k = 2$.

The Logistic Equation. The equation with $k = 1$ is called the logistic equation. It was proposed as early as 1838 by the Belgian mathematician Pierre-François Verhulst (1804–1849). We will consider this equation in more detail using different notation:

$$\boxed{u' = u(b - cu) \quad \text{with} \quad b, c > 0} \qquad \textit{logistic equation.} \tag{16}$$

This equation is the same as (15) with $k = 1$ and

$$c = \frac{c_0}{\beta - 1}, \qquad b = \beta c. \tag{17}$$

Using the methods described in VII one obtains the solutions

$$u_\gamma = \frac{b}{c} \cdot \frac{1}{1 + \gamma e^{-bt}} \quad \text{for} \quad \gamma \neq 0 \tag{18}$$

as well as two stationary solutions $u \equiv 0$ and $u \equiv b/c$ (the reader should check this). These are all of the solutions. On the one hand, every initial condition $u(t_0) = u_0$ can be satisfied by one of these solutions, on the other hand, by the results in VII and VIII, exactly one solution goes through each point.

Two simple propositions follow:

(a) *Every solution u of* (16) *with* $u(t_0) > 0$ *remains positive for* $t > t_0$ *and tends to* b/c *as* $t \to \infty$.

(b) $u_\gamma'' = 0$ *if and only if* $u_\gamma = b/(2c)$.

The proof of (a) is obvious from (18). For (b), we differentiate (16), obtaining

$$u'' = u'(b - cu) - cuu' = u'(b - 2cu). \qquad \square$$

In population models, u_γ with $\gamma > 0$ describes the growth of the population, and b/c is the carrying capacity β. We will now check how the world population $y(t) = y_0 u(t)$ has grown since 1969 ($t = 0$) according to this model. Recall that $c_0 = 0.02$. We have $u_\gamma(0) = 1$, and we obtain $\gamma = \beta - 1$ from (17), (18). Under the assumption $\beta = 3$ ($b = 0.03$) one obtains a population of $y = 5.157$ billion for the year 1990 ($t = 21$) and with $\beta = 5$ ($b = 0.025$); $y = 5.273$ billion. The actual population size in 1990 was 5.321 billion. The assumption $\beta = 5$ gives a better approximation, though $\beta = 3$ corresponds to the carrying capacity $N = y_0\beta \approx 10$ billion which is sometimes used by demographers.

The *point of inflection* marks the turning point where the second derivative becomes negative, and hence the point beyond which the yearly population growth rate begins to decrease. It occurs in this model at $\beta/2$ according to (b). Applying this result to the world population under the assumption that $\beta = 3$ would mean that we have already passed this point ($N/2 = y_0\beta/2 \approx 5$ billion). The situation fits $\beta = 5$ better. One should, however, not forget that we are dealing with the simplest growth model with bounded growth.

§ 2. The Linear Differential Equation. Related Equations

A linear differential equation is an equation of the form

$$y' + g(x)y = h(x); \tag{1}$$

we assume that the two given functions $g(x)$ and $h(x)$ are continuous on an interval J. If $h(x) \equiv 0$, then equation (1) is called *homogeneous*, otherwise *nonhomogeneous* or *inhomogeneous*. The differential operator

$$Ly := y' + g(x)y \tag{2}$$

can be used to write the differential equation (1) in the form

$$Ly = 0 \quad (\textit{homogeneous}) \quad \text{and} \quad Ly = h(x) \quad (\textit{nonhomogeneous}).$$

Thus, to each function $\phi \in C^1(J)$ the operator L associates a function $\psi = L\phi = \phi' + g\phi \in C^0(J)$. The value of the function $L\phi$ at the point x will be denoted by $(L\phi)(x)$.

The operator L is linear; i.e., if ϕ, ψ belong to $C^1(J)$ and $a, b \in \mathbb{R}$ are arbitrary constants, then

$$L(a\phi + b\psi) = aL\phi + bL\psi.$$

I. $\boxed{Ly := y' + g(x)y = 0}$ **The Homogeneous Equation.**

This is an equation with separated variables which can be solved using the techniques discussed in 1.VII and 1.VIII. From formula (1.8) we obtain the family of solutions

$$y(x; C) = C \cdot \mathrm{e}^{-G(x)} \text{ with } G(x) = \int_\xi^x g(t)\,dt \quad (\xi \in J \text{ fixed}) \tag{3}$$

(recall that g is continuous in J). It is easy to check that (3) gives a solution for every real C and that exactly one solution from this family passes through a given point $(\xi, \eta) \in J \times \mathbb{R}$. There are no other solutions, since by the theorems proved in 1.VII, VIII there exists exactly one solution through each point in $J \times \mathbb{R}$. The fact that every solution is given by (3) can also be verified directly: If ϕ is a solution of $L\phi = 0$ and $u(x) := \mathrm{e}^{G(x)}\phi(x)$, then $u' = \mathrm{e}^{G(x)}(g\phi + \phi') = 0$; i.e., u is constant and hence ϕ has the form (3).

The unique solution satisfying the initial condition $y(\xi) = \eta$ is given by

$$y(x) = \eta \cdot \mathrm{e}^{-G(x)} \quad \text{with} \quad G(x) = \int_\xi^x g(t)\,dt. \tag{4}$$

It exists in all of J.

II. $\boxed{Ly = h(x)}$ The Nonhomogeneous Equation.

Solutions to the nonhomogeneous equation can be obtained with the help of an ansatz that goes back to Lagrange, the *method of variation of constants.* In this method, the constant C in the general solution $y(x; C) = C\mathrm{e}^{-G(x)}$ of the homogeneous equation is replaced by a function $C(x)$. The calculation of an appropriate choice of $C(x)$ gives a solution of the nonhomogeneous equation. Indeed, the ansatz

$$y(x) = C(x)\mathrm{e}^{-G(x)} \quad \text{with} \quad G(x) = \int_\xi^x g(t)\,dt$$

leads to

$$Ly \equiv y' + gy = (C' - gC + gC)\mathrm{e}^{-G(x)} = C'\mathrm{e}^{-G(x)}.$$

Hence $Ly = h$ holds if and only if

$$C' = h(x)\mathrm{e}^{G(x)}, \quad \text{or equivalently,} \quad C(x) = \int_\xi^x h(t)\mathrm{e}^{G(t)}\,dt + C_0. \tag{5}$$

Theorem. *If the functions $g(x)$, $h(x)$ are continuous in J and $\xi \in J$, then the initial value problem*

$$Ly = y' + g(x)y = h(x), \quad y(\xi) = \eta \tag{6}$$

has exactly one solution,

$$y(x) = \eta \cdot \mathrm{e}^{-G(x)} + \mathrm{e}^{-G(x)} \int_\xi^x h(t)\mathrm{e}^{G(t)}\,dt. \tag{7}$$

The solution exists in all of J.

The discussion leading up to formula (5) shows that (7) is a solution to $Ly = h$; it is clear that the initial conditions are satisfied. Uniqueness is a consequence of (a) below.

Remark on Linearity. If y, $\bar{y}$ are two solutions to the nonhomogeneous equation $Ly = h$, then $L(y - \bar{y}) = Ly - L\bar{y} = 0$, i.e., $z(x) = y - \bar{y}$ is a solution of the homogeneous equation $Ly = 0$. Thus all solutions $y(x)$ of the nonhomogeneous equation can be written in the form

$$y(x) = \bar{y}(x) + z(x), \tag{8}$$

where $\bar{y}(x)$ is a fixed solution of the nonhomogeneous equation and $z(x)$ runs through all solutions of the homogeneous equation. In other words,

$$y(x; C) = \bar{y}(x) + C\mathrm{e}^{-G(x)} \quad (C \in \mathbb{R}) \tag{8'}$$

is the general solution of the inhomogeneous equation.

It follows from (4) that a solution z of the homogeneous equation that vanishes at a point ξ is identically zero (note that ξ can be any point in J). Using (8), this result implies

(a) Two solutions y, $\bar{y}$ of the inhomogeneous equation that coincide at one point in J are identical.

Example.

$$y' + y \sin x = \sin^3 x.$$

Here $G(x) = -\cos x$. Hence $z(x; C) = C\mathrm{e}^{\cos x}$ is the general solution of the homogeneous equation $Lz = 0$ and

$$\begin{aligned}
\bar{y}(x) &= \int_0^x \sin^3 t \cdot \mathrm{e}^{\cos x - \cos t}\, dt \\
&= \mathrm{e}^{\cos x} \int_1^{\cos x} (s^2 - 1)\mathrm{e}^{-s}\, ds \\
&= -\mathrm{e}^{\cos x}((s^2 - 1) + 2s + 2)\mathrm{e}^{-s}\Big|_1^{\cos x} \\
&= \sin^2 x - 2\cos x - 2 + 4\mathrm{e}^{\cos x - 1}
\end{aligned}$$

is a solution to the nonhomogeneous differential equation. It follows that the general solution of the nonhomogeneous equation is given by

$$y(x; C) = \sin^2 x - 2\cos x - 2 + C \cdot \mathrm{e}^{\cos x}.$$

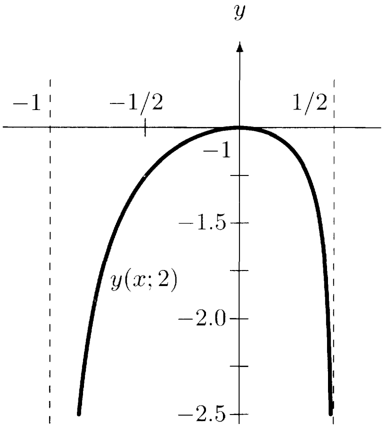
III. $\boxed{y' + g(x)y + h(x)y^\alpha = 0,\ \alpha \neq 1}$ **Bernoulli's Equation.**

This differential equation, named after Jacob Bernoulli (1654–1705), can be transformed into a linear differential equation. Let us assume that the functions g, h are continuous in J and that $y > 0$. If the equation is multiplied by $(1-\alpha)y^{-\alpha}$ and the relation $(1-\alpha)y^{-\alpha}y' = (y^{1-\alpha})'$ is used, then one obtains

$$(y^{1-\alpha})' + (1-\alpha)g(x)y^{1-\alpha} + (1-\alpha)h(x) = 0.$$

Thus the function $z = y^{1-\alpha}$ satisfies a linear differential equation,

$$z' + (1-\alpha)g(x)z + (1-\alpha)h(x) = 0. \tag{9}$$



Solution of the initial value problem
$y' + \dfrac{y}{1+x} + (1+x)y^4 = 0,\ y(0) = -1$

Conversely, if $z(x)$ is a positive solution of (9), then the function $y(x) = (z(x))^{1/(1-\alpha)}$ is a positive solution of Bernoulli's differential equation. For $\eta > 0$, the initial condition $y(\xi) = \eta$ transforms into $z(\xi) = \eta^{1-\alpha} > 0$. By Theorem II, this condition uniquely defines a solution z of (9). Hence each initial value problem for the Bernoulli equation with a positive initial value at the point ξ is uniquely solvable.

The cases where also nonpositive solutions occur will be discussed now.

(a) $\alpha > 0$: Then the differential equation is defined for $y \geq 0$, and $y \equiv 0$ is a solution. Since all positive solutions can be given explicitly, it is easy to determine, on a case by case basis, whether or not solution curves run into the x-axis from above. This is the case, for example, for $g = 0$, $h = -1$, $\alpha = \frac{1}{2}$ (Example 2 from 1.V).

(b) α *an integer:* Then $y < 0$ is also permitted. There are two cases.
α *odd:* It follows from the Bernoulli equation that

$$(-y)' + g(x)(-y) + h(x)(-y)^\alpha = 0.$$

So if $y(x)$ is a positive solution of the Bernoulli equation, then $u(x) = -y(x)$ is a negative solution. Hence initial value problems with $\eta < 0$ can be easily handled.
α *even:* Since $1-\alpha$ is odd, $y < 0$ implies $z = y^{1-\alpha} < 0$, which in turn yields $y = -|z|^{1/(1-\alpha)}$. So for a negative initial value η, the negative solution z of (9) with $z(\xi) = \eta^{1-\alpha}$ leads to a negative solution $y = -|z|^{1/(1-\alpha)}$ with $y(\xi) = \eta$.

In both cases the solution y satisfying $y(\xi) = \eta < 0$ is unique.

Exercise. Show directly (without using the uniqueness theorem) that for $\alpha \geq 2$, $\alpha \in \mathbb{N}$, a solution $y \not\equiv 0$ of Bernoulli's equation has no zero in J.

Example.

$$y' + \frac{y}{1+x} + (1+x)y^4 = 0.$$

The differential equation is defined for both positive and negative y. Using $z = \dfrac{1}{y^3}$ gives, according to (9),

$$z' - \frac{3}{1+x} z - 3(1+x) = 0.$$

Clearly, $\phi = C(1+x)^3$ is the general solution of the homogeneous equation. Thus the ansatz for the nonhomogeneous equation (by variation of constants) is $z = C(x)(1+x)^3$. After a simple calculation, one obtains

$$C' = \frac{3}{(1+x)^2} \Rightarrow C(x) = \frac{-3}{1+x}.$$

Therefore, the general solution of the nonhomogeneous equation is

$$z(x;C) = C(1+x)^3 - 3(1+x)^2 = (1+x^2)(Cx + C - 3).$$

Since $\alpha = 4$ is even, one has

$$y(x;C) = \frac{\operatorname{sgn}(Cx + C - 3)}{\sqrt[3]{(1+x)^2 |Cx + C - 3|}}.$$

The solution through the point $(0, -1)$ is given by

$$y(x;2) = -\frac{1}{\sqrt[3]{(1+x)^2(1-2x)}} \quad (-1 < x < \tfrac{1}{2}).$$

IV. $\boxed{y' + g(x)y + h(x)y^2 = k(x)}$ **Riccati's Equation.**

In this equation, which is named after the Italian mathematician Jacopo Francesco Riccati (1676–1754), the functions $g(x)$, $h(x)$, $k(x)$ are assumed to be continuous in an interval J. Except in special instances, the solutions cannot be given in closed form. However, if one solution is known, then the remaining solutions can be explicitly calculated. For proof, we consider the difference of two solutions y and ϕ, $u(x) = y(x) - \phi(x)$; it satisfies the equation

$$u' + gu + h(y^2 - \phi^2) = 0.$$

Since $y^2 - \phi^2 = (y - \phi)(y + \phi) = u(u + 2\phi)$, one has

$$u' = [g(x) + 2\phi(x)h(x)]u + h(x)u^2 = 0. \tag{10}$$

Thus the difference satisfies a Bernoulli differential equation which can be converted, using the techniques described in III, into the linear differential equation

$$z' - [g(x) + 2\phi(x)h(x)]z = h(x), \quad \text{where} \quad z(x) = \frac{1}{u(x)}. \tag{11}$$

Summary. If a solution $\phi(x)$ of the Riccati equation is known, then all of the other solutions can be obtained in the form

$$y(x) = \phi(x) + \frac{1}{z(x)}, \tag{12}$$

where $z(x)$ is an arbitrary solution of the linear equation (11).

Example. $$y' - y^2 - 2xy = 2.$$

The function $\phi(x) = -\dfrac{1}{x}$ is a particular solution. Formula (11) then gives the linear differential equation

$$z' + z\left(2x - \frac{2}{x}\right) + 1 = 0.$$

The general solution to the homogeneous equation is $z(x) = Cx^2 e^{-x^2}$, from which a particular solution $\bar{z}$ of the nonhomogeneous equation can be obtained using (7) with $h = -1$:

$$\begin{aligned}
\bar{z}(x) &= -x^2 e^{-x^2} \int \frac{e^{x^2}}{x^2}\,dx \\
&= -x^2 e^{-x^2}\left(-\frac{1}{x}e^{x^2} + 2\int e^{x^2}\,dx\right) \\
&= x - 2x^2 e^{-x^2} E(x) \quad \text{with} \quad E(x) = \int_0^x e^{t^2}\,dt.
\end{aligned}$$

The integral $E(x)$ can be expressed in terms of the error function with imaginary argument. The general solution of the original Riccati equation is now obtained from (12),

$$\begin{aligned}
y(x;C) &= -\frac{1}{x} + \frac{1}{x + x^2 e^{-x^2}(C - 2E(x))} \\
&= \frac{-e^{-x^2}(C - 2E(x))}{1 + xe^{-x^2}(C - 2E(x))}.
\end{aligned}$$

Since $y(0;C) = C$, every initial value problem $y(0) = \eta$ can be immediately solved.

V. Exercises. (a) *Isoclines.* Isoclines of a differential equation $y' = f(x,y)$ are the curves $f(x,y) = \text{const}$, on which the direction field has constant slope. Sketch the direction field for the differential equation

$$y' = y^2 + 1 - x^2$$

making use of the isoclines $y^2 + 1 - x^2 = \text{const}$. Determine all solutions (one solution is evident from the direction field). Which solutions exist on an infinite interval; which exist in $\mathbb{R}$?

(b) Determine all solutions of the differential equations

$$y' + y\sin x = \sin 2x \quad \text{and} \quad y' - 3y\tan x = 1.$$

(c) Solve the initial value problem

$$y' = x^4 y + x^4 y^4,\ y(0) = \eta.$$

VI. Exercise. Suppose $f(x)$ is continuous on the half-open interval $0 < x \le 1$. What additional conditions must $f(x)$ satisfy so that every solution of the differential equation

$$y' = f(x)y \quad \text{for} \quad 0 < x \le 1$$

has the property

$$\text{(a)} \quad y(x) \to 0 \quad \text{as} \quad x \to +0\,? \qquad \text{(b)} \quad \frac{y(x)}{x} \to 0 \ \text{ as } \ x \to +0\,?$$

Investigate the same question for the differential equation

$$y' = f(x)y \log \frac{1}{y} \quad \text{for} \quad 0 < x \le 1,$$

where only solutions with $0 \le y(x) \le 1/\mathrm{e}$ are taken into consideration.

VII. Exercise. *The Riccati Differential Equation and Linear Differential Equations of Second Order.* Show that the Riccati differential equation

$$y' + g(x)y + h(x)y^2 = k(x)$$

with $g, h \in C^0(J)$, $h \in C^1(J)$, $h(x) \neq 0$ in J, can be transformed into the linear differential equation of second order

$$u'' + u'\left(g - \frac{h'}{h}\right) - khu = 0 \tag{13}$$

using the transformation

$$u(x) = \exp\left(\int h(x)y(x)\,dx\right)$$

and that conversely, a positive solution u of (13) produces a solution $y = (\log u)'/h$ of the Riccati equation. Use this relationship to solve the initial value problem

$$y' - y + \mathrm{e}^x y^2 + 5\mathrm{e}^{-x} = 0, \ y(0) = \eta.$$

Supplement: The Generalized Logistic Equation.

We consider a generalization of the logistic differential equation $u' = u(b - cu)$, where, in contrast to 1.XIII, b and c depend on t. Our objective is to derive some theorems on the asymptotic behavior of the solutions as $t \to \infty$ and on the existence of a class of distinctive, in particular periodic, solutions.

VIII. The Generalized Logistic Differential Equation

$$u'(t) = u(b(t) - c(t)u). \tag{14}$$

The functions b and c are assumed to be continuous and positive in $\mathbb{R}$. We consider only positive solutions.

Equation (14) is a Bernoulli equation. If u is a solution with $u(\tau) = u_0 > 0$, then by III, the function $y = 1/u$ satisfies

$$y' = -by + c, \quad y(\tau) = \frac{1}{u_0} = y_0, \tag{15}$$

and hence

$$y(t) = \mathrm{e}^{-B(t)} \left(y_0 + \int_\tau^t c(s)\mathrm{e}^{B(s)}\, ds \right) \tag{16}$$

with

$$B(t) = \int_\tau^t b(s)\, ds.$$

Since $u = 1/y$, we obtain the following results.

(a) A solution u with $u(\tau) > 0$ exists and is positive for all $t > \tau$. The solution also remains positive "to the left." Either it exists for all $t < \tau$ or there exists a $t_1 < \tau$ such that $u(t) \to \infty$ as $t \to t_1+$. The latter case occurs if there is a t_1 such that $y(t_1) = 0$, i.e., if $y_0 < \displaystyle\int_{-\infty}^{\tau} c(s)\mathrm{e}^{B(s)}\, ds$.

(b) If u, v are two solutions with $u(\tau) < v(\tau)$, then $u < v$ in their common interval of existence. If $u(t_0) = v(t_0)$, it follows from (16) that $u \equiv v$.

Theorem 1. *Let $\lim_{t\to\infty} B(t) = \infty$. If u is a positive solution, then*

$$\lim_{t\to\infty} u(t) = \lim_{t\to\infty} \frac{b(t)}{c(t)},$$

provided that the limit on the right side exists.

Proof. This theorem is a substantial generalization of 1.XIII.(a). It can be proved by writing y as the quotient $Z(t)/N(t)$ with $N(t) = \mathrm{e}^{B(t)}$. The result then follows using l'Hospital's rule; since both $B(t)$ and $N(t)$ tend to ∞, the rule applies. One gets $Z'(t)/N'(t) = c(t)/b(t)$, which gives the conclusion immediately. ∎

In what follows, a function g will be called T-periodic ($T > 0$) if it is defined on all of $\mathbb{R}$ and $g(t+T) = g(t)$ holds for all $t \in \mathbb{R}$.

Theorem 2. *If the coefficients b and c are T-periodic, then there exists exactly one positive T-periodic solution of* (14).

Proof. It is sufficient to show that there is exactly one solution with $u(0) = u(T) > 0$. Under this assumption $v(t) := u(t+T)$ is a solution of (14) with $v(0) = u(0)$. Then $y = 1/u$ and $z = 1/v$ both satisfy the same linear differential equation and have the same initial values. It follows that $y = z$ and hence $u = v$ i.e., u is T-periodic. If we set $\tau = 0$ in (15), then the relation $u(0) = u(T)$ leads to the equation

$$y_0 \left(\mathrm{e}^{B(T)} - 1\right) = \int_0^T c(x)\mathrm{e}^{B(s)}\, ds > 0,$$

which can be solved uniquely for y_0 because $\mathrm{e}^{B(T)} > 1$. ∎

In the classical case (b, c constant), the constant solution $u = b/c$ is distinctive. It is the only solution for which both the limits as $t \to \infty$ and as $t \to -\infty$ are positive. Moreover, as $t \to \infty$, all positive solutions tend to this solution; cf 1.XIII.(a). There also exists a distinctive solution in the general case. To investigate it we introduce a new concept. We call a function $g : \mathbb{R} \to \mathbb{R}$ *positively bounded* if there are two positive constants α, β, such that $\alpha < g(t) < \beta$ for $t \in \mathbb{R}$. Clearly, if g_1, g_2 are positively bounded, then so are $g_1 g_2$, $g_1 + g_2$, g_1/g_2.

Theorem 3. *Let the coefficients b, c be positively bounded. Then equation* (13) *has exactly one positively bounded solution u^* on $\mathbb{R}$; and if u is any positive solution, then $u(t) - u^*(t) \to 0$ as $t \to \infty$.*

Proof. Let $\alpha, \beta, \gamma, \delta$ be positive constants with $\alpha < b < \beta$, $\gamma < c/b < \delta$ in $\mathbb{R}$. The first set of these inequalities leads to the estimates

$$\alpha t < B(t) < \beta t \ \text{ for } \ t > 0, \quad \alpha t > B(t) > \beta t \ \text{ for } \ t < 0;$$

and the second leads to

$$I(t) := \int_{-\infty}^t c(s)\mathrm{e}^{B(s)}\, ds < \delta \int_{-\infty}^t b(s)\mathrm{e}^{B(s)}\, ds = \delta \mathrm{e}^{B(s)}\Big|_{-\infty}^t = \delta \mathrm{e}^{B(t)},$$

and similarly $I(t) > \gamma \mathrm{e}^{B(t)}$.

We have to show that the linear equation (15) for $y = 1/u$ has one and only one positively bounded solution. Let y^* be the solution (16) with $y_0 = I(0)$ and $\tau = 0$, that is,

$$y^*(t) = \mathrm{e}^{-B(t)} \int_{-\infty}^t c(s)\mathrm{e}^{B(s)}\, ds.$$

(This, by the way, is the smallest positive solution that exists in all of $\mathbb{R}$; cf. (a).) From the previous estimates it follows that $\gamma < y^* < \delta$. Since the solution $z(t) = \mathrm{e}^{-B(t)}$ of the homogeneous equation is unbounded and all solutions of the nonhomogeneous equation are given by $y = y^* + \lambda z$, it follows that y^* is the only positively bounded solution. ∎

Exercise. Prove the last assertion in Theorem 3.

§ 3. Differential Equations for Families of Curves. Exact Equations

I. The Differential Equation for a Family of Curves in Various Forms. If $f(x,y)$ is defined and continuous in a domain D (open connected set), then the solutions to the differential equation $y' = f(x,y)$ form a family of curves that covers D (that is the geometric meaning of the Peano existence theorem, which will be proved in § 7).

Conversely, if a given family of curves covers D simply, then it is possible to find a first order differential equation such that the curves in the family are the solutions of the differential equation. For proof, suppose $(\bar{x}, \bar{y}) \in D$ is an arbitrary point and $y = \phi(x)$ is the curve in the family that goes through this point. If we define the function f by setting $f(\bar{x}, \bar{y}) = \phi'(\bar{x})$, then clearly each curve in the family is a solution of the differential equation $y' = f(x,y)$ (this procedure was already mentioned in 1.III).

Example. The family of concentric circles

$$x^2 + y^2 = r^2 \quad (r > 0)$$

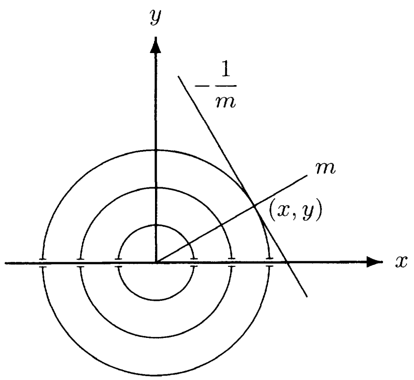
satisfies the differential equation

$$y' + \frac{x}{y} = 0, \tag{1}$$

since the slope of the line passing through the origin and the point (x,y) is $m = y/x$, and the line perpendicular to it (which is tangent to the circle) has slope $-1/m$. Technically speaking, the functions

$$y(x;r) = \pm\sqrt{r^2 - x^2} \quad (r > 0 \quad \text{is a parameter})$$

and not the circles are solutions to the differential equation (1). Moreover, the equation holds only in the open interval $-r < x < r$ because the derivative is infinite at the points $x = \pm r$. Similar problems occur whenever curves with infinite derivatives (and hence vertical line elements) are present.



Such difficulties can be overcome by representing the curves in a symmetric form, that is, either in implicit form $F(x, y) \equiv C$ or in parametric form $x = x(t)$, $y = y(t)$. A symmetric representation of a first order differential equation is given by

$$g(x, y)dx + h(x, y)dy = 0, \tag{2a}$$

or, equivalently, by

$$g(x, y)\dot{x} + h(x, y)\dot{y} = 0 \quad \text{for} \quad x = x(t),\ y = y(t), \tag{2b}$$

where $\dot{x} = dx(t)/dt$, $\dot{y} = dy(t)/dt$. Here it makes sense to assume that

$$g^2 + h^2 > 0 \tag{3}$$

(note that $g \equiv h \equiv 0$ in a domain D implies that *every* differentiable curve lying in D is a solution of (2)). Further, we require solutions in parametric form $x(t)$, $y(t)$ to be continuously differentiable and satisfy

$$(\dot{x}(t))^2 + (\dot{y}(t))^2 > 0, \tag{4}$$

which implies that the curve is a smooth curve. This assumption is also a natural one. It excludes solutions of the form $x(t) = \text{const.}$, $y(t) = \text{const.}$, and it guarantees that locally (i.e., in a neighborhood of each point of the curve) the curve can be written explicitly in the form $y = \phi(x)$ or $x = \psi(y)$ with ϕ, ψ in C^1.

The differential equation (2b) is equivalent to

$$g(x, y) + h(x, y)y' = 0 \quad \text{for} \quad y = y(x) \quad \left(y' = \frac{dy(x)}{dx}\right) \tag{2c}$$

in the following sense: If $y(x)$ is a solution of (2c), then this explicit representation can be interpreted as a parametric representation $x = t$, $y = y(t)$, for which (2b) holds. Conversely, if $x = \phi(t)$, $y = \psi(t)$ is a solution of (2b) and if $\dot{\phi}(t_0) \neq 0$, then $\dot{\phi} \neq 0$ in a neighborhood U of t_0, and the inverse function $t = t(x)$ exists. The part of the curve that corresponds to values of $t \in U$ can be expressed explicitly in the form

$$y = y(x) = \psi(t(x)).$$

The function $y(x)$ is a solution of (2c) because

$$y'(x) = \dot{\psi}(t(x))\frac{dt(x)}{dx} = \frac{\dot{\psi}(t(x))}{\dot{\phi}(t(x))}.$$

In a like manner, if $\dot{\psi}(t_0) \neq 0$, then (2b) is equivalent to the differential equation

$$g(x, y)\frac{dx}{dy} + h(x, y) = 0 \quad \text{for} \quad x = x(y). \tag{2d}$$

The case $\dot{\phi}(t_0) = \dot{\psi}(t_0) = 0$ is excluded by the requirement (4).

Summary. The solutions obtained from equations (2b), (2c), and (2d) are indeed different functions, but they give exactly the same curves with the exception that for (2c) those curve points with vertical tangents and for (2d) those curve points with horizontal tangents are missing. Consequently, whenever the emphasis is on the geometric point of view, that is, when one is interested in solution *curves*, there is no real difference between the four forms of the equation (2a), (2b), (2c), (2d).

Finally, we note that equation (2b) is invariant with respect to a change of parameter: If $x(t)$, $y(t)$ is a solution of (2b), then so is $\bar{x}(\tau) = x(h(\tau))$, $\bar{y}(\tau) = y(h(\tau))$, as long as $h(\tau) \in C^1$.

In Example 1, the symmetric form of differential equation (1) is

$$xdx + ydx = 0.$$

II. Exact Differential Equations. A differential equation of the form (2) is called an *exact* equation in the domain D if (g, h) is a gradient field, i.e., if there exists a function $F(x, y) \in C^1(D)$ such that

$$F_x(x, y) = g(x, y),\ F_y(x, y) = h(x, y) \quad \text{in} \quad D. \tag{5}$$

The function F is called a *potential function* for the field (g, h).

The *total differential* of a function F is defined as $dF = F_x dx + F_y dy$. Thus a differential equation is exact in D if and only if it can be represented in the form

$$dF(x, y) = 0 \quad \text{with} \quad F \in C^1(D). \tag{6}$$

Once a potential function has been determined, the problem of integrating the differential equation (2) is essentially settled.

Theorem. *Let the functions g, h be continuous in the domain D. If the differential equation* (2) *is exact in D and if $F \in C^1(D)$ is a potential function, then the function pair $(x(t), y(t)) \in C^1(J)$ (with values in D) is a solution of the differential equation* (2b) *if and only if $F(x(t), y(t))$ is constant in the interval J. Likewise, $y(x)$ is a solution of* (2c) *if and only if $F(x, y(x))$ is constant, and a corresponding statement holds for* (2d).

Additionally, if (3) *holds, then by solving*

$$F(x, y) = \alpha \tag{7}$$

one obtains all solution curves, and exactly one solution curve passes through each point of D.

The *proof* follows from the identity

$$g \cdot \dot{x} + h \cdot \dot{y} = F_x \dot{x} + F_y \dot{y} = \frac{d}{dx} F(x(t), y(t)).$$

Thus the pair $(x(t), y(t))$ is a solution of (2b) if and only if $F(x(t), y(t))$ is constant.

The second part of the conclusion is a consequence of the implicit function theorem. Note that by (3) and (5), $F_x^2 + F_y^2 > 0$ in D. Let $(\xi, \eta) \in D$ and $F(\xi, \eta) = \alpha$. If $F_y(\xi, \eta) \neq 0$, for instance, then by the implicit function theorem $F(x, y) = \alpha$ has a unique solution of the form $y = y(x) \in C^1$ in a neighborhood U of the point (ξ, η), and differentiation of the identity $F(x, y(x)) = \alpha$ gives equation (2c), i.e., y is a solution. ∎

Example.

$$(y^2 e^{xy} + 3x^2 y)dx + (x^3 + (1 + xy)e^{xy})dy = 0$$

is exact in $\mathbb{R}^2$. A potential function is

$$F(x, y) = y(e^{xy} + x^3).$$

The question whether a differential equation is exact and if it is, how to find a potential function is answered in the following well-known result from analysis.

III. Theorem on Potential Functions. *If $g(x, y)$, $h(x, y)$ are continuously differentiable in the simply connected*[5] *domain D, then there exists a potential function $F(x, y)$ satisfying* (5) *if and only if*

$$g_y \equiv h_x \quad \text{in} \quad D \tag{8}$$

holds.

The potential function is obtained as a line integral

$$F(\bar{x}, \bar{y}) = \int_{(\xi,\eta)}^{(\bar{x},\bar{y})} \{g(x, y)\, dx + h(x, y)\, dy\},$$

where $(\xi, \eta) \in D$ is a fixed point and the integration is carried out along an arbitrary C^1-path connecting (ξ, η) to $(\bar{x}, \bar{y})$. Equation (8) is precisely the condition required to guarantee that this integral is independent of the path.

IV. Integrating Factors (or Euler Multipliers). The differential equation

$$y\, dx + 2x\, dy = 0 \tag{9}$$

is not exact. However, it can easily be made an exact differential equation (in the domain $x > 0$) by multiplying the equation by $1/\sqrt{x}$. The resulting differential equation

$$\frac{y}{\sqrt{x}}\, dx + 2\sqrt{x}\, dy = 0$$

is exact, and a potential function is given by

$$F(x, y) = 2y\sqrt{x} \quad (x > 0).$$

An exact differential equation can also be obtained by multiplying (9) by y:

$$y^2\, dx + 2xy\, dy = 0, \quad \text{giving} \quad F(x, y) = xy^2.$$

[5] See A.VI for a definition of simple connectedness.

V. Definition and Theorem. *If the functions $g(x,y)$, $h(x,y)$ are continuous in D, then a continuous function $M(x,y) \neq 0$ defined in D is called an* integrating factor *or* Euler multiplier *for the differential equation* (2) *if the differential equation*

$$M(x,y)g(x,y)\,dx + M(x,y)h(x,y)\,dy = 0 \tag{10}$$

is exact.

If D is simply connected and $g, h, M \in C^1(D)$, then $(Mg)_y = (Mh)_x$, i.e.,

$$M_y g + M g_y = M_x h + M h_x \tag{11}$$

is necessary and sufficient for M to be a multiplier.

This follows immediately from Theorem III. Note that in general it is a difficult task to find an integrating factor M, since M is the solution to the *partial* differential equation (11). However, once a multiplier M is found, all solutions of equation (2) (which is equivalent to (10)) can be found by integration; cf. Theorems II and III.

Multipliers Depending on Only One Variable. Sometimes a multiplier can be found that depends only on x (or only on y). The ansatz $M = M(x)$, for instance, leads to

$$\frac{g_y - h_x}{h} = \frac{M'}{M} = (\log M)'. \tag{12}$$

Thus an integrating factor depending only on x exists if and only if the left-hand side of (12) depends only on x. An important example is the linear differential equation; see Exercise VIII.(e).

Example.

$$(2x^2 + 2xy^2 + 1)y\dot{x} + (3y^2 + x)\dot{y} = 0.$$

The differential equation is not exact; however (cf. (12)),

$$\frac{g_y - h_x}{h} = 2x,$$

and hence $M = \mathrm{e}^{x^2}$ is an integrating factor. A potential function $F(x,y)$ can be determined from equations (5), which for this example are given by

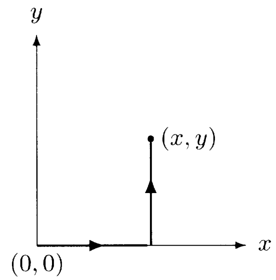
$$F_x = \mathrm{e}^{x^2} y(2x^2 + 2xy^2 + 1), \quad F_y = \mathrm{e}^{x^2}(3y^2 + x).$$

From the second of these equations it follows immediately that

$$F(x,y) = \mathrm{e}^{x^2}(y^3 + xy) + \phi(x),$$

where $\phi(x)$ is an arbitrary function of x. This function $\phi(x)$ must be determined such that the first of the two equations also holds. This is the case, as one can check, if $\phi \equiv 0$. Thus the solutions are given by

$$F(x,y) \equiv y\mathrm{e}^{x^2}(x + y^2) = C.$$



Note that it is just about as easy to get the potential function from the line integral in Theorem III. Choosing $(\xi, \eta) = (0, 0)$ and taking the path given in the figure, one sees that the integral vanishes along the x-axis, since $g = 0$ there and $dy = 0$, and that the integral over the vertical segment gives precisely the function F obtained earlier.

VI. A System of Two Autonomous Differential Equations. In connection with equation (2), we consider the system of two differential equations

$$\dot{x} = h(x, y), \quad \dot{y} = -g(x, y) \tag{13}$$

for the pair of functions $(x(t), y(t))$. Such a system is called an *autonomous* system, because the variable t does not appear explicitly in the right-hand side. A consequence of this fact is the following property:

(a) If $(x(t), y(t))$ is a solution, then so is $(x(t + c), y(t + c))$ (c arbitrary).

Phase Plane and Phase Portraits. A solution $(x(t), y(t))$ of the system (13) can also be interpreted as the parametric representation of the corresponding solution curve in the xy-plane. In this context the xy-plane is also referred to as the *phase plane*, and the curves generated by the solutions are called the *trajectories* (or the *orbits*) of the differential equation. A sketch of several of these trajectories is called a *phase portrait* or *phase diagram* of the system (13). Arrows are added to the trajectories to give the orientation of the curve in the sense of increasing t. In addition, the velocity of motion, that is, the vector $(\dot{x}(t), \dot{y}(t))$, can also be given approximately by placing special dots along the trajectory at solution points corresponding to a sequence of *equidistant* t-values. Where the dots are close together, the solution is changing slowly, where they are far apart, the change is correspondingly faster.

A phase portrait gives an excellent overview of the qualitative behavior of the solutions. Thus it is of great significance that in some examples the trajectories can be given without first determining the solutions. What is needed is a function $F(x, y)$ that is constant along each solution (i.e., on each trajectory). Such a function is called a *first integral* of the system (13). The trajectories are then given implicitly by the equation $F(x, y) = \alpha$.

There is a relationship between equations (2) and (13). A solution of (13) is clearly a solution of (2b) and also, more generally, of the equation

$$M(x,y)g(x,y)\dot{x} + M(x,y)h(x,y)\dot{y} = 0. \tag{14}$$

If $M \neq 0$ is an integrating factor that makes this equation exact, then there exists a potential function F with the desired property. To make this precise:

(b) *If equation* (14) *is exact and F is a potential function, that is, if* $\operatorname{grad} F = (Mg, Mh)$, *then F is constant along solutions of the system* (13). *The trajectories of the system* (13) *are obtained as the level sets*

$$K_\alpha = F^{-1}(\alpha) = \{(x,y) \in D : F(x,y) = \alpha\}.$$

Example. For the system $\dot{x} = y$, $\dot{y} = -x$, the corresponding equation (2) is $x\,dx + y\,dy = 0$. The function $F(x,y) = x^2 + y^2$ is constant along each solution, and therefore the trajectories are circles centered at the origin.

Some questions arise at this point.

(c) If the level sets K_α are curves, is it possible to formulate general theorems about their structure? Locally, if $\operatorname{grad} F \neq 0$, then K_α is a curve because of the implicit function theorem. Global statements, particularly statements about closed Jordan curves, are proved in the Appendix in sections A.VII–VIII.

(d) Does a solution that starts on a level curve always trace out the entire curve, or can it just stop somewhere? Statements related to this question are proved in A.IX.

(e) How can the direction of the arrows on the trajectories be obtained? This can usually be done without difficulty by considering the sign of g and h. The markings of points for equidistant t-values, on the other hand, have to be obtained numerically.

As an illustration of these ideas, we will consider a famous example from biomathematics.

VII. The Predator–Prey Model of Lotka–Volterra. We consider an ecological model consisting of two species, a predator species and a prey species, which goes back to the American biophysicist Alfred J. Lotka (1880–1949) and the Italian mathematician Vito Volterra (1860–1940). The size of the predator population will be denoted by $y(t)$, that of the prey by $x(t)$. In the system of differential equations

$$\dot{x} = x(a - by), \quad \dot{y} = y(-c + dx), \tag{15}$$

which describes their interaction, a, b, c, d are positive constants. The prey population is assumed to have ample resources (e.g., food) so that in the absence of predators ($y = 0$)

$$\text{growth rate} = \text{birth rate} - \text{death rate} = a > 0,$$

and the population increases in size according to the exponential growth equation $\dot{x} = ax$. In the presence of predators the growth rate reduces from a to $a - by$

for obvious reasons; in fact, it can become negative. The situation is different for the predator population. Without prey ($x = 0$), the predator population decreases in accordance with the equation for exponential decay $\dot{y} = -cy$ (because without an adequate food supply, the death rate exceeds the birth rate), but when the prey population is present, the improvement in the food supply enlarges the growth rate to $-c+dy$, which can be positive if the prey population is large.

The following analysis of equation (15) begins with an application of the existence and uniqueness theorem 10.VI. From this theorem it follows that for every initial condition $(x(0), y(0)) = (\xi, \eta)$, there is exactly one solution of (15). Clearly, there is exactly one positive *constant*, or *stationary, solution* $(x(t), y(t)) = (x_0, y_0) = (c/d, a/b)$.

In the notation of (13), $g = y(c - dx)$ and $h = x(a - by)$. We claim that $M(x, y) = \dfrac{1}{xy}$ is an integrating factor. Indeed, the functions

$$\bar{g} = Mg = \frac{c}{x} - d, \quad \bar{h} = \frac{a}{y} - b$$

satisfy the condition $\bar{g}_y = \bar{h}_x = 0$ from Theorem III. A potential function is easily found:

$$F = G(x) + H(y) \quad \text{with} \quad G(x) = c \log x - dx, \ H(y) = a \log y - by.$$

The function G is strongly monotone increasing in the interval $(0, x_0)$, strongly decreasing in the interval $[x_0, \infty)$, and it tends to $-\infty$ as $x \to 0+$ and as $x \to \infty$; H behaves in the same manner on the intervals $(0, y_0)$, $[y_0, \infty)$. It follows that F has a maximum at (x_0, y_0),

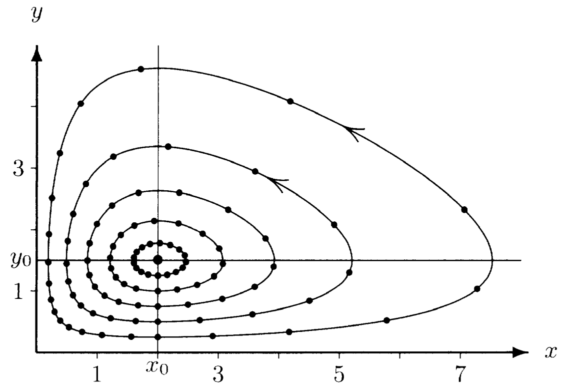
$$F(x, y) < F(x_0, y_0) =: B \quad \text{for} \quad x, y > 0, \quad (x, y) \neq (x_0, y_0),$$

and $\operatorname{grad} F \neq 0$ there. From Theorem II and Theorems A.VIII (with $A = -\infty$) and A.IX in the Appendix, we immediately obtain the following

Theorem. *For $-\infty < \alpha < B$, the level sets $K_\alpha = F^{-1}(\alpha)$ are closed Jordan curves that surround the stationary point (x_0, y_0). All positive solutions $(x(t), y(t))$ of the Lotka–Volterra equations are periodic; $x(t)$ has its largest and smallest values when $y(t) = y_0$, and $y(t)$ has its largest and smallest values when $x(t) = x_0$.*

We briefly describe the evolution of a solution $(x(t), y(t))$ with initial value (x_0, η), $\eta < y_0$, at $t = 0$. The solution traces out the curve K_α, where $\alpha = F(x_0, \eta)$. When $t = 0$, the predator population y is at its lowest point; as t increases, y begins to increase, while at the same time the growth of the prey population x slows and comes to a halt at the value $y = y_0$. At this point x begins to decrease and y continues to grow, but more slowly, reaching a maximum when x sinks to the value x_0, etc.

It can be shown directly, without appealing to A.IX, that the above solution actually makes a complete rotation in finite time and does not just come to a halt



The predator–prey model with $(a, b, c, d) = (3, 2, 2, 1)$, $(x_0, y_0) = (2, \frac{3}{2})$

somewhere on the curve. Since the solution values $(x(t), y(t))$ lie on a bounded curve K_α, the solution itself is bounded; thus it exists for all $t \geq 0$ by Theorem 10.VI. As long as $y(t) < y_0$, $\dot{x}(t)$ is positive, whence $x(t) > x_0$ for small positive t. But then from $x \geq x_0 + \varepsilon$ it follows that $\dot{y} \geq y(-c + d(x_0 + \varepsilon)) = d\varepsilon y > 0$. Thus $y > \eta$, and hence $\dot{y} > d\varepsilon\eta > 0$. Therefore, y takes on the value y_0 at a time t_1, where $x(t_1) > x_0$. In a similar manner, one shows that the remaining three parts of the curve K_α are traced out in this way and that there exists a smallest positive number T with $(x(T), y(T)) = (x(0), y(0))$. By V.(a) the function $(\bar{x}(t), \bar{y}(t)) = (x(t+T), y(t+T))$ is also a solution of the differential equation with the same initial conditions at $t = 0$ as $(x(t), y(t))$, and by the uniqueness theorem both solutions must be identical, i.e., the solution under consideration is periodic with period T.

(a) *Exercise.* Consider the mean value (x_m, y_m) of a T-periodic solution $(x(t), y(t))$ of equation (15); i.e.,

$$x_m = \frac{1}{T}\int_0^T x(t)\,dt, \quad y_m = \frac{1}{T}\int_0^T y(t)\,dt.$$

Show that $x_m = x_0$, $y_m = y_0$. Thus the mean value of a solution over a period is equal to the value of the stationary solution.

Hint: Integrate $\dot{x}/x$ and $\dot{y}/y$ from 0 to T.

VIII. Generalized Predator–Prey Models. Exercise. (a) Show that the same qualitative statements hold for the nonnegative solutions of the system

$$\dot{x} = x(a - by^2), \quad \dot{y} = y(-c + dx^2)$$

as for the Lotka–Volterra model; cf. Theorem VI. Is VI.(a) still valid?

The statements of Theorem VI can be essentially carried over to more general systems of the form

$$\dot{x} = \phi(x)\alpha(y), \quad \dot{y} = -\psi(y)\beta(x) \tag{16}$$

and, in fact, can be carried a step further.

(b) We consider an autonomous system of the form

$$\dot{x} = W(x,y)\bar{h}(y), \quad \dot{y} = -W(x,y)\bar{g}(y) \tag{17}$$

with $W > 0$ (in the case of equation (16), $W = \phi(x)\psi(y)$). Let the functions $\bar{g}$, $\bar{h}$ be continuous and strongly monotone decreasing in $[0,\infty)$ and let each function have a positive zero, say $\bar{g}(x_0) = 0$, $\bar{h}(y_0) = 0$. Show directly that:

(i) The function $F(x,y) = G(x) + H(y)$ with

$$G(x) = \int_{x_0}^{x} \bar{g}(s)\,ds, \quad H(y) = \int_{y_0}^{y} \bar{h}(s)\,ds$$

is constant along each solution of the system (17).

(ii) If it is assumed that $G(0+) = H(0+) = -\infty$, then the statements of Theorem VI are valid. In particular, all positive solutions are periodic.

(iii) Using $\bar{g}(x) = 2(1-x)$, $\bar{h}(y) = 2(1-y)$, $W = 1$ as an example, discuss how the behavior of the solutions changes if the hypothesis in (ii) is violated.

IX. Exercises. (a) Determine all solutions of the differential equation

$$(\cos(x+y^2) + 3y)\,dx + (2y\cos(x+y^2) + 3x)\,dy = 0$$

in implicit form. Discuss (and sketch) the solution through the origin.

(b) Determine all solutions of the differential equation

$$(xy^2 - y^3)\,dx + (1 - xy^2)\,dy = 0.$$

(There is an integrating factor $M = M(y)$.) Sketch the direction field and draw some solution curves (with the help of the isoclines for the slopes 0, 1, -1, ∞, for instance). Determine the solution through the origin.

(c) Determine all solutions of the differential equation

$$y(1+xy)\,dx = x\,dy$$

in explicit form. There exists an integrating factor $M = M(y)$.

(d) Derive a differential equation for the family of circles

$$(x-\lambda)^2 + y^2 = \lambda^2 \quad (\lambda > 0)$$

and draw a sketch showing some of the solutions.

(e) Find an integrating factor $M = M(x)$ for the linear equation

$$y' + p(x)y = q(x)$$

and find the associated potential function. Compare the solutions obtained from $F(x,y) = \alpha$ with Theorem 2.II.

§ 4. Implicit First Order Differential Equations

We consider the implicit differential equation

$$F(x, y, y') = 0. \tag{1}$$

Throughout this section, we assume that the function $F(x, y, p)$ is continuous in a domain D of three-dimensional space (without making specific reference to the fact each time).

Just as in the case of explicit differential equations, the differential equation (1) defines a direction field. It is the set of all line elements (x, y, p), for which

$$F(x, y, p) = 0. \tag{2}$$

The new feature in the case of implicit differential equations is that now a point $(\bar{x}, \bar{y})$ can be a "carrier" of more than one line element, in contrast to the situation for explicit differential equations. This happens whenever the equation $F(\bar{x}, \bar{y}, p) = 0$ has more than one solution p.

I. Regular and Singular Line Elements. If $F(\bar{x}, \bar{y}, \bar{p}) = 0$ and if equation (2) can be rewritten in a neighborhood $U \subset \mathbb{R}^3$ of the point $(\bar{x}, \bar{y}, \bar{p})$ in a unique way in the form

$$p = f(x, y) \quad \text{with a continuous} \quad f(x, y) \quad ((x, y) \in V(\bar{x}, \bar{y})) \tag{3}$$

(this means that the line elements in U are precisely the triples $(x, y, f(x, y))$ with $(x, y) \in V$), then $(\bar{x}, \bar{y}, \bar{p})$ is called a *regular* line element. All line elements that are not regular are called *singular*. A solution curve $y(x)$ of (1) is called *regular*, respectively *singular*, if all of the line elements $(x, y(x), y'(x))$ are regular, respectively singular. Finally, (x, y) is a *singular point* of the differential equation if there exists a singular line element (x, y, p).

Theorem. *If the functions $F(x, y, p)$ and $F_p(x, y, p)$ are continuous in a neighborhood of $(\bar{x}, \bar{y}, \bar{z})$ and if*

$$F(\bar{x}, \bar{y}, \bar{p}) = 0, \quad F_p(\bar{x}, \bar{y}, \bar{p}) \neq 0, \tag{4}$$

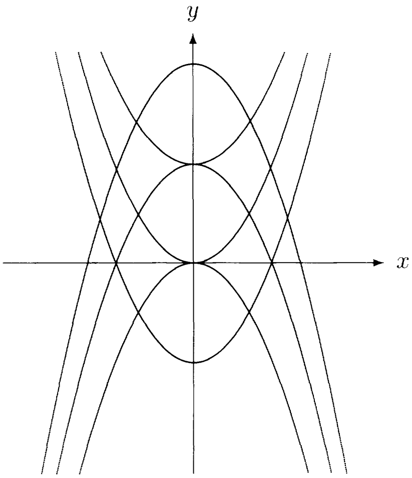
then $(\bar{x}, \bar{y}, \bar{p})$ is a regular line element.

The implicit function theorem applies under the hypotheses of this theorem and implies that a representation in the form (3) is possible.

It follows that the conditions

$$F(\bar{x}, \bar{y}, \bar{p}) = F_p(\bar{x}, \bar{y}, \bar{p}) = 0 \tag{5}$$

must hold for singular line elements $(\bar{x}, \bar{y}, \bar{p})$. Note, however, that a line element that satisfies property (5) is not necessarily singular. For instance, every line element of $F(x, y, p) = [p - f(x, y)]^2$ is regular (for continuous f), since the equation can be uniquely expressed in the form (3), even though (5) holds for all line elements.



Solution curves of the differential equation ${y'}^2 = 4x^2$

Example.

$${y'}^2 = 4x^2.$$

The equation $F(x, y, p) = p^2 - 4x^2 = 0$ is equivalent to $p = \pm 2x$, and thus the line elements are the triples $(x, y, \pm 2x)$; the solutions of the differential equation are the parabolas $y = C + x^2$ and $y = C - x^2$. The only place where (2) cannot be written (locally uniquely) in the form (3) is the y-axis; i.e., the line elements $(0, y, 0)$ are singular, as are the points $(0, y)$.

II. Parametric Representation with y' as the Parameter. In the following sections we will discuss some examples of implicit differential equations that can be solved in closed form. The ansätze that are used here all have the common property that they lead to solution curves with a special parametric representation in which $p = y'$ is the parameter.

This will now be explained. Consider pairs $(x(p), y(p))$ of continuously differentiable functions in an interval J with the property

$$\dot{y}(p) = p \cdot \dot{x}(p). \tag{6}$$

Here $\dot{x} = dx/dp$, $\dot{y} = dy/dp$. If $\dot{x}(p) \neq 0$, then the curve represented by this pair has slope p at the point $(x(p), y(p))$. Indeed, it is well known that the curve can be represented in explicit form $y = \phi(x)$, and moreover, from $y(p) = \phi(x(p))$ and (6), it follows that

$$\phi'(x(p)) = \frac{\dot{y}(p)}{\dot{x}(p)} = p. \tag{7}$$

Conversely, suppose $y = \phi(x)$ is an arbitrary curve in explicit form. A parametric representation that satisfies condition (6) can be derived by solving the equation $p = \phi'(x)$ for x. If we denote the inverse function of ϕ' by $x(p)$ and set $y(p) := \phi(x(p))$, then $(x(p), y(p))$ is a parametric representation of the curve that satisfies the relation (6),

$$\dot{y}(p) = \phi(x(p)) \cdot \dot{x}(p) = p \cdot \dot{x}(p),$$

as was expected.

Two Examples. (a) $y = x^3$ for $x \in \mathbb{R}$. From the relation $p = 3x^2 \geq 0$, we obtain, using the procedure described above, that

$$\begin{aligned} x &= \xi(p) = \pm\sqrt{p/3} \\ y &= \eta(p) = \xi^3(p) \end{aligned} \qquad (p \geq 0).$$

Thus each of the two branches $x \geq 0$, $x \leq 0$ of the cubic parabola has a parametric representation with $y' = p$ as parameter that satisfies the condition (6).

(b) $y = \sin x$, $0 \leq x \leq \pi$. From $p = \cos x$ it follows that

$$\begin{aligned} x &= \arccos p \\ y &= \sqrt{1 - p^2} \end{aligned} \qquad (-1 \leq p \leq 1).$$

What is the corresponding representation in the interval $\pi \leq x \leq 2\pi$?

Such a representation of a curve is possible only if $p = \phi'(x)$ can be solved for x (as is the case if $\phi'' \neq 0$). In particular, straight lines cannot be represented parametrically in this manner. We will have to be on the alert for this situation later on in the discussion.

The outline of a solution procedure based on the above ideas goes something like this. An implicit differential equation in the form (1) is given. Denote the solution by $\phi(x)$. If the solution curve has a parametric representation $(x(p), y(p))$ with property (6) (this would be the case if $\phi'' \neq 0$!), then the substitution $x = x(p)$ into $F(x, \phi(x), \phi'(x)) \equiv 0$ gives the equation

$$F(x(p), y(p), p) = 0 \tag{8}$$

because of (7). The functions $x(p)$, $y(p)$ can now be determined using the two equations (6) and (8).

The following types of equations can be solved using this procedure.

III. $\boxed{x = g(y')}$

Let J be an interval and $g \in C^1(J)$. Here $x(p) = g(p)$ is given and $y(p)$ is obtained from (6). Thus the solution curves are given by

$$\left\{ \begin{aligned} x(p) &= g(p), \\ y(p) &= C + \int p\dot{g}(p)\, dp. \end{aligned} \right.$$

Clearly, the set of solutions does not include line segments.

IV. $\boxed{y = g(y')}$

Let $g \in C^1(J)$. Similar to the result obtained in III, it follows from (6) that

$$\begin{cases} y(p) = g(p), \\ x(p) = C + \displaystyle\int \frac{\dot{g}(p)}{p}\, dp. \end{cases}$$

In addition, the constant function $y = g(0)$ is also a solution, provided that $0 \in J$.

V. $\boxed{y = xy' + g(y')}$ **Clairaut's Differential Equation.**[6]

Let $g \in C^1(J)$. Differentiating $y(p) = x(p)p + g(p)$ gives the relation

$$\dot{y} = p\dot{x} + x + \dot{g},$$

and hence, using (6), we have $x + \dot{g} = 0$; i.e.,

$$\begin{cases} x(p) = -\dot{g}(p), \\ y(p) = -p\dot{g}(p) + g(p). \end{cases} \tag{9}$$

However, this is only one solution. It is easy to check that the straight lines

$$y = cx + g(c) \quad (c \in J) \tag{10}$$

are also solutions. It is also not difficult to verify that the curve (9) touches each of the lines (10) at the point $(x(c), y(c))$ corresponding to the parameter value $p = c$ and that both have the same slope $m = c$ at this point. The lines (10) form a set of tangents to the curve (9), and the curve (9) is called the *envelope* of the family of lines (10).

What about the conditions that must be met before (9) defines a solution? If $g \in C^2(J)$, then clearly $x, y \in C^1(J)$. Further, if $\ddot{g} \neq 0$, then $\dot{x} \neq 0$ in J, i.e., the curve (9) can be explicitly represented in the form $y = \phi(x)$ with a continuously differentiable ϕ. If one notes that $\phi(x(p)) = y(p)$, then the proof that ϕ is actually a solution follows from the second line of (9). It can be further proved under these assumptions that all of the solutions to the Clairaut equation have been found, i.e., that every solution is either the function ϕ obtained from (9), one of the lines (10), or a function constructed by splicing ϕ to one of the lines (10). The proof (which is not exactly short) can be found in Kamke (Differentialgleichungen, Vol. I, pp. 52–54).

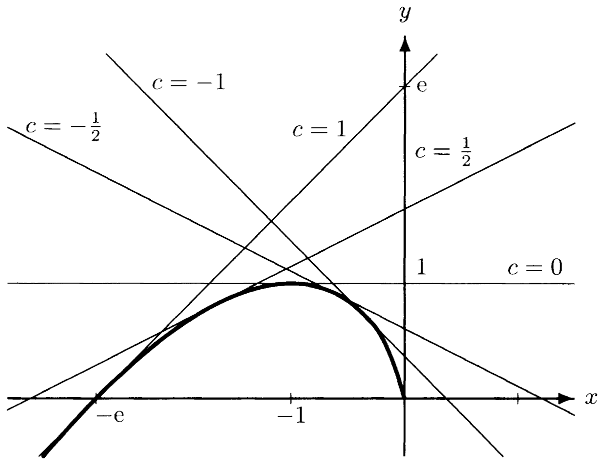
Example.

$$y = xy' + e^{y'}$$

The solution curves are the lines

$$y = cx + e^c \quad (c \in \mathbb{R})$$

[6]Clairaut, Alexis Claude, 1713–1765, French mathematician and astronomer.



Solution curves of Clairaut's equation $y = xy' + e^{y'}$

and their envelope

$$\left.\begin{array}{l} x = -e^p \\ y = (1-p)e^p \quad (p \in \mathbb{R}) \end{array}\right\} \Leftrightarrow y = x\{\log(-x) - 1\} \quad (x < 0).$$

VI. $\boxed{y = xf(y') + g(y')}$ **D'Alembert's Differential Equation.**[7]

Let $f, g \in C^1(J)$. Differentiating $y(p) = xf(p) + g(p)$ to get

$$\dot{y} = \dot{x}f + x\dot{f} + \dot{g}$$

and then using (6) leads to the linear differential equation

$$\dot{x} = \frac{x\dot{f}(p) + \dot{g}(p)}{p - f(p)},$$

from which $x(p)$ and hence also $y(p) = xf(p) + g(p)$ can be determined in closed form. A line $y = cx + d$ is a solution if and only if $f(c) = c$ and $d = g(c)$.

Example. Let

$$y = x\left(y' + \frac{1}{y'}\right) + y'^4.$$

Differentiation of $y(p) = x(p)(p + 1/p) + p^4$ leads to

$$\dot{y} = \dot{x}\left(p + \frac{1}{p}\right) + x(1 - p^{-2}) + 4p^3,$$

or, since $\dot{y} = p\dot{x}$,

$$\dot{x} = x\left(\frac{1}{p} - p\right) - 4p^4.$$

[7] D'Alembert, Jean Le Rond, 1717–1783, French mathematician, philosopher, and writer.

A solution to the homogeneous equation is $x(p) = e^{\log p - p^2/2} = p \cdot e^{-p^2/2}$, a solution to the inhomogeneous equation

$$\tilde{x}(p) = pe^{-p^2/2} \cdot (-4) \int e^{p^2/2} p^3 \, dp.$$

The substitution $s = p^2/2$, $ds = p\,dp$, transforms the integral into $2\int se^s \, ds = 2e^s(s-1)$, which gives $\tilde{x}(p) = 8p - 4p^3$. Thus the general solution reads

$$x(p) = Cpe^{-p^2/2} + 8p - 4p^3, \quad y(p) = x(p)\left(p + \frac{1}{p}\right) + p^4 \quad (C \in \mathbb{R}).$$

VII. Integration by Differentiation. This heading refers to the following procedure for solving a differential equation $F(x, y, y') = 0$. If there is a solution of the form $(x(p), y(p))$, then $F(x(p), y(p), p) = 0$ holds, and the equation

$$F_x \cdot \dot{x} + F_y \cdot \dot{y} + F_p = 0$$

follows by differentiation. Using this relation together with (6) gives the following equations for the functions $\dot{x}$, $\dot{y}$:

$$\dot{x} = -\frac{F_p}{F_x + pF_y}, \quad \dot{y} = -\frac{pF_p}{F_x + pF_y}. \tag{11}$$

This is a system of two differential equations for the two unknown functions $(x(p), y(p))$. In many cases the variables are separated; i.e., only $x(p)$ [or $y(p)$] appears on the right-hand side of the first [or second] differential equation. Two examples where this kind of separation occurs are

$$\boxed{y = G(x, y')} \Rightarrow \dot{x} = \frac{G_p(x,p)}{p - G_x(x,p)}, \quad y(p) = G(x(p), p);$$

$$\boxed{x = H(y, y')} \Rightarrow \dot{y} = \frac{pH_p(y,p)}{1 - pH_y(y,p)}, \quad x(p) = H(y(p), p).$$

The types of equations discussed above in III, IV, and VI are special cases of this type of equation in which the new differential equation for $\dot{x}$, respectively $\dot{y}$, can be solved explicitly.

VIII. Exercises. Determine all the solutions for the following Clairaut differential equations in explicit form. Sketch some of the solution curves!

(a) $y = xy' - \sqrt{y' - 1}$.

(b) $y = xy' + y'^2$.

(c) Show that all solutions of the differential equation

$$y = xy' + ay' + b \quad (a,\ b \text{ constant})$$

are lines through a fixed point. Is the "envelope" given by (9) a solution?

Determine the solutions of the following differential equations in parametric form:

(d) $y = xy'^2 + \ln(y'^2)$,

(e) $x = y(y' + 1/y') + y'^5$.

Chapter II
Theory of First Order Differential Equations

§ 5. Tools from Functional Analysis

Many questions in the theory of differential equations can be answered in a particularly elegant manner using general concepts of functional analysis. In this and later chapters, functional-analytic methods will be used to derive theorems on existence, uniqueness, and dependence of solutions on parameters. We begin by introducing the concept of a Banach space.

I. Vector Spaces. A set $L = \{a, b, c, \ldots\}$ is called a *vector space* or *linear space* if an addition and a "scalar" multiplication (scalars are real or complex numbers) are defined (i.e., to each pair of elements $a, b \in L$ there is associated a unique element $a + b \in L$, and to each element $a \in L$ and each number λ, an element $\lambda a \in L$) and if these operations satisfy the following laws:

The set L is an abelian group with respect to addition, that is to say, the following rules hold for $a, b, c \in L$:

$$\begin{aligned}(a+b)+c &= a+(b+c)\\ a+b &= b+a,\end{aligned}$$

and there is a unique zero element, denoted by θ, and to each $a \in L$ a unique inverse, denoted by $-a$, such that

$$\begin{aligned}a+\theta &= a\\ a+(-a) &= \theta.\end{aligned}$$

For $a, b \in L$ and arbitrary numbers λ, μ, scalar multiplication satisfies the rules

$$\begin{aligned} \lambda(a+b) &= \lambda a + \lambda b \\ (\lambda+\mu)a &= \lambda a + \mu a \\ \lambda(\mu a) &= (\lambda\mu)a \\ 1 \cdot a &= a \\ 0 \cdot a &= \theta. \end{aligned}$$

The space is called a *real* or a *complex* vector space, depending on whether the scalars λ, μ come from the field of real or complex numbers.

A nonempty subset of L that (with the previously described operations) is again a linear space is called a (linear) *subspace* of L.

II. Normed Space. Let L be a real or complex linear space. A real-valued function $\|a\|$, defined for $a \in L$, is called a *norm* if it has the properties

$$\begin{aligned} &\|\theta\| = 0, \quad \|a\| > 0 \quad \text{for} \quad a \neq \theta && \textit{definiteness,} \\ &\|\lambda a\| = |\lambda| \cdot \|a\| && \textit{homogeneity,} \\ &\|a+b\| \leq \|a\| + \|b\| && \textit{triangle inequality.} \end{aligned}$$

The space L is sometimes said to be "normed" by $\|\cdot\|$. We have used a special symbol θ here for the zero element of L in order to avoid confusion with the number 0. From now on we shall take the commonly used symbol 0 for the zero element in any vector space; the reader should have no problem with equations like $0 \cdot a = 0$ from Section I.

For future reference we mention two simple consequences of the triangle inequality:

$$\|a_1 + \cdots + a_n\| \leq \|a_1\| + \cdots + \|a_n\|, \tag{1}$$

$$\Big| \|a\| - \|b\| \Big| \leq \|a - b\|. \tag{2}$$

Note that a norm defines a *distance function* (or *metric*) $\rho(a,b) = \|a-b\|$ that satisfies the axioms of a metric space:

$$\begin{aligned} &\rho(a,b) > 0 \;\text{ for }\; a \neq b, \;\; \rho(a,a) = 0 && \textit{positivity,} \\ &\rho(a,b) = \rho(b,a) && \textit{symmetry,} \\ &\rho(a,b) \leq \rho(a,c) + \rho(c,b) && \textit{triangle inequality.} \end{aligned}$$

Thus a normed space is a metric space. Using this "canonical" distance function and proceeding in a natural manner, we can extend the definition of familiar mathematical objects from the Euclidean space $\mathbb{R}^n$ to any normed space L: balls, ε-neighborhoods, neighborhoods, interior points, boundary points, open and closed sets... and, not least, convergence.

III. Examples. (a) *The n-dimensional Euclidean space* $\mathbb{R}^n$. This space is the set of all n-tuples of real numbers

$$\mathbf{a} = (a_1, \ldots, a_n) = (a_i).$$

Addition and scalar multiplication (λ real) are defined by

$$\mathbf{a} + \mathbf{b} = (a_i + b_i), \quad \lambda\mathbf{a} = (\lambda a_i).$$

The space $\mathbb{R}^n$ can be normed in many ways, for example, by any one of the following:

$$|\mathbf{a}|_e = \sqrt{a_1^2 + \cdots + a_n^2} \qquad \textit{Euclidean norm,}$$

$$|\mathbf{a}| = |a_1| + \cdots + |a_n|,$$

$$|\mathbf{a}| = \max_i |a_i| \qquad \textit{maximum norm.}$$

In this text, elements from $\mathbb{R}^n$ are denoted by boldface italic type and norms in $\mathbb{R}^n$ by the ordinary absolute value symbol.

(b) *The n-dimensional complex*, or *unitary, space*, $\mathbb{C}^n$ is defined in the same manner as example (a) except that a_i and λ are complex numbers. In the definition of *Euclidean norm*, it is necessary to use absolute value bars:

$$|\mathbf{a}|_e = \sqrt{|a_1|^2 + \cdots + |a_n|^2}.$$

(c) Let $K \subset \mathbb{R}^n$ be a compact set and $C(K)$ the set of all continuous real-valued functions $f(x) = f(x_1, \ldots, x_n)$ on K. Addition $h = f + g$ and scalar multiplication $k = \lambda f$ are defined for $f, g \in C(K)$ and real numbers λ in the natural way:

$$h(x) = f(x) + g(x); \quad k(x) = \lambda f(x) \quad \text{for} \quad x \in K.$$

As a norm one can choose, for instance,

$$\|f\|_0 = \max\{|f(x)| : x \in K\} \qquad \textit{maximum norm,}$$

or, more generally, a *weighted maximum norm*

$$\|f\|_1 = \sup\{|f(x)|p(x) : x \in K\},$$

where $p(x)$ is a given, fixed function with $0 < \alpha \le p(x) \le \beta < \infty$.

(d) This final example is needed in the investigation of complex differential equations. Let $G \subset \mathbb{C}$ be a domain in the complex plane and $H_0(G)$ the set of holomorphic (i.e., regular analytic) and bounded functions $u(z) : G \to \mathbb{C}$. If $p(z)$ is a real valued function in G and $0 < \alpha \le p(z) \le \beta$ for suitable positive constants α, β, then

$$\|u\| = \sup\{|u(z)|p(z) : z \in G\}$$

defines a norm in $H_0(G)$.

It is easy to verify the norm properties in each of these examples. The norms are homogeneous, nonnegative, they vanish only for the zero function, and they are finite. The triangle inequality is well known for the Euclidean norm $|\mathbf{a}|_e$, and easily verified for the other two norms in (a). In function spaces like (c) and (d), the triangle inequality holds under quite general hypotheses: G can be any set, the functions f, g, p need to be defined on G; f and g can be complex valued, p real valued and ≥ 0. If one sets

$$\|f\| = \sup\{|f(x)|p(x) : x \in G\},$$

then

$$|f(x) + g(x)|p(x) \leq |f(x)|p(x) + |g(x)|p(x) \leq \|f\| + \|g\| \text{ for } x \in G.$$

Therefore, the triangle inequality $\|f + g\| \leq \|f\| + \|g\|$ holds if the norms of f and g are finite.

IV. Convergence and Completeness. The notion of the convergence of a sequence of numbers can be extended in a natural way to a normed space L. A sequence $x_1, x_2, \ldots$ of elements of L converges "strongly" or "in the norm" to an element $x \in L$ if

$$\|x_n - x\| \to 0 \quad \text{as} \quad n \to \infty.$$

In this case, we also write

$$x_n \to x \quad (n \to \infty) \quad \text{or} \quad \lim_{n\to\infty} x_n = x.$$

Convergence for infinite series is defined similarly:

$$\sum_{k=1}^{\infty} x_k = x \Longleftrightarrow \left\| \sum_{k=1}^{n} x_k - x \right\| \to 0 \quad \text{as} \quad n \to \infty.$$

A sequence $x_1, x_2, \ldots$ is called a *Cauchy sequence* or a *fundamental sequence* if it satisfies the *Cauchy convergence criterion: For every $\varepsilon > 0$, there exists an $N_0(\varepsilon)$ such that $\|x_n - x_m\| < \varepsilon$ for $n, m \geq N_0(\varepsilon)$; or more briefly,*

$$\lim_{n,m\to\infty} \|x_n - x_m\| = 0.$$

It is well known that every Cauchy sequence of real or complex numbers has a limit (that is the essence of the Cauchy convergence criterion). However, the same is not true for all normed spaces; instead, it is a special property, called the *completeness property*, of certain linear spaces.

A normed linear space L is called *complete* if every Cauchy sequence of elements of L has a limit in L (in the sense of convergence in the norm).

V. Banach Spaces. A *Banach space* is a complete normed linear space, that is, a set with the properties given in sections I, II, and IV.

Examples III.(a),(c) are real Banach spaces, and III.(b),(d) are complex Banach spaces. In the first two examples the completeness follows immediately from the completeness of the spaces $\mathbb{R}$ and $\mathbb{C}$. In the third example, we rely on the following important observation:

Convergence in the maximum norm is equivalent to uniform convergence.

Indeed, if (f_n) is a Cauchy sequence, then the statement $\|f_n - f_m\|_0 < \varepsilon$ for $m, n \geq n_0$ is precisely the Cauchy criterion for uniform convergence:

$$|f_n(x) - f_m(x)| < \varepsilon \quad \text{for} \quad m, n \geq n_0 \ \text{ and all } \ x \in K. \tag{$*$}$$

The completeness then follows from the well-known theorem that the limit of a uniformly convergent sequence of continuous functions is again continuous. Therefore, there exists a function $f \in C(K)$ such that $\lim\limits_{n\to\infty} f_n(x) = f(x)$ uniformly in K. If x and n in $(*)$ are fixed and $m \to \infty$, then it follows that

$$|f_n(x) - f(x)| \leq \varepsilon \quad \text{for} \quad n \geq n_0 \quad \text{and} \quad x \in K;$$

i.e., $\|f_n - f\|_0 \leq \varepsilon$ for $n \geq n_0$. Thus $f_n \to f$ in the sense of convergence in the norm, and hence $C(K)$ is complete.

This argument is also valid for the weighted maximum norm $\|f\|_1$ in III.(c). There we assumed that $0 < \alpha \leq p(x) \leq \beta$ in K, so that

$$\alpha\|f\|_0 \leq \|f\|_1 \leq \beta\|f\|_0.$$

It follows that these two norms are equivalent:

Equivalence of Norms. Two norms $\|\cdot\|$ and $\|\cdot\|'$ on a vector space L are said to be *equivalent* if $\|x\| \leq \alpha\|x\|'$ and $\|x\|' \leq \beta\|x\|$ for all $x \in L$ (α, β constant), which means that convergence in one norm implies and is implied by convergence in the other norm. The two norms $\|f\|_0$ and $\|f\|_1$ from III.(c) are equivalent. It will be shown in 10.III that all norms in $\mathbb{R}^n$ are equivalent. Equivalence of norms is discussed in more detail in D.II.

Completeness of the space in Example (d) is obtained in a similar manner; here, however, one needs the theorem that the limit of a uniformly convergent sequence of holomorphic functions is holomorphic; cf. C.VI.

VI. Operators and Functionals. Continuity and Lipschitz Condition. Let E, F be two real or complex normed spaces and $T : D \to F$ a function with $D \subset E$. It is customary to refer to such mappings as operators or, in the case where $F = \mathbb{R}$ or $F = \mathbb{C}$, as functionals. An operator $T : D \to F$ is called linear if D is a linear subspace of E and $T(\lambda x + \mu y) = \lambda T(x) + \mu T(y)$ holds for $x, y \in D$ and $\lambda, \mu \in \mathbb{R}$ or $\mathbb{C}$, respectively. The value of T at x is usually written Tx instead of $T(x)$.

An operator $T : D \to F$ is said to be *continuous* at a point $x_0 \in D$ if $x_n \in D$, $x_n \to x_0$ implies that $Tx_n \to Tx_0$. The equivalent δ, ε-formulation reads: *For every $\varepsilon > 0$, there exists $\delta > 0$ such that from $x \in D$, $\|x - x_0\| < \delta$, it follows* *that* $\|Tx - Tx_0\| < \varepsilon$. An operator T satisfies a *Lipschitz condition* in D (with *Lipschitz constant* q) if

$$\|Tx - Ty\| \leq q\|x - y\| \quad \text{for} \quad x, y \in D. \tag{3}$$

It is easy to check that such an operator is continuous in D.

Note that the norms of two spaces, E and F, appear in this inequality. We make no distinction between these two norms in the notation because $E = F$ in most applications.

Remark. If T satisfies a Lipschitz condition, there is always a smallest Lipschitz constant. Let q_0 be the infimum of all numbers q for which (3) holds (for all $x, y \in D$); then for fixed x and y the inequality (3) clearly also holds with q_0 in place of q.

The Operator Norm. If T is linear, then (3) follows from

$$\|Tx\| \leq q\|x\| \quad \text{for} \quad x \in D, \tag{3'}$$

and conversely (take $y = 0$ in (3) for the latter case). The smallest Lipschitz constant in this case is called the *operator norm* of T and is denoted by $\|T\|$. It is given by $\|T\| = \sup\{\|Tx\| : \|x\| \leq 1\}$; cf. D.III.

VII. Some Examples. (a) In the special case $E = F = \mathbb{R}$, every linear operator T is of the form $Tx = cx$ with $c \in \mathbb{R}$, and $\|T\| = c$.

(b) Let $D = E = C(J)$, $J = [a, b]$, $F = \mathbb{R}$, and

$$Tf = \int_a^b f(t)\,dt.$$

Clearly, T is a linear functional. It satisfies a Lipschitz condition (3) with $q = b - a$ if the maximum norm is used in E and the absolute value in $F = \mathbb{R}$. This is the smallest Lipschitz constant, i.e., $\|T\| = b - a$.

(c) Let $D = E = F = C(J)$ with $J = [a, b]$, and

$$(Tf)(x) = \int_a^x f(t)\,dt.$$

The operator T is linear and satisfies a Lipschitz condition (3) with $q = b - a$ (maximum norm). Furthermore, if the weighted maximum norm with $p(x) = \mathrm{e}^{-x}$ is used (cf. III.(c)), then $q = 1 - \mathrm{e}^{-(b-a)}$ (Exercise!).

(d) We consider the functional $Tx = \|x\|$ from E to $\mathbb{R}$ $(= F)$. From inequality (2) it follows immediately that a Lipschitz condition is satisfied with $q = 1$: *The norm in E is a continuous functional; it satisfies a Lipschitz condition with Lipschitz constant* 1.

VIII. Iteration in Banach Spaces. Contractive Mappings. Many existence problems in analysis — and this also includes, as we will see, existence problems for ordinary differential equations — can be written as an equation of the form

$$x = Tx \tag{4}$$

in a suitably chosen Banach space B. Here T is an operator $D \to B$ with $D \subset B$. A solution of (4) is called a *fixed point* of T; it is a point which remains "fixed" under the map T.

Fixed points are frequently found using an iteration procedure called the *method of successive approximation*: starting from an element $x_0 \in D$, one forms successively the elements

$$x_1 = Tx_0,\ x_2 = Tx_1,\ \ldots,\ x_{n+1} = Tx_n, \ldots. \tag{5}$$

The central question of whether the sequence (x_n) converges to a solution of equation (4), is intimately connected to the concept of a contractive mapping. The mapping $T : D \to B$ is called *contracting* or a *contraction* if it satisfies a Lipschitz condition (3) with a Lipschitz constant $q < 1$. In this case, (3) says that the distance between the image points Tx, Ty under the mapping is smaller by a factor q than the distance between the two original points x, y, and hence T "contracts" distances between points.

The following result is fundamental for later existence proofs.

IX. Fixed Point Theorem for Contractive Mappings (The Contraction Principle). *Let D be a nonempty, closed set in a Banach space B. Let the operator $T : D \to B$ map D into itself, $T(D) \subset D$, and be a contraction, i.e., satisfy a Lipschitz condition* (3) *with constant $q < 1$. Then equation* (4) *has exactly one solution $x = \bar{x}$ in D.*

If a sequence (x_n) of "successive approximations" is formed according to (5), *beginning with an arbitrary element $x_0 \in D$, then the sequence converges (in the norm) to $\bar{x}$, and we have the estimate*

$$\|x_n - \bar{x}\| \le \frac{1}{1-q}\|x_{n+1} - x_n\| \le \frac{q^n}{1-q}\|x_1 - x_0\|. \tag{6}$$

Proof. First we note that $x_n \in D$ implies $x_{n+1} \in D$ because $T(D) \subset D$; thus the sequence (x_n) can be constructed as indicated in (5) and is contained in D.

We will first establish the estimate

$$\|x_{n+1} - x_n\| \le q^n \|x_1 - x_0\| \quad (n = 0, 1, 2, \ldots). \tag{7}$$

This inequality is clearly true for $n = 0$ and can be easily proved by induction. Assume that (7) is true for index n. Then from (3) it follows that

$$\begin{aligned} \|x_{n+2} - x_{n+1}\| &= \|Tx_{n+1} - Tx_n\| \le q\|x_{n+1} - x_n\| \\ &\le q^{n+1}\|x_1 - x_0\|. \end{aligned}$$

Thus (7) holds for the index $n+1$ and hence, by induction, for all n.

Applying (1) to the identity $x - y = (x - Tx) + (Tx - Ty) + (Ty - y)$, one obtains

$$\|x - y\| \leq \|x - Tx\| + \|Tx - Ty\| + \|Ty - y\| \quad \text{for} \quad x, y \in D.$$

The right-hand side of this inequality becomes larger if the term $\|Tx - Ty\|$ is replaced by $q\|x - y\|$ (cf. (3)). Moving this expression to the left-hand side and dividing both sides by $1 - q$ gives the inequality

$$\boxed{\|x - y\| \leq \frac{1}{1-q}\{\|x - Tx\| + \|y - Ty\|\} \quad \text{for} \quad x, y \in D.} \tag{8}$$

All of the assertions follow very quickly from (8). The quantity $x - Tx$ is called the *defect* of x relative to equation (4), and correspondingly, the inequality (8) is called a *defect inequality*.

If x and y are fixed points of T, then (8) implies that $\|x - y\| = 0$ and hence the uniqueness of the fixed point. Setting $x = x_{n+p}$ in (8) with $p > 0$ and $y = x_n$, then using (7) and $Tx = x_{n+p+1}$, $Ty = x_{n+1}$, one obtains

$$\begin{aligned}
\|x_{n+p} - x_n\| &\leq \frac{1}{1-q}\{\|x_{n+p+1} - x_{n+p}\| + \|x_{n+1} - x_n\|\} \\
&\leq \frac{1}{1-q}(q^{n+p} + q^n)\|x_1 - x_0\| \leq Cq^n,
\end{aligned}$$

where $C = 2\|x_1 - x_0\|/(1-q)$. This implies that (x_n) is a Cauchy sequence, which then has a limit $\bar{x}$ because of the completeness of B. Since D is closed, $\bar{x}$ is in D. On the one hand, we have $Tx_n \to T\bar{x}$ from the continuity of T, and on the other hand, $Tx_n = x_{n+1} \to \bar{x}$. Thus $\bar{x}$ is a fixed point of T. The first inequality in (6) is obtained from (8) when $x = x_n$, $y = \bar{x}$ are substituted, and the second inequality then follows from (7). This completes the proof of this important theorem. ∎

X. Remarks. (a) The method of successive approximation is easy to illustrate graphically in the special case $B = \mathbb{R}$. Suppose $T(x)$ is a real function of a real variable x defined in an interval $D = [a, b]$. The assumption $T(D) \subset D$ means that $a \leq T(x) \leq b$. The Lipschitz condition (3) is an estimate on the size of the difference quotients

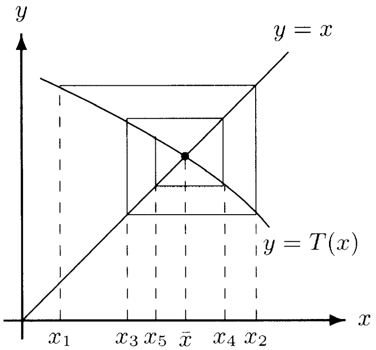
$$\left|\frac{T(x) - T(y)}{x - y}\right| \leq q < 1 \quad \text{for } x, y \in D,\ x \neq y.$$

If $T \in C^1(D)$, it is equivalent to $|T'(x)| \leq q$ in D.

A solution $\bar{x}$ of the equation

$$x = T(x)$$

corresponds geometrically to the intersection of the line $y = x$ with the curve $y = T(x)$. The construction (5) can be carried out graphically: See the figure,



The method of successive approximation in the case $T : \mathbb{R} \to \mathbb{R}$

in which $-1 < T'(x) < 0$. The reader should make sketches for each of the other three cases $0 < T'(x) < 1$, $T'(x) \geq 1$, $T'(x) \leq -1$, and verify that in the first of these cases the method of successive approximations converges but that it does not converge in the other two cases.

(b) The preceding fixed point theorem goes back to Banach (1922) and is also called the *Banach fixed point theorem*. Stefan Banach (1892–1945, Polish mathematician) is one of the founders of functional analysis. In the work cited (his dissertation) he introduced the central concept of a normed space and laid the foundations of the corresponding theory.

XI. Exercises. (a) Let $M \subset \mathbb{R}^n$ be an arbitrary set and $p(x)$ a positive, continuous function on M. Show that the subset $C(M;p)$ of all functions $f \in C(M)$ for which the norm

$$\|f\| := \sup\{|f(x)|p(x) \,:\, x \in M\}$$

is finite, forms a real (or complex) Banach space.

Hint: A Cauchy sequence in this norm is locally uniformly convergent, i.e., for $x \in M$ there exists a neighborhood $U(x)$ such that the sequence converges uniformly in $U(x) \cap M$. Don't forget to check that $C(M;p)$ is a linear space.

This result is false if p has zeros; see (b).

(b) Let L be the space of continuous functions $f(x)$ on $0 \leq x \leq 1$ and $\|f\| = \max|x^2 f(x)|$. Show that this defines a norm, but that the space L is not complete.

Hint: Consider the sequence f_n with $f_n(x) = \dfrac{1}{x}$ for $\dfrac{1}{n} \leq x \leq 1$, $f_n(x) = n$ for $0 \leq x \leq \dfrac{1}{n}$.

(c) Let $C(M;p)$ be the Banach space from (a) and ϕ, ψ two real-valued functions defined in M with $\phi(x) \leq \psi(x)$ in M. Show that the set of all $f \in C(M;p)$ with $\phi(x) \leq f(x) \leq \psi(x)$ for $x \in M$ is closed.

XII. Exercises. (a) Let $J = [0, a]$. We define three norms in $C(J)$, the maximum norm $\|f\|_0$ and

$$\|f\|_1 = \max_J |f(x)| e^{-ax}, \quad \|f\|_2 = \max_J |f(x)| e^{-x^2}.$$

Determine the corresponding operator norms $\|T\|_0$, $\|T\|_1$, $\|T\|_2$ for the operator

$$(Tf)(x) = \int_0^x t f(t)\, dt.$$

(b) Show that the integral equation

$$y(x) = \frac{1}{2}x^2 + \int_0^x t y(t)\, dt, \ x \in J = [0, a],$$

has exactly one solution and determine it (i) by rewriting the equation as an initial value problem and solving it, and (ii) by using the results from (a) and explicitly calculating the successive approximations (5), beginning with $y_0 = 0$.

(c) In $C^1(J)$, $J = [a, b]$, let $\|f\|_0$ be the maximum norm and $\|f\|_1 := \|f\|_0 + \|f'\|_0$. Show that this space is a Banach space with the norm $\|\cdot\|_1$, but not with the norm $\|\cdot\|_0$.

§ 6. An Existence and Uniqueness Theorem

All of the functions in this section are assumed to be real valued. We consider the following initial value problem

$$y' = f(x, y) \quad \text{for} \quad \xi \le x \le \xi + a, \quad y(\xi) = \eta. \tag{1}$$

The main assumptions in the following theorem are that f is continuous in the strip $S = J \times \mathbb{R}$ with $J = [\xi, \xi + a]$ and satisfies a *Lipschitz condition with respect to y in S:*

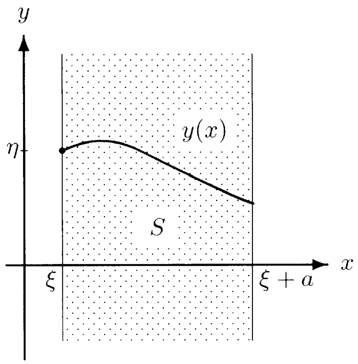
$$|f(x, y) - f(x, \bar{y})| \le L|y - \bar{y}|. \tag{2}$$

No restrictions are placed on the value of the *Lipschitz constant* $L \ge 0$.

I. Existence and Uniqueness Theorem. *Let $f \in C(S)$ satisfy the Lipschitz condition* (2). *Then the initial value problem* (1) *has exactly one solution $y(x)$. The solution exists in the interval $J : \xi \le x \le \xi + a$.*

The *proof* is essentially an application of the fixed point theorem 5.IX. As a preliminary step, the initial value problem is transformed into an equivalent fixed point equation $y = Ty$. Let $y(x)$ be differentiable on the interval J and satisfy the initial value problem (1). Because of the continuity of f, $u(x) := f(x, y(x))$ is continuous in J, so $y(x)$ is actually continuously differentiable. Therefore, by the fundamental theorem of calculus

$$y(x) = \eta + \int_\xi^x f(t, y(t))\, dt. \tag{3}$$



Conversely, if $y(x)$ is a solution of (3) that is merely *continuous* in J, then the right-hand side of (3) is continuously differentiable; hence so is $y(x)$, and $y' = f(x,y)$ holds. Furthermore, y satisfies the initial condition $y(\xi) = \eta$. Therefore, the initial value problem (1) is equivalent to the integral equation (3), which can be written in the form of an operator equation

$$y = Ty \quad \text{with} \quad (Ty)(x) = \eta + \int_\xi^x f(t, y(t))\, dt. \tag{3'}$$

The integral operator T maps each function y from the Banach space $C(J)$ of continuous functions (cf. Example 5.III.(c)) to a function Ty in the same space.

It follows that the solutions to the initial value problem (1) are precisely the fixed points of the operator T, considered as a mapping $B \to B$ with $B = C(J)$. To complete the proof of Theorem I, we show that the operator T satisfies a Lipschitz condition (5.3) with a Lipschitz constant $q < 1$, and we then apply the fixed point theorem 5.IX.

If the space $C(J)$ is normed with the maximum norm $\|y\|_0 = \max\{|y(x)| : x \in J\}$, then (2) implies that for $x, y \in C(J)$,

$$\begin{aligned} |(Ty)(x) - (Tz)(x)| &= \left| \int_\xi^x \{f(t, y(t)) - f(t, z(t))\}\, dt \right| \\ &\le \int_\xi^x L|y(t) - z(t)|\, dt \le L\|y - z\|_0 (x - \xi), \end{aligned} \tag{4}$$

and hence, because $x - \xi \le a$,

$$\|Ty - Tz\|_0 \le La\|y - z\|_0.$$

Hence T satisfies a Lipschitz condition with Lipschitz constant La. Note, however, that this Lipschitz constant is less than 1 only if the interval is small, since $La < 1$ implies that $a < \dfrac{1}{L}$. One way to handle the case where $a \ge \dfrac{1}{L}$ is to find an n such that $b = \dfrac{a}{n} < \dfrac{1}{L}$ and then use the above procedure to determine the solution successively on the intervals

$$\xi \le x \le \xi + b,\ \xi + b \le x \le \xi + 2b, \ldots, \xi + (n-1)b \le x \le \xi + nb = \xi + a.$$

To do this one needs the result in VI.(b).

A more elegant way is to work with a weighted maximum norm:

$$\|y\| = \max\{|y(x)|e^{-\alpha x} : x \in J\} \quad (\alpha > 0). \tag{5}$$

The last integral in (4) is now estimated as follows:

$$L\int_\xi^x |y(t) - z(t)|e^{-\alpha t}e^{\alpha t}\,dt \le L\|y - z\| \int_\xi^x e^{\alpha t}\,dt \le L\|y - z\|\frac{e^{\alpha x}}{\alpha}.$$

Then we conclude from (4) that

$$|(Ty)(x) - (Tz)(x)|e^{-\alpha x} \le \frac{L}{\alpha}\|y - z\|,$$

and hence

$$\|Ty - Tz\| \le \frac{L}{\alpha}\|y - z\|.$$

Thus if one chooses $\alpha = 2L$, for example, then T satisfies a Lipschitz condition with Lipschitz constant $\frac{1}{2}$. This variant of the proof gives existence for the whole interval (of arbitrary length) in a *single* step. ■

II. Comments. (a) The theorem shows that starting with an arbitrary function $y_0(x) \in C(J)$ and calculating the sequence of "successive approximations" given by

$$y_{k+1}(x) = \eta + \int_\xi^x f(t, y_k(t))\,dt \quad (k = 0, 1, 2, \ldots), \tag{6}$$

one obtains a sequence that converges in the norm, and hence uniformly in J, to the solution $y(x)$ of the initial value problem. This iteration procedure can also be used to determine a numerical approximation to the solution. In numerical approximations, it is a good idea to start with a function $y_0(x)$ that is as close as possible to the solution. However, if nothing is known about the solution, then $y_0(x) = \eta$ is not a bad choice.

(b) The following is a sufficient condition for the Lipschitz condition (2) to hold: *f is differentiable with respect to y, and* $|f_y(x, y)| \le L$ (the proof uses the mean value theorem).

(c) *Existence and Uniqueness Theorem to the Left of the Initial Value.* Let $J_- = [\xi - a, \xi]$ $(a > 0)$. If f is continuous in the strip $S_- := J_- \times \mathbb{R}$ and the Lipschitz condition (2) holds in S_-, then the initial value problem

$$y' = f(x, y) \quad \text{for} \quad \xi - a \le x \le \xi, \quad y(\xi) = \eta \tag{1_-}$$

has exactly one solution in J_-.

This result can be proved by

(d) *Reflection about the Line $x = \xi$.* We introduce the functions $\bar{y} := y(2\xi - x)$, $\bar{f}(x, y) := -f(2\xi - x, y)$ to transform the problem (1_-) in J_- into the initial value problem in J

$$\bar{y}' = \bar{f}(x, \bar{y}) \quad \text{for} \quad \xi \le x \le \xi + a, \quad \bar{y}(\xi) = \eta. \tag{1_+}$$

Clearly, $\bar{f}$ satisfies the hypotheses of Theorem I. In addition, one can see at once that $\phi(x) \mapsto \bar{\phi} := \phi(2\xi - x)$ defines a bijective mapping of $C(J_-)$ onto $C(J)$ that maps solutions of (1_-) into solutions of (1_+) (and conversely). The conclusion then follows from Theorem I.

(e) We note that an alternative approach is to carry the original proof directly over to the present case. Existence to the left and to the right can both be proved using the norm

$$\|y\| = \max |y(x)| e^{-\alpha|x-\xi|}$$

(equation (3) holds in both cases).

Frequently, f is not defined in the whole strip, but only in a neighborhood of the point (ξ, η). The following result deals with this situation.

III. Theorem. *Let R be the rectangle $\xi \le x \le \xi + a$, $|y - \eta| \le b$ $(a, b > 0)$ and let $f \in C(R)$ satisfy a Lipschitz condition* (2) *in R. Then there exists exactly one solution to the initial value problem* (1). *The solution exists (at least) in an interval $\xi \le x \le \xi + \alpha$, where*

$$\alpha = \min\left(a, \frac{b}{A}\right), \quad \textit{with} \quad A = \max_R |f|.$$

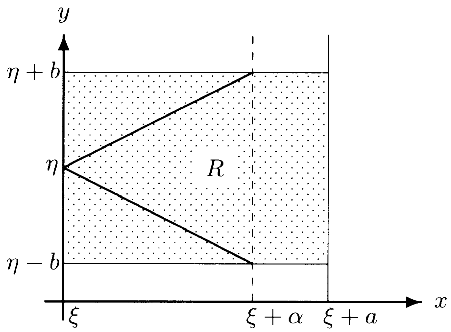
A corresponding statement holds for $\xi - a \le x \le \xi$, with the rectangle R lying to the left of the point (ξ, η).

For the *proof*, we extend f continuously to the strip $\xi \le x \le \xi + a$, $-\infty < y < \infty$; for example, by setting

$$\bar{f}(x, y) = \begin{cases} f(x, \eta - b) & \text{for} \quad y < \eta - b, \\ f(x, y) & \text{in} \quad R, \\ f(x, \eta + b) & \text{for} \quad y > \eta + b. \end{cases}$$

The function $\bar{f}$ is clearly continuous in the strip and satisfies the same Lipschitz condition (2) as f with the same Lipschitz constant. By Theorem I, there exists exactly one solution $y(x)$ of the initial value problem with right-hand side $\bar{f}$. As long as this solution remains in R, it is also a solution of the original initial value problem. Since $|\bar{f}| \le A$, we have $|y'| \le A$; i.e., the solution remains in the angular region formed by the two lines through the point (ξ, η) with slopes $\pm A$ (see the figure). Hence the solution does not leave R as long as $\xi \le x \le \xi + \alpha$, where α is the smaller of the two numbers a and b/A. ■

Remark. We sketch another proof that does not require a continuous extension of f outside of R. Let $J' = [\xi, \xi + \alpha]$. One considers the Banach space $B = C(J')$ and the operator T as defined in (3′) on the subset D of all $\phi \in B$ with $|\phi(x) - \eta| \le b$. In order to apply the fixed point theorem 5.IX, one has to show that D is closed, $T(D) \subset D$, and T is a contraction (proof as in Theorem I). The details of this proof are recommended as an exercise.



IV. Local Lipschitz Condition. Definition. The function $f(x,y)$ is said to satisfy a local Lipschitz condition with respect to y in $D \subset \mathbb{R}^2$ if for every $(x_0, y_0) \in D$ there exists a neighborhood $U = U(x_0, y_0)$ and an $L = L(x_0, y_0)$ such that in $U \cap D$ the function f satisfies the Lipschitz condition

$$|f(x,y) - f(x,\bar{y})| \leq L|y - \bar{y}|. \tag{2}$$

Criterion. *If D is open and if $f \in C(D)$ has a continuous derivative f_y in D, then f satisfies a local Lipschitz condition in this set.*

Suppose, namely, that U is a circular neighborhood of $(x_0, y_0) \in D$ with $\overline{U} \subset D$; then f_y is bounded in U, say $|f_y| \leq L$, and for (x,y), $(x,\bar{y}) \in U$, the relation

$$f(x,y) - f(x,\bar{y}) = (y - \bar{y}) f_y(x, y^*) \quad \text{with} \quad y^* \in (y, \bar{y})$$

follows from the mean value theorem. Hence, (2) holds. ∎

A local Lipschitz condition is a weak requirement compared to the global Lipschitz condition in Theorem I. For instance, $f(x,y) = y^2$ satisfies

$$|f(x,y) - f(x,\bar{y})| = |y^2 - \bar{y}^2| = |y + \bar{y}||y - \bar{y}|.$$

Therefore, f satisfies a local Lipschitz condition in $\mathbb{R}^2$ (or in the strip $J \times \mathbb{R}$) but does not satisfy a Lipschitz condition in this set.

Theorem on Local Solvability. *If D is open and $f \in C(D)$ satisfies a local Lipschitz condition in D, then the initial value problem* (1) *is locally uniquely solvable for $(\xi, \eta) \in D$; i.e., there is a neighborhood I of ξ such that exactly one solution exists in I.*

This follows immediately from Theorem III. A rectangle like the one that appears in III is constructed to the right of the point (ξ, η). If the rectangle is chosen small enough, then a Lipschitz condition holds in this rectangle and Theorem III applies. A corresponding argument holds to the left. ∎

Our next objective is to derive some global statements on the unique extension of these local solutions. We begin with a lemma that looks awkward at first sight. It will later be used in different situations in connection with global existence.

V. Lemma. *Let f be defined in D and let $\Phi = \{\phi_\alpha\}_{\alpha \in A}$ $(A \neq \emptyset)$ be a set of functions, where ϕ_α is a solution to the initial value problem* (1) *in the interval J_α containing ξ. Assume that the following property* (U) *holds:*

$$\phi_\alpha(x) = \phi_\beta(x) \quad \textit{for} \quad x \in J_\alpha \cap J_\beta \quad (\alpha, \beta \in A). \tag{U}$$

Then there exists exactly one solution ϕ defined in the interval $J := \bigcup_{\alpha \in A} J_\alpha$ with the property $\phi\,|_{J_\alpha} = \phi_\alpha$ for every $\alpha \in A$.

This solution can be constructed as follows: For each $x \in J$ determine an $\alpha \in A$ such that $x \in J_\alpha$ and then define $\phi(x) = \phi_\alpha(x)$. If β is another index with $x \in J_\beta$, then by hypothesis $\phi_\alpha(x) = \phi_\beta(x)$, i.e., $\phi(x)$ is well-defined.

If a is an arbitrary point from J, then there exists an $\alpha \in A$ such that $a \in J_\alpha$. Thus $[\xi, a] \subset J_\alpha$ and $\phi(x) = \phi_\alpha(x)$ holds in $[\xi, a]$. It follows that ϕ is a solution of (1) in J. ∎

If this lemma is applied to the set of *all* solutions to the initial value problem, then (U) implies the uniqueness of the solution. In summary, one obtains the following

Corollary. *If the initial value problem* (1) *has at least one solution and if the uniqueness statement* (U) *holds for every pair of solutions, then there exists a solution of* (1) *which cannot be extended. All other solutions are restrictions of this solution.*

VI. Lemma on the Extension of Solutions. *Let $D \subset \mathbb{R}^2$ and $f \in C(D)$.*

(a) *If ϕ is a solution of the differential equation $y' = f(x, y)$ in the interval $\xi \le x < b$ such that* graph ϕ *remains in a compact set $A \subset D$, then ϕ can be extended to the closed interval $[\xi, b]$.*

(b) *If ϕ is a solution in the interval $[\xi, b]$ and ψ a solution in the interval $[b, c]$ with $\phi(b) = \psi(b)$, then the function*

$$u(x) := \begin{cases} \phi(x) & \text{for} \quad \xi \le x \le b, \\ \psi(x) & \text{for} \quad b < x \le c \end{cases}$$

is a solution in the interval $[\xi, c]$.

Proof. (a) The function f is bounded on A, say $|f| \le C$. This means that $|\phi'| \le C$ and therefore ϕ is uniformly continuous in $[\xi, b)$. It follows that $\beta = \lim_{x \to b-} \phi(x)$ exists, and $(b, \beta) \in A$. If one sets $\phi(b) = \beta$, then $\phi(x)$, and hence also $f(x, \phi(x))$, is continuous in $[\xi, b]$. The equation

$$\phi(x) = \phi(\xi) + \int_\xi^x f(t, \phi(t))\, dt$$

holds for $\xi \le x < b$. The limit as $x \to b-$ shows that this is also valid for $x = b$. Therefore ϕ is differentiable (to the left) at b and $\phi'(b) = f(b, \phi(b))$.

(b) It is sufficient to check that u satisfies the differential equation at b. The function u is differentiable to the left and the right at this point, and both derivatives are equal to $f(b, \phi(b))$. ∎

We come now to the main theorem of this section.

VII. Existence and Uniqueness Theorem. *Let $f \in C(D)$ satisfy a local Lipschitz condition with respect to y in D, where $D \subset \mathbb{R}^2$ is open. Then for every $(\xi, \eta) \in D$ the initial value problem*

$$y' = f(x, y), \quad y(\xi) = \eta \tag{7}$$

has a solution ϕ that cannot be extended and that to the left and to the right comes arbitrarily close to the boundary of D. The solution is uniquely determined in the sense that every solution of (7) is a restriction of ϕ.

Definition. The statement "ϕ comes arbitrarily close to the boundary of D *to the right*" is defined as follows: If G is the closure of graph ϕ and if G_+ is the set of points $(x, y) \in G$ with $x \ge \xi$, then

(a) G_+ is not a compact subset of D.

An equivalent formulation that gives a better understanding reads as follows: ϕ exists to the right in an interval $\xi \le x < b$ ($b = \infty$ is allowed), and one of the following cases applies:

(b) $b = \infty$; the solution exists for all $x \ge \xi$.

(c) $b < \infty$ and $\limsup\limits_{x \to b-} |\phi(x)| = \infty$; the solution "becomes infinite."

(d) $b < \infty$ and $\liminf\limits_{x \to b-} \rho(x, \phi(x)) = 0$, where $\rho(x, y)$ denotes the distance from the point (x, y) to the boundary of D; the solution "comes arbitrarily close to the boundary of D."

Indeed, statement (a) says that G_+ is either unbounded (case (b) or (c)) or is bounded and contains boundary points of D (case (d)).

We have repeatedly encountered these three types of behavior. In the example $y' = e^y \sin x$ of 1.VIII, (b) or (c) holds to the left and to the right, depending on the value of $y(0)$. For the equation $y' = (2y)^{-1}$ in the upper half plane $y > 0$, all solutions are given by $y = \sqrt{x + c}$ ($x > -c$). Here, case (b) prevails to the right and case (d) to the left.

Proof. Uniqueness. We prove the statement "If ϕ and ψ are two solutions of the initial value problem and if J is a common interval of existence of both solutions with $\xi \in J$, then $\phi = \psi$ in J."

Let us assume on the contrary that there exist, say to the right of ξ, points $x \in J$ with $\phi(x) \ne \psi(x)$. Then there also exists a first point $x_0 \in J$ to the right of ξ where the two solutions separate. This x_0 is the largest number with the property that $\phi(x) = \psi(x)$ for $\xi \le x \le x_0$ ($x_0 = \xi$ is not excluded).

However, we know from IV that there exists a local solution through the point $(x_0, \phi(x_0))$ and that it is uniquely determined. In other words, $\phi(x) = \psi(x)$ in a right neighborhood of x_0. This is a contradiction to our assumption about x_0. The uniqueness to the left is proved similarly.

Existence. By Theorem IV there exists a local solution to (7), and as we have just proved, the uniqueness statement (U) of V holds. Thus Corollary V guarantees the existence of a nonextendable solution ϕ, and we have only to show that it comes arbitrarily close to the boundary of D (we consider only the case "to the right" in the direction of increasing x, $x \geq \xi$).

Assume that (a) is false. Then G_+ is a compact subset of D, and ϕ exists in a finite interval $\xi \leq x < b$ or $\xi \leq x \leq b$. In the first case, Lemma VI.(a) can be applied, i.e., ϕ can be extended to $[\xi, b]$. In the second case, $(b, \phi(b)) \in D$, and there exists a local solution ψ that passes through this point. Applying VI.(b), one again obtains an extension of ϕ.

In either case, we have a contradiction to the assumption that ϕ cannot be extended. This completes the proof of the theorem. ∎

VIII. Exercise. Let $k(x, t, z)$ be continuous for $0 \leq t \leq x \leq a$, $-\infty < z < \infty$ and satisfy a Lipschitz condition in z,

$$|k(x, t, z) - k(x, t, \bar{z})| \leq L|z - \bar{z}|,$$

and let $g(x)$ be continuous for $0 \leq x \leq a$. Show, by applying the fixed point theorem 5.IX, that the *Volterra integral equation*"

$$u(x) = g(x) + \int_0^x k(x, t, u(t))\, dt$$

has exactly one continuous solution in $0 \leq x \leq a$.

IX. Exercise. Prove: If $f(x, y)$ satisfies a local Lipschitz condition with respect to y in the set $D \subset \mathbb{R}^2$ and if $A \subset D$ is compact and f bounded on A, then f satisfies a Lipschitz condition with respect to y in A. In particular, if $v, w \in C([a, b])$ and $\operatorname{graph} v$, $\operatorname{graph} w \subset D$, then there exists $L > 0$ such that

$$|f(x, v(x)) - f(x, w(x))| \leq L|v(x) - w(x)| \quad \text{in} \quad [a, b].$$

X. Exercise. Prove: If f is continuous in the open set D and ϕ is a solution of (7) in the interval $[\xi, b)$ with $b < \infty$ that comes arbitrarily close to the boundary of D to the right, then at least one of the following two cases applies (both can happen at the same time):

(c′) $\phi(x) \to +\infty$ or $-\infty$ as $x \to b-$;

(d′) $\rho(x, \phi(x)) \to 0$ as $x \to b-$.

This sharpens the statement in VII.

Hint: Show: If G_b is the intersection of $\overline{\operatorname{graph} \phi}$ with the line $x = b$, then $G_b \subset \partial D$ (the boundary of D).

XI. Exercise. Rosenblatt's Condition. Let the function $f(x,y)$ be continuous in the strip $S = J \times \mathbb{R}$, $J = [0,a]$ and satisfy the condition

$$|f(x,y) - f(x,z)| \le \frac{q}{x}|y-z| \quad \text{for} \quad 0 < x \le a \quad \text{and} \quad y,z \in \mathbb{R}$$

with $q < 1$. Show that the initial value problem

$$y' = f(x,y) \quad \text{in } J, \quad y(0) = \eta$$

has exactly one solution and that this solution can be obtained by the method of successive approximations. The above condition was introduced by Rosenblatt (1909).

Hint. In the Banach space B of all functions $u \in C(J)$ with *finite* norm

$$\|u\| := \sup\{|u(x)|/x : 0 < x \le a\},$$

the operator T,

$$(Tu)(x) := \int_0^x f(t, \eta + u(t))\,dt,$$

satisfies the Lipschitz condition (5.3). If u is a fixed point of T, then $y = u + \eta$ is a solution of the initial value problem.

Supplement: Singular Initial Value Problems

Here we consider a singular initial value problem for a differential equation of second order,

$$y'' + \frac{\alpha}{x} y' = f(x,y) \quad \text{in} \quad J_0 = (0,b], \quad y(0) = \eta, \quad y'(0) = 0. \tag{8}$$

This problem is closely connected to the problem of finding rotationally symmetric solutions of the nonlinear elliptic equation

$$\Delta u = f(r,u),$$

where $x \in \mathbb{R}^n$ and $r = |x|$.

XII. The Operators $\mathbf{L}_\alpha$ and $\mathbf{I}_\alpha$. In what follows, $J = [0,b]$, $J_0 = (0,b]$, $\alpha > 0$, and L_α is the differential operator

$$L_\alpha y = y'' + \frac{\alpha}{x} y' = x^{-\alpha}(x^\alpha y')'.$$

Lemma. *Let $y \in C(J) \cap C^2(J_0)$, y' bounded, and $f(x) \in C(J)$. If*

$$L_\alpha y = f(x) \quad \text{in} \quad J_0, \quad y(0) = \eta, \tag{9}$$

then $y \in C^2(J)$, $y'(0) = 0$, $\lim_{x\to 0+} y'(x)/x = y''(0) = f(0)/(\alpha+1)$, and

$$y(x) = \eta + (I_\alpha f)(x) \quad \text{with} \quad I_\alpha f = \int_0^x s^{-\alpha} \int_0^s t^\alpha f(t)\,dt\,ds. \tag{10}$$

Conversely, if y is defined by (10), *then y is a solution of* (9) *with the above properties; in particular, $y \in C^2(J)$, $y'(0) = 0$.*

Proof. Because y' is bounded, we have $x^\alpha y'(x) \to 0$ as $x \to 0+$. By integrating $(x^\alpha y')' = x^\alpha f$, one obtains

$$x^\alpha y'(x) = \int_0^x t^\alpha f(t)\,dt. \tag{11}$$

The substitution $t = x\tau$, $dt = x\,d\tau$ in (11) leads to

$$\frac{y'(x)}{x} = \int_0^1 \tau^\alpha f(\tau x)\,d\tau. \tag{11'}$$

Since $f(x\tau) \to f(0)$ as $x \to 0+$ uniformly in $\tau \in [0,1]$, one derives from (11′) that $y'(x)/x \to f(0)/(\alpha+1)$ and, in particular, $y'(x) \to 0$ as $x \to 0+$. By a well-known theorem from analysis (see C.VI.(b)), $y \in C^1(J)$ and $y'(0) = 0$. Using equation (9) to determine y'', we get $y''(x) \to f(0)/(\alpha+1)$ as $x \to 0+$. A second application of the previously mentioned lemma shows that $y \in C^2(J)$ and that $y''(0)$ has the specified value.

Solving equation (11) for y' and then integrating gives (10). The integrand $s^{-\alpha}\displaystyle\int_0^s t^\alpha f(t)\,dt$ in I_α is a continuous function in J vanishing at 0. This follows from the boundedness of f together with the inequality $s^{-\alpha}t^\alpha \le 1$.

Conversely, one obtains (11) by differentiating equation (10). As we have seen, the specified properties of y follow from (11). Finally, if equation (11) for $x^\alpha y'$ is differentiated, (9) follows. ∎

XIII. Existence and Uniqueness Theorem. *Let the function $f(x,y)$ be continuous in $J \times \mathbb{R}$ and satisfy a Lipschitz condition* (2) *in y. Then for given $\alpha > 0$, the initial value problem* (8) *has exactly one solution $y \in C^2(J)$.*

Proof. By the previous lemma, (8) is equivalent to the Volterra integral equation $y = \eta + I_\alpha f(\cdot, y)$, which can be written in the form

$$y(x) = \eta + \int_0^x k(x,t) f(t, y(t))\,dt \tag{12}$$

with

$$k(x,t) = t^\alpha \int_t^x s^{-\alpha} ds.$$

Since $(t/s)^\alpha \le 1$ implies that $0 \le k(x,t) \le x - t$, it follows that the "kernel" $k(x,t)$ is continuous in the triangle $D: \ 0 \le t \le x \le b$. The assertion now follows from the theorem in Exercise VIII. ∎

XIV. Rotationally Symmetric Solutions of Elliptic Differential Equations. In the following, $x \in \mathbb{R}^n$, $n \ge 2$, $r = |x|$ (Euclidean norm), and Δ is the Laplace operator

$$\Delta u = u_{x_1x_1} + u_{x_2x_2} + \cdots + u_{x_nx_n}.$$

The Laplace operator for *rotationally symmetric* functions $u(x) = y(|x|)$ (they are also called *radial* functions) is given by

$$\Delta u = y'' + \frac{n-1}{r} y' = L_{n-1} y;$$

for proof, use the formulas for u_i, u_{ij} given below. The results established here lead at once to the following

Existence and Uniqueness Theorem. *Let the function $f(r, z)$ be continuous in $J \times \mathbb{R}$ and Lipschitz continuous in z. Denote by B_b the closed ball $B_b : |x| \leq b$. Then the differential equation*

$$\Delta u = f(|x|, u) \quad \textit{in} \quad B_b$$

has exactly one rotationally symmetric solution $u \in C^2(B_b)$ satisfying the initial condition $u(0) = u_0$.

This result follows immediately from the previous theorem. The solution u is obtained in the form $u(x) = y(|x|)$, where $y(r)$ is the solution of (12) with $\alpha = n - 1$, $\eta = u_0$, and r in the place of x. One question needs clarification. While the C^2-property of y carries over at once to u as long as $x \neq 0$, the same is not immediately obvious for $x = 0$. The next lemma gives information about this.

Lemma. *The function $u(x) = y(|x|)$ ($x \in \mathbb{R}^n$) is twice continuously differentiable in the ball $B_b : |x| \leq b$ if and only if $y \in C^2(J)$ and $y'(0) = 0$.*

Proof. Since $u(t, 0, \ldots, 0) = y(|t|)$ is an even function of t, it follows easily that $u \in C^2(B_b)$ implies $y \in C^2(J)$ and $y'(0) = 0$. For the proof of the converse proposition, the partial derivatives of u will be denoted by u_i and u_{ij}. For $x \neq 0$, we have

$$u_i = \frac{x_i}{r} y', \quad u_{ij} = \frac{\delta_{ij}}{r} y' + \frac{x_i x_j}{r^2} \left(y'' - \frac{y'}{r} \right).$$

Because $|x_i/r| \leq 1$, $u_i(x) \to 0$ as $x \to 0$. Setting $u_i(0) := 0$, one obtains a continuous function in B_b for which $\partial u(0)/\partial x_i = 0$; cf. B.VI.(a) and (b). Thus $u \in C^1(B_b)$. One proceeds in exactly the same manner with the second derivatives. From $y'(r)/r = (y'(r) - y'(0))/r \to y''(0)$ it follows that $y'' - y'/r \to 0$ as $r \to 0+$, thus $u_{ij}(x) \to \delta_{ij} y''(0)$ as $x \to 0$. Taking this value to define $u_{ij}(0)$ and using Theorem VI from Appendix B again, we see that this defines a continuous function in B_b and that $\partial u_i/\partial x_j = u_{ij}$ at $x = 0$. This completes the proof of this lemma and also the proof of the previous theorem. ∎

Radial solutions of elliptic equations have been studied extensively. Such solutions are important in differential geometry and in many areas of applied mathematics. The question of the existence of *entire* radial solutions (i.e., those that exist in $\mathbb{R}^n$) has been completely solved in the case of the equation $\Delta u + u^p = 0$. The first comprehensive results for more general equations of the form $\Delta u + K(r) u^p = 0$ were given by Ni (1982).

XV. Exercise. Comparison Theorem. *Assume that $v, w \in C^2(J)$ satisfy*

$$\begin{aligned} L_\alpha v &\le f(x,v), \quad L_\alpha w \ge f(x,w) \quad \text{in} \quad J_0 = (0,b], \\ v(0) &< w(0), \quad v'(0) = w'(0) = 0, \end{aligned}$$

where $f(x,y)$ is increasing in y. Then $v' \le w'$ and $v < w$ in J.

Hint: Show that $v' \le w'$ as long as $v < w$.

XVI. Exercise. Let the functions p and p_1 be continuous in $J = [0,b]$ and positive in $J_0 = (0,b]$ and let $p_1(t)/p(x) \le Cx^{-\gamma}$ for $0 < t \le x \le b$ with $0 \le \gamma < 1$. We consider the initial value problem

$$Ly = f(x,y) \quad \text{in} \quad J_0, \quad y(0) = \eta, \quad y'(0) = 0,$$

where

$$Ly = \frac{1}{p_1(x)}(p(x)y')'.$$

Prove: If the function $f(x,y)$ is continuous in $J \times \mathbb{R}$ and satisfies a Lipschitz condition in y, then the initial value problem has exactly one solution. For a solution we require $y \in C^1(J)$ and $py' \in C^1(J_0)$.

Hint: Reduce the problem to a Volterra integral equation and use the theorem from Exercise VIII.

§ 7. The Peano Existence Theorem

In Chapter I we dealt with instances where the right-hand side of the differential equation

$$y' = f(x,y) \tag{1}$$

does not satisfy a Lipschitz condition. An example is the equation $y' = \sqrt{|y|}$. The fundamentally important question whether continuity of $f(x,y)$ is sufficient for existence of a solution was first answered in the affirmative by the Italian mathematician and logician Giuseppe Peano (1858–1932). Peano's paper (1890) is written in logical symbols and was later "translated" by G. Mie (1893) into German.

I. The Peano Existence Theorem. *If $f(x,y)$ is continuous in a domain D and (ξ,η) is any point in D, then at least one solution of the differential equation* (1) *goes through (ξ,η). Every solution can be extended to the right and to the left up to the boundary of D.*

The last part of the statement of this theorem means that every solution has an extension that comes arbitrarily close to the boundary of D both to the right and the left as explained in 6.VII.

The proof of this theorem requires some additional concepts and lemmas.

II. Equicontinuity. A set $M = \{f, g, \ldots\}$ of continuous functions on the interval $J: a \le x \le b$ is called *equicontinuous* if for every $\varepsilon > 0$ there exists a number $\delta = \delta(\varepsilon)$ such that for any $f \in M$,

$$|f(x) - f(\bar{x})| < \varepsilon \quad \text{for} \quad |x - \bar{x}| < \delta, \quad (x, \bar{x} \in J). \tag{2}$$

It is important to note in this definition that for a given $\varepsilon > 0$, the same δ works for every function in the family M.

Example. Let M be a set of functions $f(x)$ that satisfy a Lipschitz condition with a common Lipschitz constant L; i.e.,

$$|f(x) - f(\bar{x})| \le L|x - \bar{x}| \quad \text{for} \quad x, \bar{x} \in J \quad \text{and} \quad f \in M.$$

The set M is equicontinuous. Clearly, one can set $\delta(\varepsilon) = \varepsilon/L$ here.

III. Lemma. *Let $J = [a, b]$ and let $A \subset J$ be a dense set of points in J. If the sequence of functions $f_1(x)$, $f_2(x)$, $\ldots$ is equicontinuous in J and converges for every $x \in A$, then it converges uniformly in J. Hence the limit $f(x)$ is continuous in J.*

A point set A is said to be dense in J if every subinterval of J contains at least one point of A (example: $A =$ the set of all rational numbers in J).

Proof. Given $\varepsilon > 0$, let $\delta = \delta(\varepsilon)$ be determined such that (2) holds for all functions f_n $(n \ge 1)$. Now partition the interval J into p closed subintervals $J_1, \ldots, J_p$ such that each J_i is less that δ in length. For each J_i, choose an $x_i \in J_i \cap A$ (there exists at least one such point for each i). By hypothesis, there exists an $n_0 = n_0(\varepsilon)$ such that

$$|f_m(x_i) - f_n(x_i)| < \varepsilon \quad \text{for} \quad m, n \ge n_0 \quad \text{and} \quad i = 1, \ldots, p.$$

Now let x be an arbitrary point of J and q be such that $x \in J_q$. It follows from the inequality $|x - x_q| < \delta$, property (2), and the above inequality that for $m, n \ge n_0$,

$$\begin{aligned} |f_m(x) - f_n(x)| \le |f_m(x) - f_m(x_q)| + |f_m(x_q) - f_n(x_q)| \\ + |f_n(x_q) - f_n(x)| < 3\varepsilon. \end{aligned}$$

This shows that the sequence $f_n(x)$ converges uniformly in J. ∎

As a further tool we require the

IV. Ascoli–Arzelà Theorem. *Every bounded and equicontinuous sequence of functions (f_n) in $C(J)$ contains a subsequence that converges uniformly in J. (Boundedness means that there exists a constant M such that $|f_n(x)| \le M$ for all $n \ge 1$ and $x \in J$.)*

Proof. Let $A = \{x_1, x_2, \ldots\}$ be a countable dense point set in J (for instance, the set of all rational numbers in J). The sequence of numbers $a_n = f_n(x_1)$ $(n = 1, 2, \ldots)$ is bounded and hence contains a convergent subsequence, say

$$f_{p_1}(x_1), f_{p_2}(x_1), f_{p_3}(x_1), \ldots.$$

The sequence of numbers $b_n = f_{p_n}(x_2)$ is likewise bounded and has, accordingly, a convergent subsequence, say

$$f_{q_1}(x_2), f_{q_2}(x_2), f_{q_3}(x_2), \ldots.$$

Note that (q_n) is a subsequence of (p_n). The sequence $c_n = f_{q_n}(x_3)$ is again a bounded sequence and therefore has a convergent subsequence

$$f_{r_1}(x_3), f_{r_2}(x_3), f_{r_3}(x_3), \ldots.$$

By continuing this process one obtains a series of sequences of the form

$$\begin{array}{lll} f_{p_1}, f_{p_2}, f_{p_3}, f_{p_4}, \ldots, & \text{which converges for} & x = x_1, \\ f_{q_1}, f_{q_2}, f_{q_3}, f_{q_4}, \ldots, & \text{which converges for} & x = x_1, x_2, \\ f_{r_1}, f_{r_2}, f_{r_3}, f_{r_4}, \ldots, & \text{which converges for} & x = x_1, x_2, x_3, \\ \ldots\ldots\ldots\ldots\ldots\ldots & & \end{array}$$

For each $k > 1$ the sequence that appears in the kth line is a subsequence of the sequence in the previous, $(k-1)$st, line and converges for $x = x_1, \ldots, x_k$. It follows that the diagonal sequence

$$f_{p_1}(x), f_{q_2}(x), f_{r_3}(x), \ldots.$$

converges for every $x = x_k$; i.e., for all $x \in A$, because it is a subsequence of the sequence in the kth line, at least from the kth term onward $(k = 1, 2, \ldots)$. The uniform convergence of this diagonal sequence now follows from Lemma III. ∎

We first give a proof of a weaker version of the Peano existence theorem.

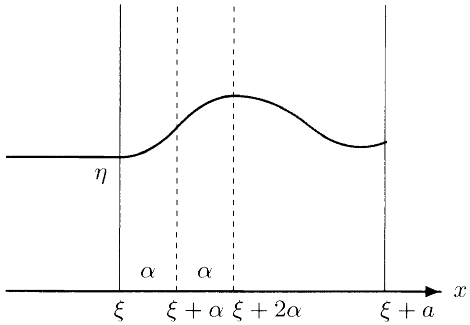
V. Theorem. *Let the function $f(x, y)$ be continuous and bounded in the strip $S = J \times \mathbb{R}$ with $J = [\xi, \xi + a]$, $a > 0$. Then there exists at least one function $y(x)$ defined and differentiable in J for which*

$$y' = f(x, y) \quad \text{in} \quad J, \qquad y(\xi) = \eta \tag{3}$$

(and, as a consequence, $y(x)$ is continuously differentiable in J).

Proof. The theorem is proved by finding a function $y(x) \in C(J)$ that satisfies the integral equation

$$y(x) = \eta + \int_\xi^x f(t, y(t))\, dt \quad \text{in} \quad J; \tag{4}$$



The approximate solutions $z_\alpha(x)$

see 6.I. With this objective in mind we construct, for every $\alpha > 0$, an "approximate solution" $z_\alpha(x) \in C(J)$ using the formula

$$z_\alpha(x) = \begin{cases} \eta & \text{for} \quad x \le \xi, \\ \eta + \displaystyle\int_\xi^x f(t, z_\alpha(t-\alpha)\, dt & \text{for} \quad x \in J. \end{cases} \tag{5}$$

This formula defines z_α unambiguously for $x \le \xi + a$. Indeed, if $\xi \le x \le \xi + \alpha$, then $t - \alpha \le \xi$ and $z_\alpha(t-\alpha) \equiv \eta$ in the integrand; i.e., the integral is well-defined. If $\xi + \alpha \le x \le \xi + 2\alpha$, then $t - \alpha \le x - \alpha \le \xi + \alpha$. Therefore, $z_\alpha(t-\alpha)$ is determined from the previous step; i.e., the integral is well-defined and so on. After a finite number of such steps we have constructed a continuous function $z_\alpha(x)$ that satisfies the integral equation (5). It follows from $|f| \le C$ that $|z'_\alpha(x)| \le C$; i.e., the functions $z_\alpha(x)$ satisfy the Lipschitz condition

$$|z_\alpha(x) - z_\alpha(\bar{x})| \le C|x - \bar{x}|$$

in J. If we denote the set of functions z_α in $C(J)$ (we consider only their restrictions to J) by M, then M is equicontinuous. Therefore, by the Ascoli–Arzelà theorem IV, the sequence $z_1(x), z_{1/2}(x), z_{1/3}(x), \ldots$ has a uniformly convergent subsequence $(z_{\alpha_n}(x))$ $(n = 1, 2, 3, \ldots; \alpha_n \to 0)$, which we relabel as $(z_n(x))$ to simplify the notation. Denote the continuous limit of this sequence by $y(x)$. Then by (5),

$$z_n(x) = \eta + \int_\xi^x f(t, z_n(t - \alpha_n))\, dt. \tag{6}$$

It follows now from the inequality

$$\begin{aligned} |z_n(t-\alpha_n) - y(t)| &\le |z_n(t-\alpha_n) - z_n(t)| + |z_n(t) - y(t)| \\ &\le C\alpha_n + |z_n(t) - y(t)| \end{aligned}$$

that $z_n(t - \alpha_n)$ converges uniformly in J to $y(t)$ and hence that $f(t, z_n(t-\alpha_n))$ converges uniformly in J to $f(t, y(t))$ (here we have used the fact that $f(x, y)$ is uniformly continuous on bounded sets). Therefore, passage to the limit under the integral sign in equation (6) is allowed, and equation (4) follows as a result. ∎

This special case of the Peano theorem can now be used in a manner that parallels the role of Theorem 6.I in the development of the Lipschitz case. But in other respects, the proof of theorem I differs essentially from that in the corresponding theorem 6.VII.

One first proves, exactly as in §6, that the existence theorem holds for a strip to the left of ξ and also for a rectangle. We formulate the latter case, which corresponds to Theorem 6.III.

VI. Theorem. *If f is continuous in the rectangle $R : \xi \le x \le \xi + a$, $|y - \eta| \le b$, then the initial value problem $y' = f(x, y)$, $y(\xi) = \eta$ has a solution $y(x)$ existing (at least) in the interval $\xi \le x \le \xi + \alpha$, where $\alpha = \min(a, b/A)$, $A = \max_R |f|$. A corresponding result holds for a rectangle to the left of ξ.*

This establishes the first part of the Peano existence theorem, that an integral curve passes through every point of D.

The second part, the proof of the assertion that every solution can be extended to the boundary, is more difficult in this setting in comparison to the proof in §6. The difficulties come from the fact that we do not have a uniqueness statement at our disposal.

Only the extendibility to the right will be discussed.

We first prove the following intermediate result.

(Z) *If ϕ is a solution in the interval $\xi \le x < b$ and A is a compact subset of D, then ϕ can be extended beyond A, i.e., there exists an extension $\bar{\phi}$ to the right with* $\operatorname{graph} \bar{\phi} \not\subset A$.

The distance from the set A to the boundary of D is positive, say $3\rho > 0$ (if $D = \mathbb{R}^2$, one can take $\rho = 1$). If $A_{2\rho}$ is the set of points whose distance from A is $\le 2\rho$, then $A_{2\rho}$ is likewise a compact subset of D, and hence f is bounded in $A_{2\rho}$, say $|f| \le C$. Denote by $R(x_0, y_0)$ the rectangle $x_0 \le x \le x_0 + \rho$, $|y - y_0| \le \rho$. Then $R(x_0, y_0) \subset A_{2\rho}$ for any $(x_0, y_0) \in A$.

If $\operatorname{graph} \phi \subset A$, we begin by extending ϕ to $[\xi, b]$ in the manner described in 6.VI.(a). Then we continue ϕ further to the right, by applying Theorem VI in the rectangle $R(b, \phi(b))$. This results in a solution in $b \le x \le b + \alpha =: b_1$ with $\alpha := \min(\rho, \rho/C)$. If this extension still lies entirely in A, then the process is repeated with $R(b_1, \phi(b_1))$, etc. Since at each step the interval of existence increases by a fixed number $\alpha > 0$ (as long as $\operatorname{graph} \phi \subset A$), one obtains in this manner a solution that extends beyond A after a finite number of steps. This proves (Z). ■

The remainder of the proof is straightforward. We consider a sequence (A_n) of compact sets with $A_n \subset A_{n+1} \subset D$ for all n and such that every compact $B \subset D$ is contained in one of the A_n (for example, let A_n be the set of points from D with distance $\ge \dfrac{1}{n}$ from the boundary of D and distance $\le n$ from the origin). Let ϕ with $\phi(\xi) = \eta$ be a solution in an interval to the right of ξ that does not approach the boundary of D. Then $\overline{\operatorname{graph}\phi}$ is a compact subset of D; i.e., $\operatorname{graph}\phi \subset A_p$ for a suitable p.

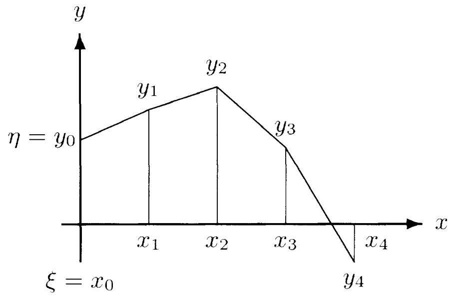
We continue ϕ beyond A_p using (Z) and call the extension ϕ_p; it exists in an interval $J_p = [\xi, b_p]$. We now construct ϕ_{p+1}: If ϕ_p does not belong entirely to A_{p+1}, then we set $\phi_{p+1} = \phi_p$; if ϕ_p does lie in A_{p+1}, then we continue the function ϕ_p to the right until it leaves A_{p+1} and call the continuation ϕ_{p+1}. It exists in an interval $[\xi, b_{p+1}]$ with $b_{p+1} \geq b_p$. Continuing inductively in this manner, one obtains a sequence of functions (ϕ_n) such that ϕ_n is defined in $[\xi, b_n]$ and (b_n) is a monotone increasing sequence of numbers. If $p \leq n < m$, then $\phi_n \equiv \phi_m$ in J_n.

Thus, by Lemma 6.V, there exists exactly one solution y defined in $[\xi, b)$, where $b = \lim_{n\to\infty} b_n$ ($b = \infty$ allowed), with the property that $y\,|_{J_n} = \phi_n$ for every $n \geq p$. This function y is the extension to the right of the original solution ϕ mentioned in the conclusion of the theorem. Clearly, graph y is not contained entirely in any A_n and hence is not contained in any compact subset of D. ∎

Remark on constructive proofs. The proof of Theorem V is different from that of 6.I in one important respect. In 6.I it was possible to calculate explicitly a sequence of successive approximations that converges to the solution (this is the essence of the contraction principle). Here, on the other hand, the initial value problem has several solutions in general, and it cannot be expected that a sequence of approximations like the one we constructed even has a limit, let alone tends toward any particular one of these solutions. Now one applies the Ascoli–Arzelà theorem to a sequence of such approximations. It says that there *exists* at least one convergent subsequence. However, no procedure is given that would allow the particular subsequence to be identified. An existence proof like the earlier one in §6 is called a constructive proof. By contrast, the proof of the Peano existence theorem given in this section is nonconstructive.

In the proof of the Peano existence theorem that we have just given the approximate solutions z_α were computed from the integral relation (5). Another frequently used method of obtaining approximations is

VII. The Euler–Cauchy Polygon Method. In this method, polygonal approximations u_α ($\alpha > 0$) to the solution to the initial value problem (3) are constructed in the following manner. Let $x_i = \xi + \alpha i$ ($i = 0, 1, 2, \ldots$). For $\xi = x_0 \leq x \leq x_1$ we set $u_\alpha(x) = \eta + (x - x_0)f(\xi, \eta)$; i.e., u_α is the straight line through the point $(\xi, \eta) = (x_0, y_0)$ with slope $f(x_0, y_0)$. In the interval $x_1 \leq x \leq x_2$, u_α is the straight line through the point $(x_1, y_1) := (x_1, u_\alpha(x_1))$ with slope $f(x_1, y_1)$. In general, one arrives after p steps at a point (x_p, y_p) with $y_p = u_\alpha(x_p)$ and defines then for $x_p \leq x \leq x_{p+1}$ that u_α is the straight line through (x_p, y_p) with slope $f(x_p, y_p)$. The advantage to this construction is two-fold. It is based on a relatively simple idea that is easily carried out numerically. However, the final step in the proof (passing to the limit to get the solution) is more difficult.



The Cauchy polygon method

VIII. Exercise. Prove that the Volterra integral equation

$$y(x) = g(x) + \int_0^x k(x, t, y(t))\, dt$$

has at least one continuous solution in $J = [0, a]$, provided that the function $g(x)$ is continuous in J and the "kernel" $k(x, t, z)$ is continuous for $0 \le t \le x \le a$, $-\infty < z < \infty$ and satisfies a growth condition $|k(x, t, z)| \le L(1 + |z|)$.

Hint: Let $C = \max |g(x)|$ and $D = \{v \in C(J) : |v(x)| \le \rho(x) \text{ in } J\}$, where ρ is determined by $\rho' = L(1+\rho)$, $\rho(0) = C$. If the integral equation is written in the form $u = Tu$, then $T(D) \subset D$. Now apply the Schauder fixed point theorem 7.XII.

Application to an Elliptic Problem. Show that if f is continuous in $J \times \mathbb{R}$ and satisfies $|f(x, y)| \le L(1 + |y|)$, then the initial value problem (6.8),

$$L_\alpha y = f(x, y) \quad \text{in} \quad J, \quad y(0) = \eta, \quad y'(0) = 0,$$

where L_α is the operator defined in 6.XII and $\alpha > 0$, has at least one solution.

As in 6.XIV, this result leads to an existence theorem for rotationally symmetric solutions of the elliptic differential equation $\Delta u = f(|x|, u)$. In particular, if f is continuous in $[0, \infty) \times \mathbb{R}$ and satisfies an estimate $|f(r, y)| \le L(r)(1 + |y|)$, where $L(r)$ is a continuous function in $[0, \infty)$, then there exists an entire, rotationally symmetric solution for every initial value η.

IX. Divergence of the Successive Approximations. If the right-hand side $f(x, y)$ is not Lipschitz continuous in y, then, as a general rule, the sequence of functions generated by the method of successive approximation does not converge to a solution. The example

$$y' = 2x - 2\sqrt{y_+}, \quad y(0) = 0 \quad \text{with} \quad y_+ = \max\{y, 0\}$$

shows that this behavior can also occur even if the solution is uniquely determined.

Indeed, beginning with $y_0 = 0$ one obtains $y_1 = x^2$, $y_2 = 0$, ..., and in general, $y_{2n} = 0$, $y_{2n+1} = x^2$. Nevertheless, since f is monotone decreasing in y, the uniqueness of the solution to the right follows from the uniqueness theorem 9.X.

Exercise. (a) If $y_0 = \alpha x^2$ ($\alpha > 0$) is chosen as a starting value in the method of successive approximations for the above equation, then one obtains $y_1 = \phi(\alpha)x^2$, hence $y_n = \alpha_n x^2$, where α_n is defined inductively by $\alpha_0 = \alpha$, $\alpha_{n+1} = \phi(\alpha_n)$ ($n = 0, 1, 2, \ldots$). Determine ϕ and show that ϕ has exactly one fixed point $\bar{\alpha}$ ($\bar{\alpha} = \phi(\bar{\alpha})$), and compute the fixed point. This gives a solution $y = \bar{\alpha}x^2$ to the problem. A more difficult problem is to show that the convergence is alternating for $0 < \alpha_0 < \bar{\alpha}$: $0 < \alpha_0 < \alpha_2 < \cdots < \bar{\alpha} < \cdots < \alpha_3 < \alpha_1 < 1$ and that $\lim \alpha_n = \bar{\alpha}$.

(b) Analyze the above problem "to the left" in a similar manner, i.e., consider for $z(x) := y(-x)$ the problem

$$z' = 2x + 2\sqrt{z_+} \quad \text{for} \quad x \geq 0,\ z(0) = 0$$

and show that starting with $z_0 = \beta_0 x^2$ and using the method of successive approximations, one obtains $z_n = \beta_n x^2$, where $\beta_{n+1} = \psi(\beta_n)$, $\lim \beta_n = \bar{\beta}$ (independent of $\beta_0 \in \mathbb{R}$), and that $z = \bar{\beta}x^2$ is the unique solution.

Hint: Show that the condition of Rosenblatt in 6.XI is satisfied (one can assume that $z \geq 2x^2$). The uniqueness and convergence of the iteration procedure for every starting value $z_0(x)$ can then be obtained from the theorem in 6.XI.

Supplement: Methods of Functional Analysis

We briefly develop some concepts and theorems from functional analysis in order to deepen our understanding of the Peano existence theorem and its proof. Fundamental to this effort is the concept of

X. Compactness. A subset A of a normed linear space B is called *compact* if every sequence (x_n) in A has a convergent subsequence with limit in A. The set $A \subset B$ is called *relatively compact* if $\bar{A}$ is compact.

In $\mathbb{R}^n$ a set is compact if and only if it is closed and bounded. In the general case, a compact set is always closed and bounded, but the converse is not always true.

An operator $T : D \to B$ with $D \subset B$ is called a *compact operator* in D if $T(D)$ is relatively compact.

If T is a compact operator and if it is possible to find approximating solutions, then the equation $x = Tx$ has a solution. This goes as follows. An equation $x = Tx$ is called *approximately solvable* if for every $\varepsilon > 0$, there exists $x \in D$ with $\|x - Tx\| < \varepsilon$.

Fixed Point Theorem. *Let the operator $T : D \to B$ be continuous and compact in D, where D is a closed subset of the normed space B. If the equation*

$$x = Tx$$

is approximately solvable in D, then there exists a solution in D.

The *proof* is elementary. There exists a sequence (x_n) from D with $x_n - Tx_n \to 0$. The sequence $(y_n) = (Tx_n)$ has a convergent subsequence, since $T(D)$ is relatively compact. If we denote this subsequence again by (y_n) to simplify the notation, then $y_n = Tx_n \to y \in B$ and hence $x_n = y_n + (x_n - Tx_n) \to y$. Since D is closed, it follows that $y \in D$. Then because of the continuity of T, $Tx_n \to Ty$, from which $y = Ty$ follows. ∎

XI. Example. Let J be a compact interval and M a subset of the Banach space $C(J)$ with the maximum norm.

If M is bounded and equicontinuous, then M is relatively compact.

This statement expresses precisely the content of the Ascoli–Arzelà theorem IV in the terminology of compact sets.

The Peano existence theorem, in the form of Theorem V, can be derived using the fixed point theorem X. One sets $D = B = C(J)$ and defines T to be the integral operator

$$(T\phi)(x) = \eta + \int_\xi^x f(t, \phi(t))\,dt \quad \text{for} \quad \phi \in B.$$

The following must now be shown:

(a) T is continuous in B;

(b) T is compact in B;

(c) the equation $x = Tx$ is approximately solvable.

The reader should carry out this proof. The approximations $z_\alpha(x)$ (cf. (5)) can be used to verify (c).

The proof can be made even simpler if one makes use of the following, significantly deeper, tool from functional analysis.

XII. The Schauder Fixed Point Theorem. *Let D be a closed and convex set in a Banach space B and let $T : D \to B$ be a continuous and compact operator in D with $T(D) \subset D$. Then T has at least one fixed point in D.*

A set D is called *convex* if, whenever $a, b \in D$, the line segment $\overline{ab} = \{x = \lambda a + (1-\lambda)b : 0 \le \lambda \le 1\}$ also lies in D.

A proof of the Schauder fixed point theorem is given in D.XII.

In order to derive the Peano existence theorem from the Schauder fixed point theorem, one has only to verify the two properties XI.(a), (b).

XIII. A Proof Based on Zorn's Lemma. Theorem I was derived from Theorem V by constructing the extensions explicitly. This result can also be proved using a theorem from set theory, Zorn's lemma. We introduce some additional concepts. A set M, or more precisely the pair $(M, \le)$, is called an ordered set if $\le$ is a transitive, antisymmetric relation in M (i.e., $x \le y$ and $y \le z$ implies $x \le z$; $x \le y$ and $y \le x$ implies $x = y$). A subset $N \subset M$ is called totally ordered if every pair of elements of N is comparable (either $x \le y$ or $y \le x$ for all $x, y \in N$). An element $m \in M$ is called an upper bound of N if $x \le m$ holds for all $x \in N$. The element $m \in M$ is called a maximal element of M if, except for m itself, there does not exist an $x \in M$ with $m \le x$.

Zorn's Lemma. *If $(M, \le)$ is ordered and if every totally ordered subset of M has an upper bound in M, then there exists a maximal element in M.*

We give here a brief indication of how Theorem I can be proved using this lemma. Let M be the set consisting of the graphs of *all* solutions of the given initial value problem, ordered by inclusion $\subset$. A set of solutions with the uniqueness property (E) of 6.V is a totally ordered subset of M. It follows from Lemma 6.V that this set of solutions has an upper bound. Thus Zorn's lemma can be applied and there exists a maximal element, i.e., a nonextendible solution. The proof that this solution extends to the boundary in both directions now follows, word for word, as in 6.VII.

XIV. Delay-Differential Equations. Let $\tau(x)$ be a given function in $C(J)$, $J = [\xi, \xi + a]$, with $0 \le \tau(x) \le b$. The equation

$$y'(x) = f(x, y(x - \tau(x))) \quad \text{for} \quad x \in J \tag{7}$$

is called a delay-differential equation. It is necessary to specify the function $y(x)$ in the interval $J_- = [\xi - b, \xi]$ as the "initial value"; otherwise, the right-hand side of (7) would not be defined for x close to ξ if, e.g., $\tau(x) = b$. Thus an initial condition for (7) reads

$$y(x) = \phi(x) \quad \text{for} \quad x \in J_- = [\xi - b, \xi], \tag{8}$$

where ϕ is a given function.

The functions z_α constructed for the proof of Theorem V are, in fact, solutions of just such an initial value problem with delay arguments where $\phi(x) = \eta$ and $\tau(x) = \alpha$.

Theorem. *We consider the initial value problem (7), (8), where f is continuous in the strip $S = J \times \mathbb{R}$ and τ is continuous in J with $0 \le \tau(x) \le b$.*

(a) *If $\tau(x) > 0$ in J, then there exists exactly one solution.*

(b) *If f satisfies a Lipschitz condition in S,*

$$|f(x, y) - f(x, z)| \le L|y - z| \quad \text{with} \quad L \ge 0,$$

then there exists exactly one solution, and it can be obtained by successive approximation.

(c) *If f is bounded in S, then there exists at least one solution.*

The *proof* of (a) proceeds like the proof in V. Theorem (b) corresponds to Theorem 6.I and is proved in exactly the same way. Theorem (c) corresponds to the Peano existence theorem V; the proof can be carried over. ∎

XV. Exercise. Let the initial value problem

$$y'(x) = y(x-\alpha) \quad \text{for} \quad x > \alpha, \quad y(x) = 1 \quad \text{for} \quad 0 \le x \le \alpha,$$

for a linear delay-differential equation be given, where $\alpha > 0$ is a given constant. Show that the solution can be represented in the form

$$y(x) = \sum_{k=0}^{\infty} a_k |(x - ka)_+|^k \quad \text{with} \quad s_+ = \max(0, s).$$

Determine the a_k.

Show that the solution approaches the solution to the initial value problem

$$y' = y, \quad y(0) = 1$$

as $\alpha \to 0$ and that the convergence is uniform in bounded intervals.

XVI. An Elementary Proof of the Peano Existence Theorem. We sketch a constructive proof of the Peano existence theorem that does not use a compactness argument.

Suppose that the assumptions of Theorem V hold; in particular, let $|f(x,y)| \le A$. For $h > 0$ let

$$f_h(\bar{x}, \bar{y}) := \max\{f(x,y) : \bar{x} \le x \le x + h, \, \bar{y} - 3Ah \le y \le y + Ah\}.$$

We apply the polygon method (cf. VII), first with step size h and the function f_h (instead of f), and then with step size $h/2$ and the function $f_{h/2}$. Denote the polygonal curves obtained in this manner by $y(x)$ and $z(x)$, respectively. Then it follows that $z \le y$ in J. This can be proved using an induction argument on the grid points of y. If $\bar{x}$ is a grid point and $z(x) \le y(x)$ for $x \le \bar{x}$ and $z(\bar{x}) \le y(\bar{x}) - 2Ah$, then the inequality $z \le y$ also holds up to the next grid point $\bar{x}+h$. On the other hand, if $y(\bar{x})-2Ah < z(\bar{x}) \le y(\bar{x})$, then $f_{h/2}(\bar{x}, z(\bar{x})) \le f_h(\bar{x}, y(\bar{x}))$, and from this it follows that $z \le y$ up to the point $\bar{x} + h/2$, and then, using a similar argument, up to $\bar{x} + h$.

If the polygon method is applied in the manner described above for $h = 2^{-n}$, then one obtains a monotone decreasing sequence of polygonal curves y_n. The proof that the limit of these sequences is a solution of the initial value problem is carried out in the usual way by going over to an integral equation. Note that the solution that is obtained in this way is the maximal integral (§ 9); cf. Walter (1971).

§ 8. Complex Differential Equations. Power Series Expansions

In this section z, w denote complex numbers and $w(z)$, $f(z,w)$ complex-valued functions of one and two complex variables, respectively. We begin with a few definitions and facts from the theory of functions of complex variables.

I. Properties of Holomorphic Functions. A continuous function $g(z) : G \to \mathbb{C}$, where G is an open set in $\mathbb{C}$, is called *holomorphic* (also *analytic*) *in* G if the (complex) derivative $g'(z) = \lim_{h \to 0} (g(z+h) - g(z))/h$ exists and is continuous in G. We denote by $H(G)$ the complex vector space of holomorphic functions in G. If the closed disk $Z : |z - z_0| \leq a$ belongs to G, then every function $g \in H(G)$ has an absolutely and uniformly convergent power series expansion

$$g(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \quad \text{in} \quad Z.$$

Similarly, a function $f(z, w)$ is holomorphic in an open set $D \subset \mathbb{C}^2$ (written more briefly $f \in H(D)$) if f and the partial derivatives f_z and f_w are continuous in D. If this is the case and if the disk product $Z : |z - z_0| \leq a$, $|w - w_0| \leq b$ lies in D, then $f(z, w)$ admits an absolutely and uniformly convergent power series representation

$$f(z, w) = \sum_{j,k=0}^{\infty} c_{jk} (z - z_0)^j (w - w_0)^k \ \text{ in } \ Z.$$

Thus holomorphic functions have continuous derivatives of all orders. The composition $g(z) := f(h_1(z), h_2(z))$ is holomorphic whenever the functions $f(z, w)$, $h_1(z)$, $h_2(z)$ are holomorphic (naturally, it is also assumed that the range of $(h_1(z), h_2(z))$ is contained in the domain of f), and the chain rule applies,

$$g'(z) = f_z(h_1(z), h_2(z)) h_1'(z) + f_w(h_1(z), h_2(z)) h_2'(z). \tag{1}$$

Suppose $h(z) \in H(G)$ and $z = \zeta(t)$, $t \in I = [\alpha, \beta]$, is a smooth path in G. Then the *path integral* along ζ is given by

$$\overset{\zeta}{\int} h(z)\, dz = \int_{\alpha}^{\beta} h(\zeta(t)) \zeta'(t)\, dt. \tag{2}$$

If G is *simply connected* and the path closed (i.e., $\zeta(a) = \zeta(b)$), then $\overset{\zeta}{\int} h(z)\, dz = 0$ by the *Cauchy integral theorem*. It follows easily that the function

$$H(z) := \int_{z_0}^{z} h(\zeta)\, d\zeta \quad (z_0 \in G) \tag{3}$$

is well-defined in G, i.e., that the integral has the same value for all paths that run in G from z_0 to z. The function H is an antiderivative of h, i.e., it is holomorphic in G, and

$$H'(z) = h(z) \quad \text{in} \quad G.$$

As in the real case, the fundamental theorem of calculus holds:

$$h(z) = h(z_0) + \int_{z_0}^{z} h'(\zeta)\, d\zeta \quad (z, z_0 \in G). \tag{4}$$

However, the mean value theorem of differential calculus is valid only in the following special form:

If the line segment connecting z_0 and z lies in G and if $|h'(\zeta)| \leq L$ on this segment, then

$$|h(z) - h(z_0)| \leq L|z - z_0|. \tag{5}$$

This result can be obtained immediately from (4) by taking the path of integration to be the straight line connecting the points z_0, z and making use of a general estimate for the path integral in (2),

$$\left|\int^{\zeta} h(z)\, dz\right| \leq \int_{\alpha}^{\beta} |h(\zeta(t))|\, |\zeta'(t)|\, dt \leq \max_{C} |h(z)| \cdot l(\zeta). \tag{6}$$

Here $l(\zeta) = \int_{\alpha}^{\beta} |\zeta'(t)|\, dt$ is the length of the path, and $C = \zeta(I)$ is the curve generated by the path.

These facts are assumed to be known; cf. Appendix C.

II. Existence and Uniqueness Theorem in $\mathbb{C}$. *Let the function $f(z, w)$ be holomorphic in a domain $D \subset \mathbb{C}^2$ that contains the set*

$$Z: \; |z - z_0| \leq a, \quad |w - w_0| \leq b$$

and let $|f| \leq M$ in Z.

Then there is a holomorphic solution $w(z)$ of the initial value problem

$$w' = f(z, w(z)), \quad w(z_0) = w_0, \tag{7}$$

and this solution exists at least in the disk $K: \; |z - z_0| < \alpha = \min(a, b/M)$.

If v and w are solutions of (7) existing in a domain G that contains the point z_0, then $v = w$ in G.

Proof. Suppose $|f_w| \leq L$ on the set

$$Z_1: \; |z - z_0| \leq \alpha, \; |w - w_0| \leq b.$$

Then f satisfies a Lipschitz condition with respect to w in Z_1:

$$|f(z, w_1) - f(z, w_2)| \leq L|w_1 - w_2|. \tag{8}$$

This follows from the mean value theorem in the form (5) (with w as the independent variable and z as a parameter). Because of (4), the initial value problem (7) is equivalent to the integral equation

$$w(z) = w_0 + \int_{z_0}^{z} f(\zeta, w(\zeta))\, d\zeta. \tag{9}$$

Let B be the space of functions $w(z)$ that are holomorphic and bounded in the disk K with the norm

$$\|u\| = \sup_{K} |u(z)| e^{-2L|z - z_0|}.$$

This space is complete and hence a Banach space; cf. Example 5.III.(d). Let D_T be the set of all $u \in B$ with $|u(z) - w_0| \le b$. The operator T,

$$Tu = w_0 + \int_{z_0}^{z} f(\zeta, w(\zeta))\, d\zeta,$$

is defined for $u \in D_T$, and the solutions to the initial value problem (7) are precisely the fixed points of T. We are going to show:

(a) T maps D_T into itself;

(b) T satisfies a Lipschitz condition in D_T with constant $\frac{1}{2}$.

To prove (a), if $u \in D_T$, we have

$$|(T(u)(z) - w_0| = \left| \int_{z_0}^{z} f(\zeta, u(\zeta))\, d\zeta \right| \le M|z - z_0| \le \alpha M \le b. \tag{10}$$

The argument for (b) is similar to the one in 6.I. A straight line path of integration is chosen: $\zeta(t) = z_0 + \theta \cdot t$, $0 \le t \le |z - z_0|$, where θ is the unit vector $(z - z_0)/|z - z_0|$. Then using (8), (6), and $|\zeta'| = 1$ one obtains

$$\begin{aligned} |(Tu)(z) - (Tv)(z)| &\le \left| \int_{z_0}^{z} \{f(\zeta, u) - f(\zeta, v)\}\, d\zeta \right| \\ &\le L \int_0^{|z - z_0|} |u(\zeta(t)) - v(\zeta(t))| \mathrm{e}^{-2Lt} \mathrm{e}^{2Lt}\, dt \\ &\le L\|u - v\| \int_0^{|z - z_0|} \mathrm{e}^{2Lt}\, dt \\ &\le \tfrac{1}{2} \mathrm{e}^{2L|z - z_0|} \|u - v\|. \end{aligned}$$

Statement (b), $\|Tu - Tv\| \le \frac{1}{2}\|u - v\|$ for $u, v \in D_T$, follows from this inequality.

Now, by the contraction principle 5.IX, T has exactly one fixed point w in D_T. It is the limit $w(z) = \lim u_n(z)$ in the sense of uniform convergence in K of a sequence of successive approximations (u_n), which can be constructed by first setting $u_0(z) = w_0$ (for instance) and then in succession setting

$$u_{n+1} = Tu_n, \quad \text{in full,} \quad u_{n+1}(z) = w_0 + \int_{z_0}^{z} f(\zeta, u_n(\zeta))\, dt. \tag{11}$$

Uniqueness. (i) Let v be another solution of (7) and let $\alpha' \le \alpha$ be chosen such that $|v(z) - w_0| < b$ in the disk $K' : |z - z_0| < \alpha'$. From the above proof (with K' instead of K) it follows that $v = w$ in K'.

(ii) Now let v, w be solutions in the domain G with $v(z_1) \ne w(z_1)$. If z_0 and z_1 are connected by a smooth path $z = \zeta(s)$ $(0 \le s \le l)$ that lies in G, then there exists a maximal $s' < l$ with $v(\zeta(s)) = w(\zeta(s))$ for $0 \le s \le s'$. Since v and w have the same "initial value" at the point $z' = \zeta(s')$, it follows from (i) with z' in place of z_0, that v and w agree in a neighborhood of z'. This contradiction to the maximality of s' shows that $v = w$ holds in G. Note that this result also follows from (i) and the identity theorem for holomorphic functions. ∎

III. Power Series Expansions. Like all holomorphic functions, the uniquely determined solution $w(z)$ of the initial value problem (7) can be expanded in a power series

$$w(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \quad \text{for} \quad |z - z_0| < \alpha. \tag{12}$$

Determining some coefficients of this expansion is an efficient numerical procedure, at least for z close to z_0. The following two methods can be used to this end:

Method 1. Beginning with the identity $w'(z) = f(z, w(z))$, the higher order derivatives can be calculated, one after another, by differentiation:

$$\begin{aligned} w' &= f, \\ w'' &= f_z + w' f_w, \\ w''' &= f_{zz} + 2w' f_{zw} + w'' f_w + w'^2 2 f_{ww}, \\ &\text{etc.} \end{aligned} \tag{13}$$

The coefficients

$$a_n = \frac{w^{(n)}(z_0)}{n!} \tag{14}$$

are obtained by inserting the values (z_0, w_0) into (13).

Method 2 (Power Series Ansatz). Substituting the expansion (12) into the right-hand side of the differential equation

$$f(z, w) = \sum_{j,k=0}^{\infty} c_{jk} (z - z_0)^j (w - w_0)^k$$

and differentiating termwise for the left-hand side leads to the relation

$$\sum_{j=1}^{\infty} j a_j (z - z_0)^{j-1} = \sum_{j,k=0}^{\infty} c_{jk} (z - z_0)^j \left(\sum_{n=1}^{\infty} a_n (z - z_0)^n \right)^k . \tag{15}$$

Equating coefficients of like terms gives a recursion formula for the a_i. This method is often easier to carry out than the first.

Naturally, the power series ansatz can also be used for differential equations in the real domain if the right-hand side is analytic.

Example. A special Riccati Equation (Johann Bernoulli, 1694).

$$y' = x^2 + y^2, \quad y(0) = 1.$$

The ansatz

$$y(x) = \sum_{j=0}^{\infty} a_j x^j$$

leads to the identity

$$\sum_{j=1}^{\infty} j a_j x^{j-1} = x^2 + \left(\sum_0^{\infty} a_j x^j \right)^2 = x^2 + \sum_{j=0}^{\infty} x^j \sum_{k=0}^{j} a_k a_{j-1}$$

and, when coefficients of like powers are set equal, to the relation

$$(j+1) a_{j+1} = \sum_{k=0}^{j} a_k a_{j-1} \quad \{+1 \quad \text{for } j = 2\}. \tag{16}$$

Incorporating the initial condition, one obtains for the first few terms

$$\begin{array}{rrcll} & & & & a_0 = 1; \\ j = 0: & a_1 & = & a_0^2, & a_1 = 1; \\ 1: & 2a_2 & = & 2a_0 a_1, & a_2 = 1; \\ 2: & 3a_3 & = & 2a_0 a_2 + a_1^2 + 1, & a_3 = \frac{4}{3}; \\ 3: & 4a_4 & = & 2a_0 a_3 + 2a_1 a_2, & a_4 = \frac{7}{6}. \end{array}$$

Thus the power series expansion begins with

$$y(x) = 1 + x + x^2 + \frac{4x^3}{3} + \frac{7x^4}{6} + \cdots.$$

One can see immediately from the recursion formula for the a_i that all a_i are positive. An inspection of the first few terms suggests that in fact $a_i \geq 1$ ($i \geq 0$). This inequality is valid for small i and can be proved in general by induction (exercise!). Therefore, we have

$$y(x) > 1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x} \quad \text{for} \quad x > 0. \tag{17}$$

This inequality shows that the solution can exist to the right at most up to the point $x = 1$. Better methods for the estimation of solutions are derived in § 9; cf. Example 9.V.

IV. Exercises. (a) Give the first few terms in the power series expansion (up to the fourth power) of the solution of the initial value problem

$$y' = e^x + x \cos y, \quad y(0) = 0.$$

(b) Determine the first terms in the power series expansion $y = \sum a_k x^k$ for the solution to the initial value problem

$$y' = x^3 + y^3, \quad y(0) = 1.$$

Determine the power series expansion of the solution $u = \sum b_k x^k$ of

$$u' = u^3, \quad u(0) = 1$$

and show that $a_k \geq b_k$. From this derive an upper bound for the number a, where $[0, a)$ is the maximal interval of existence of the solution y to the right.

V. Growth Estimates. The growth of solutions of complex differential equations can be estimated using the following theorem, which depends on a result from § 9.

Theorem. *Let w be a holomorphic solution of the initial value problem* (7)*: $w'(z) = f(z, w)$ in G, $w(z_0) = w_0$ (G a convex domain), where*

$$|f(z, w)| \le h(|z - z_0|, |w - w_0|)$$

($h(t, y)$ is real-valued and locally Lipschitz continuous in y). Then $|w(z) - w_0| \le \phi(|z - z_0|)$, where ϕ is differentiable and satisfies

$$\phi'(t) \ge h(t, \phi(t)), \quad \phi(0) \ge 0.$$

Hint: Apply Theorem 9.VIII to $v(t) = |w(z_0 + e^{i\alpha}t) - w_0|$ with $0 \le \alpha < 2\pi$, $w(t) = \phi(t)$. Note that $v'_+(t) \le h(t, v(t))$; cf. B.IV.

§ 9. Upper and Lower Solutions. Maximal and Minimal Integrals

In this section all quantities are again real-valued.

I. Lemma. *Let $\phi(x)$, $\psi(x)$ be differentiable in the half-open interval J_0 : $\xi < x \le \xi + a$ ($a > 0$) and suppose that $\phi(x) < \psi(x)$ in an interval $\xi < x < \xi + \varepsilon$ ($\varepsilon > 0$). Then one of the following two cases holds:*

(a) *$\phi < \psi$ in J_0;*

(b) *there exists an $x_0 \in J_0$ such that $\phi(x) < \psi(x)$ for $\xi < x < x_0$ and*

$$\phi(x_0) = \psi(x_0) \quad \text{and} \quad \phi'(x_0) \ge \psi'(x_0). \tag{1}$$

The *proof* is simple. If (a) does not hold, then there exists a first point $x_0 > \xi$ where $\phi(x_0) = \psi(x_0)$. Since $\phi < \psi$ to the left of x_0, the left-sided difference quotients at the point x_0 satisfy

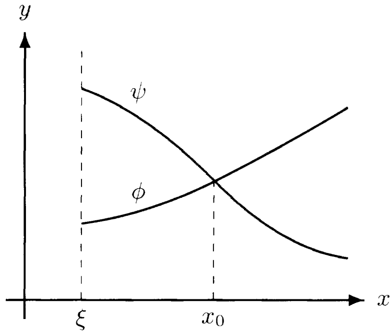
$$\frac{\phi(x_0) - \phi(x_0 - h)}{h} > \frac{\psi(x_0) - \psi(x_0 - h)}{h} \quad \text{for} \quad h > 0. \tag{2}$$

The second inequality in (1) follows by taking the limit in (2) as $h \to 0+$. ■

(c) It is possible to weaken the hypotheses in the above result. The conclusion of the lemma remains true if the functions $\phi(x)$, $\psi(x)$ are assumed only to be continuous. In this case the relations (1) in (b) are replaced by

$$\phi(x_0) = \psi(x_0), \; D^-\phi(x_0) \ge D^-\psi(x_0), \; \text{and} \; D_-\phi(x_0) \ge D_-\psi(x_0). \tag{1'}$$

Here D^-, D_- are the left-sided upper and lower *Dini derivatives*, defined in section B.I of the Appendix. Clearly, (1′) also follows from (2).



II. Defect. The *defect* $P\phi$ of a function ϕ with respect to the differential equation $y' = f(x, y)$ is the function

$$P\phi = \phi' - f(x, \phi). \tag{3}$$

The defect tells "how close" ϕ is to satisfying the differential equation. Solutions $y(x)$ are characterized by $Py = 0$ (the defect is 0).

Every theorem on differential inequalities for the interval $J = [\xi, \xi + a]$ has a counterpart for the interval $J_- = [\xi - a, \xi]$ to the left of ξ. The corresponding defect inequality is obtained by reflection about the point ξ, and the proof is carried out by reducing the case "to the left" to the earlier case; cf. 6.II.(d). The reflection transformation, $\bar{\phi}(x) = \phi(2\xi - x)$, $\bar{f}(x, y) = -f(2\xi - x, y)$, introduces a minus sign to the defect,

$$(\bar{P}\bar{\phi})(x) := \bar{\phi}'(x) - \bar{f}(x, \bar{\phi}(x)) = -(P\phi)(2\xi - x) \tag{4}$$

with the result that in the theorem for J_-, the differential inequalities are reversed. A first example is the

III. Comparison Theorem. *Let the functions $\phi(x)$, $\psi(x)$ be differentiable in $J_0 : \xi < x \le \xi + a$ and let the following hold:*

(a) $\phi(x) < \psi(x)$ *for* $\xi < x < \xi + \varepsilon$ $(\varepsilon > 0)$;

(b) $P\phi < P\psi$ *in* J_0.

Then

$$\phi < \psi \quad \text{in} \quad J_0.$$

There is no assumption on f. The theorem remains true for continuous functions ϕ, ψ with $D^-\phi$, $D^-\psi$ or $D_-\phi$, $D_-\psi$ instead of ϕ', ψ'.

The *proof* is based on showing that case (b) of Lemma I cannot occur. Suppose $\phi(x_0) = \psi(x_0)$. Then at x_0,

$$\phi'(x_0) = P\phi + f(x_0, \phi(x_0)) < P\psi + f(x_0, \psi(x_0)) = \psi'(x_0)$$

holds because of hypothesis (b). Thus (1) or (1′) certainly does not hold. ∎

We formulate the corresponding theorem for an interval lying to the left of the point ξ.

Corollary. *If ϕ, ψ are differentiable in $J_0^- = [\xi - a, \xi)$ and satisfy*
(a) $\phi(x) < \psi(x)$ *for* $\xi - \varepsilon < x < \xi$ $(\varepsilon > 0)$,
(b) $P\phi > P\psi$ *in* J_0^-,
then it follows that

$$\phi < \psi \quad \text{in} \quad J_0^- .$$

IV. Upper Solutions, Lower Solutions. Let $f(x, y)$ be defined in D, $D \subset \mathbb{R}^2$ arbitrary. The function $v(x)$ is called a *lower solution* (or *subsolution*) and $w(x)$ is called an *upper solution* (or *supersolution*) of the initial value problem

$$y' = f(x, y) \quad \text{in} \quad J = [\xi, \xi + a], \quad y(\xi) = \eta, \tag{5}$$

if it is differentiable in J and

$$\begin{aligned} v' &< f(x, v) \quad \text{in } J, \quad v(\xi) \le \eta, \qquad \textit{lower solution,} \\ w' &> f(x, w) \quad \text{in } J, \quad w(\xi) \ge \eta, \qquad \textit{upper solution.} \end{aligned} \tag{6}$$

Naturally, it is assumed that $\operatorname{graph} v \subset D$ and $\operatorname{graph} w \subset D$. These concepts were introduced (in a somewhat more general way) by Perron (1915). An upper solution runs above a solution, a lower solution below. More precisely: If v is a lower solution, w an upper solution, and if y is a solution to the initial value problem (5), then

$$v(x) < y(x) < w(x) \quad \text{in} \quad J_0 : \xi < x \le \xi + a. \tag{7}$$

These inequalities follow immediately from Theorem III. Setting $\phi = v$, $\psi = y$, one obtains $P\phi < 0 = P\psi$; thus (b) holds. If $v(\xi) < \eta = y(\xi)$, then clearly (a) holds, and if $v(\xi) = \eta = y(\xi)$, then by (6), $v'(\xi) < f(\xi, \eta) = y'(\xi)$. Therefore, $v < y$ for $\xi < x < \xi + \varepsilon$ $(\varepsilon > 0)$. The second inequality in (7) is proved in a similar manner. ∎

Upper and Lower Solutions to the Left. If J_- is an interval to the left $\xi - a \le x \le \xi$, then the conditions that define lower and upper solutions in the initial value problem (5) read

$$\begin{aligned} v' &> f(x, v) \quad \text{in} \quad J_-, \quad v(\xi) \le \eta, \qquad \textit{lower solution,} \\ w' &< f(x, w) \quad \text{in} \quad J_-, \quad w(\xi) \ge \eta, \qquad \textit{upper solution} \end{aligned} \tag{6'}$$

and the conclusion is

$$v(x) < y(x) < w(x) \quad \text{in} \quad J_0^- : \xi - a \le x < \xi. \tag{7'}$$

A common method of determining upper and lower solutions is to make small changes in f to find functions f_1, f_2 such that

$$f_1(x, y) < f(x, y), \quad \text{respectively} \quad f_2(x, y) > f(x, y),$$

and such that the corresponding initial value problems for the equations

$$v' = f_1(x, v), \text{ respectively } w' = f_2(x, w),$$

can be explicitly solved.

V. Example. We consider Bernoulli's example from 8.III,

$$y' = x^2 + y^2, \quad y(0) = 1.$$

For positive x, $y^2 < x^2 + y^2$, thus using $f_1(x, y) = y^2$, one obtains a lower solution

$$v' = v^2, \quad v(0) = 1 \quad \Rightarrow \quad v(x) = \frac{1}{1-x}.$$

We see from this that the solution exists to the right at most up to the point $x = 1$. Thus one can assume that $0 \le x < 1$ and obtain an upper solution by setting $f_2(x, y) = y^2 + 1$:

$$w' = w^2 + 1, \quad w(0) = 1 \quad \Rightarrow \quad w(x) = \tan\left(x + \frac{\pi}{4}\right).$$

Thus without computational effort one obtains the estimates

$$\frac{1}{1-x} < y(x) < \tan\left(x + \frac{\pi}{4}\right)$$

as well as an estimate for the asymptote $x = b$ ($y \to \infty$ as $x \to b-$)

$$0.78 < \frac{\pi}{4} \le b \le 1.$$

A significantly better upper solution can be obtained using the ansatz

$$w_1 = \frac{1}{1 - cw} \qquad (c > 1).$$

The inequality $w_1' > w_1^2 + x^2$ is equivalent to $c - 1 > (1 - cx)^2 x^2$ for $0 \le x < 1/c$. For example, one can take $c = 17/16$. It follows that

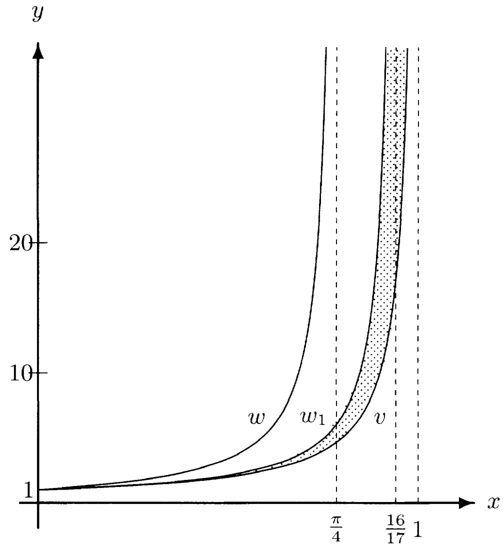
$$w_1(x) = \frac{16}{16 - 17x} \quad \text{and} \quad 0.94 < \frac{16}{17} < b \le 1.$$

The solution remains in the domain illustrated in the figure. Better bounds for b are obtained by computing the solution y with the Lohner algorithm, which gives *exact bounds*; cf. XVI. Let us assume that $y_0 < y(a) < y_1$. Then a lower solution v for $x > a$ is obtained from $v(a) = y_0$, $v' = a^2 + v^2$; i.e.,

$$v(x) = a \tan(ax + c) \quad \text{with} \quad a^2 + c = \arctan y_0/a.$$

Solving $ax + c = \pi/2$, the asymptote $x = b_1$ of v is obtained; it is an upper bound for b. Now we can construct an upper solution w from $w(a) = y_1$, $w' = b_1^2 + w^2$, i.e.,

$$w(x) = b_1 \tan(b_1 x + d) \quad \text{with} \quad b_1 a + d = \arctan y_1/b_1.$$



Here the asymptote b_0 of w, a lower bound of b, comes from the equation $b_1 x + d = \pi/2$. For example, the Lohner algorithm produces

$$a = \frac{31}{32} = 0.96875, \quad y(a) \in 942.81425692\,{}^{308}_{082},$$

and one obtains in the way just described,

$$b \in (b_0, b_1) = 0.96981\,06539\,3\,{}^{13}_{04}$$

(the notation used above represents an interval; for example, $4.5\,{}^{356}_{130}$ is the interval $(4.5130, 4.5346)$). The success of the method, here shown impressively, depends on good upper and lower solutions. For example, it works with the tangent function used above for the equation $y' = g(x) + y^2$.

VI. Maximal and Minimal Solutions. Definition and Theorem. *If $f(x, y)$ is continuous in a domain D, then the initial value problem*

$$y' = f(x, y), \quad y(\xi) = \eta \quad \text{with} \quad (\xi, \eta) \in D$$

has two solutions $y_*(x)$, $y^*(x)$ that come arbitrarily close to the boundary of D both to the left and the right and that have the following property:

If $y(x)$ is any solution of the initial value problem, then

$$y_*(x) \le y(x) \le y^*(x) \tag{8}$$

(each of the inequalities holds as long as the functions involved are defined). The solution $y_*(x)$ is called the minimal solution (minimal integral), and $y^*(x)$ is called the maximal solution (maximal integral) of the initial value problem.

Proof. First assume that $f(x,y)$ is continuous and bounded in the strip $J \times \mathbb{R}$, $J = [\xi, \xi + a]$. Let $y(x)$ be a solution of the initial value problem (5) and $w = w_n(x)$ be a solution of the initial value problem

$$w' = f(x,w) + \frac{1}{n} \quad \text{in } J, \qquad w(\xi) = \eta + \frac{1}{n} \quad (n = 1, 2, 3, \ldots) \tag{9}$$

(if (9) has more than one solution, then one of them is chosen). If Theorem III is first applied with $\phi = y$, $\psi = w_{n+1}$ and then with $\phi = w_{n+1}$, $\psi = w_n$, then one obtains the inequalities

$$y(x) < w_{n+1}(x) < w_n(x) \quad \text{in } J.$$

Thus the sequence w_n is monotone decreasing and has a limit

$$y^*(x) = \lim_{n \to \infty} w_n(x) \geq y(x). \tag{10}$$

In fact, by Lemma 7.III, this limit is uniform, since (9) implies that $|w_n'| \leq C := \sup |f| + 1$, and hence the sequence of functions (w_n) is equicontinuous. Thus one can pass to the limit as $n \to \infty$ under the integral sign in the integral equation equivalent to (9),

$$w_n(x) = \eta + \frac{1}{n} + \frac{1}{n}(x - \xi) + \int_\xi^x f(t, w_n(t))\, dt, \tag{9'}$$

and obtain that the limit satisfies the integral equation

$$y^*(x) = \eta + \int_\xi^x f(t, y^*(t))\, dt.$$

Therefore, the function $y^*(x)$ is a solution to the initial value problem, and because of (10), it has the property $y(x) \leq y^*(x)$ proposed in (8).

In a corresponding manner the minimal integral $y_*(x)$ can be obtained as the limit of a related sequence $v_n(x)$; here one has to replace the term $1/n$ by $-1/n$ in both places in (9).

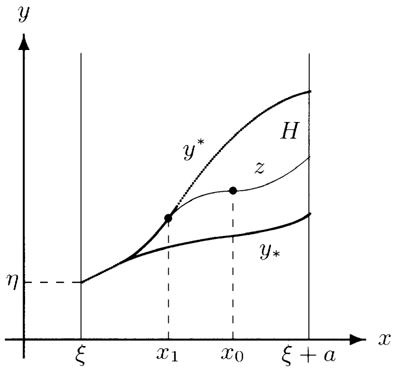
This proves the theorem under the special assumptions. The general theorem can now be derived from this result following the procedure described in § 7 for the Peano existence theorem. This procedure has to be modified to the extent that instead of an arbitrary solution, the maximal solution is chosen in each step of the extension. ∎

VII. Remarks. (a) If f is continuous, then the initial value problem is uniquely solvable if and only if $y_*(x) = y^*(x)$.

(b) If the initial value problem is not uniquely solvable, then the whole region between the maximal solution and the minimal solution is filled up with solutions. To be more explicit:

If $f(x,y)$ is continuous in the strip $J \times \mathbb{R}$, $J = [\xi, \xi + a]$, and if $y_*(x)$ is the minimal and $y^*(x)$ the maximal integral of the initial value problem (5), then a solution of the initial value problem (5) goes through each point of the set

$$H = \{(x,y) : x \in J,\ y_*(x) \leq y \leq y^*(x)\}.$$



The proof is an exercise. *Hint:* If a solution z, which starts at a point $(x_0, z_0) \in H$, meets, e.g., y^* to the left of x_0, say, $z(x_1) = y^*(x_1)$, $x_1 < x_0$, then one may set $z(x) = y^*(x)$ in $[\xi, x_1]$.

VIII. Theorem. *Let f be continuous in the rectangle $R = J \times [b, c]$, $J = [\xi, \xi + a]$. Let the functions v, w be differentiable in J and let*

$$\begin{aligned} &v' \le f(x, v) \quad \text{in} \quad J, \quad v(\xi) \le \eta \qquad \text{or} \\ &w' \ge f(x, w) \quad \text{in} \quad J, \quad w(\xi) \le \eta. \end{aligned} \tag{11}$$

If $y_* \in C^1(J)$ is the minimal solution and $y^* \in C^1(J)$ the maximal solution of the initial value problem (5)*, then*

$$v \le y^* \quad \text{or} \quad w \ge y_* \quad \text{in} \quad J.$$

This remains true for continuous functions with Dv, Dw (D any Dini derivative). It it is assumed that the graphs of all four functions lie in R.

For the *proof* one extends f as a continuous and bounded function to the strip $J \times \mathbb{R}$ (cf. the proof of 6.III) and determines $w_n(x)$ using equation (9). By Theorem III, $v < w_n$ (note that $Dv \le f(t, v)$ implies $D^- v \le f(t, v)$ by B.II). Since the w_n converge to $y^*(x)$, it follows that $v \le y^*$ (it is easy to see that y^* is also the maximal solution relative to the extended function f). The second inequality $y_* \le w$ is handled in a similar manner. ∎

Corollary on Upper and Lower Solutions. *If f is continuous and the solution y of the initial value problem* (5) *is unique, then upper and lower solutions can be characterized by the weak inequalities* (11) *(with $\le$ instead of $<$). It follows immediately from the above theorem that the inequalities* (11) *imply $v \le y \le w$ in J.*

Example.

$$y' = \sqrt{|y|}, \quad y(0) = 0.$$

The maximal solution y^* and the minimal solution y_* are given by

$$y^*(x) = \frac{x^2}{4}, \quad y_*(x) = 0 \qquad \text{for} \quad x \geq 0,$$

$$y^*(x) = 0, \qquad y_*(x) = -\frac{x^2}{4} \quad \text{for} \quad x \leq 0$$

(see Example 2 of 1.V). Make a sketch of all solutions for the proof.

We conclude the discussion of differential inequalities with a variant of Theorem III, in which it is assumed that the function f satisfies a local Lipschitz condition in y (the reader will recall the definition from 6.IV). This case applies to most applications, in particular those where f_y is continuous. Our proof is independent of the earlier results, and the conclusion is sharper with respect to strict inequalities.

IX. Comparison Theorem. *Let $f : D \to \mathbb{R}$ satisfy a local Lipschitz condition in y. Let ϕ, ψ be differentiable in $J = [\xi, \xi + a]$, and let*
(a) $\phi(\xi) \leq \psi(\xi)$,
(b) $P\phi \leq P\psi$ *(which is $\phi' - f(x, \phi) \leq \psi' - f(x, \psi)$)* in J.
Then $\phi \leq \psi$ in J, and with respect to strict inequalities,

$$\phi < \psi \text{ in } J \quad \text{or} \quad \phi = \psi \text{ in } [\xi, c], \;\; \phi < \psi \text{ in } (c, \xi + a] \; (c \in J).$$

Corollary. *For an interval $J_- = [\xi - a, \xi]$ to the left of ξ,* (a) $\phi(\xi) \leq \psi(\xi)$ *and* (b$'$) $P\phi \geq P\psi$ *in J_- implies $\phi < \psi$ in J_- or $\phi = \psi$ in $[c, \xi]$, $\phi < \psi$ in $[\xi - a, c]$ with $c \in J_-$.*

Proof. We write $f(\phi)$, $f(\psi)$ as an abbreviation for $f(x, \phi(x))$, $f(x, \psi(x))$. There exists $L > 0$ such that $|f(\phi) - f(\psi)| \leq L|\phi - \psi|$ in J; cf. Exercise 6.IX. Hence the function $w = \psi - \phi$ satisfies

$$w' = \psi' - \phi' = P\psi + f(\psi) - P\phi - f(\phi) \geq f(\psi) - f(\phi) \geq -L|w|.$$

Assume that $w(d) < 0$ for some $d \in J$ and that $I = [b, d] \subset J$ is the largest interval to the left of d where $w \leq 0$. In this interval $w' \geq Lw$, hence

$$\left(w(x)\mathrm{e}^{-Lx}\right)' = (w' - Lw)\mathrm{e}^{-Lx} \geq 0.$$

Thus the function $w(x)\mathrm{e}^{-Lx}$ is monotone increasing in I and therefore negative in I. This shows that $b = a$ and $w(a) < 0$, which is a contradiction. Hence $\phi \leq \psi$ in J.

Now assume that $w(d) > 0$ and that $I' = [d, d'] \subset J$ is the largest interval to the right of d where $w \geq 0$. In I' we have $w' \geq -Lw$, which implies

$$(\mathrm{e}^{Lx} w)' = \mathrm{e}^{Lx}(w' + Lw) \geq 0.$$

Reasoning as before, we conclude first that $\mathrm{e}^{Lx} w(x)$ is increasing in I', then that $w(d') > 0$, and finally that d' must be the right endpoint $\xi + a$ of J. This completes the proof of the theorem. The corollary is left to the reader's care. ∎

Consequences of the Comparison Theorem. (a) *Upper and lower solutions.* Let y be a solution of the initial value problem

$$y' = f(x, y) \quad \text{in} \quad J, \qquad y(\xi) = \eta, \tag{5}$$

where f is locally Lipschitz continuous in y. Then upper and lower solutions can be characterized by the inequalities (11) (with equality permitted), that is to say, these inequalities imply

$$v \le y \le w \qquad \text{in} \quad J.$$

But now the theorem gives a stronger statement than the one in section VIII. For example, if $y = v$ in $J' = [\xi, c]$, then $v'(x) = f(x, v(x))$ holds in J'. Thus the strong inequality $v < y$ holds in $J_0 = (\xi, \xi + a]$ if either $v(\xi) < \eta$ or if there exists a sequence (x_n) in J_0 with $\lim x_n = \xi$ such that $v'(x_n) < f(x_n, v(x_n))$ (since v is usually given in applications and y is the unknown solution, these conditions can be checked). This follows readily from Theorem IX and applies also for upper solutions.

(b) A uniqueness theorem for the initial value problem (5) follows immediately from this theorem.

(c) If y and z are solutions of the differential equation and if $y(x_0) < z(x_0)$, then it follows that $y < z$ in the common interval of existence of both solutions (to the right and to the left of x_0).

Part of the previous theorem, the uniqueness to the right, can be proved under the weaker assumption that f only satisfies a one-sided Lipschitz condition

$$f(x, y) - f(x, z) \le L(y - z) \quad \text{for} \quad y > z. \tag{12}$$

While the usual Lipschitz condition says that the difference quotients

$$\frac{f(x, y) - f(x, z)}{y - z}$$

lie between $-L$ and L, the one-sided condition implies only that they are $\le L$. The proof of the following theorem is left as an exercise for the reader. *Hint:* Study the previous proof

X. Theorem. *Let f satisfy a local one-sided Lipschitz condition of the form* (12). *Then the inequality $\phi \le \psi$ in $J = [\xi, \xi + a]$ follows from*

(a) $\phi(\xi) \le \psi(\xi)$,

(b) $P\phi \le P\psi$ in J.

In particular, a uniqueness theorem "to the right" holds for the corresponding initial value problem.

Remark. This theorem applies in particular to functions f that are monotone decreasing in y. Such functions satisfy a one-sided Lipschitz condition with $L = 0$.

Exercise. Give an example of a monotone decreasing, continuous function $f(y)$, for which the stronger hypotheses of Theorem IX are not satisfied.

XI. Exercise. Let the function $f(x,y)$ be defined for $x \geq 0$ by

$$f(x,y) = \begin{cases} 2x & \text{for} \quad y \geq x^2, \quad x \geq 0, \\ 2y/x & \text{for} \quad |y| < x^2, \quad x > 0, \\ -2x & \text{for} \quad y \leq -x^2, \quad x \geq 0. \end{cases}$$

Is f continuous in $[0,\infty) \times \mathbb{R}$? Find all of the solutions to the initial value problem

$$y' = f(x,y) \quad \text{for} \quad x \geq 0,\ y(0) = \eta.$$

For which values of η is the solution unique? Give the maximal and minimal solutions in the nonuniqueness case.

XII. Exercise. Construct upper and lower solutions for the following initial value problems

(a) $y' = x^3 + y^3, \quad y(0) = 1$.

(b) $y' = x + \sqrt{1+y^2}, \quad y(0) = 1$.

In the case (a), if $0 \leq x < a$ is the maximal interval of existence to the right, calculate two bounds $a_1 \leq a \leq a_2$ with $a_2 - a_1 < 0.05$. Compare this to Exercise 8.IV.(b).

Supplement: The Separatrix

Here we consider differential equations

$$y' = f(x,y) \quad \text{for} \quad x \geq 0 \tag{13}$$

that have the following (at first imprecisely formulated) property: There exists a special global solution ϕ (i.e., one that exists in $[0,\infty)$) which is distinguished by the property that the solutions above ϕ and the solutions below ϕ form two classes of solutions such that within the classes solutions have similar behavior for large x, whereas two solutions taken from different classes have completely different behavior. We give two examples to illustrate this property.

Example 1. $y' = x - 1/y \qquad (y > 0)$.

Example 2. $y' = x^3 + y^3$.

In the first example, the special solution ϕ is the only bounded global solution. The solutions above ϕ tend to ∞ as $x \to \infty$, while every positive solution beneath ϕ exists only in a finite interval $[0,b)$ and tends to 0 as $x \to b-$.

In the second example, ϕ is the only global solution. The solutions above ϕ tend to $+\infty$ and those below ϕ tend to $-\infty$ as x approaches the right endpoint of the (finite) maximal interval of existence.

In this connection, the solution curve $C = \operatorname{graph} \phi$ is called a *separatrix:* It "separates" solutions with different behavior. Various definitions of the separatrix can be found in the literature.

Differential inequalities provide a powerful tool for dealing with such questions.

XIII. Existence of Global Solutions. Solutions of (13) that exist in $[0,\infty)$ are called *global solutions.* Let f be continuous and locally Lipschitz continuous with respect to y (for instance, suppose f_y is continuous) on a set $D: x \geq 0$, $\alpha < y < \beta$ $(-\infty \leq \alpha < \beta \leq \infty)$. If v, w are two functions with the properties $v \leq w$ and $Pv \leq 0 \leq Pw$ in $[0,\infty)$, then every solution y of (13) with $v(0) \leq y(0) \leq w(0)$ lies between v and w on $[0,\infty)$. This follows immediately from Theorem IX. The following theorem with reversed differential inequalities also follows from Theorem IX, but the proof is less trivial.

Theorem. *Let v and w be functions that are differentiable in $[0,\infty)$ and satisfy the inequalities*

$$v \leq w \quad \text{and} \quad Pw \leq 0 \leq Pv \quad \text{in} \quad [0,\infty).$$

Then the differential equation $y' = f(x,y)$ has a global solution ϕ with $v \leq \phi \leq w$ for $x \geq 0$.

Proof. Let y_n denote the (unique) solution of the initial value problem

$$y' = f(x,y), \quad y(n) = w(n) \quad (n = 1,2,3,\ldots). \tag{A$_n$}$$

It follows from applying Corollary IX twice on the interval $[0,n]$, which lies to the left of the point $\xi = n$, that the inequalities $v \leq y_n \leq w$ hold in $[0,n]$. In particular, $y_{n+1}(n) \leq w(n) = y_n(n)$, and hence, again using Corollary IX, $y_{n+1} \leq y_n$ in $[0,n]$.

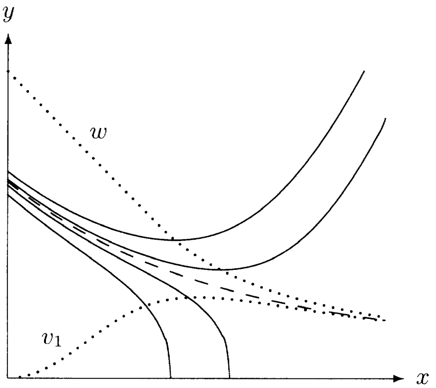
We fix a number $a > 0$. If $n > a$, then the inequalities $v \leq y_{n+1} \leq y_n \leq w$ hold in the interval $[0,a]$. Since the sequence (y_n) is monotone decreasing, $\phi(x) := \lim y_n(x)$ exists in $[0,a]$ and satisfies the inequalities $v \leq \phi \leq w$ there. The set $M_a = \{(x,y) : 0 \leq x \leq a,\ v(x) \leq y \leq w(x)\}$ is compact, and the function f is bounded on M_a, say $|f| \leq L$. Thus, for $n > a$, the functions y_n are Lipschitz continuous with Lipschitz constant L in the interval $[0,a]$. Uniform convergence on the interval $[0,a]$ now follows from Lemma 7.III.

Passing to the limit in the corresponding integral equation

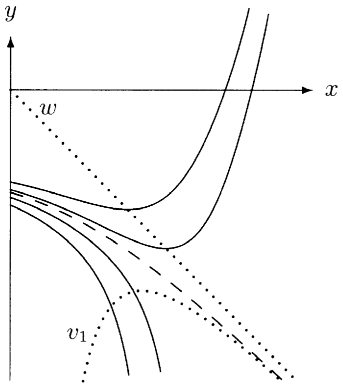
$$y_n(x) = y_n(0) + \int_0^x f(t_n, y_n(t))\,dt$$

shows that ϕ is a solution of (13) in the interval $[0,a]$. Since a is arbitrary, the rest of the theorem follows. ∎

Remark. If there exist sequences (x_n), (x_n') tending to ∞, with $(Pv)(x_n) > 0$, $(Pw)(x_n') < 0$, then the strong inequalities $v < \phi < w$ hold in $[0,\infty)$. This follows again from Theorem IX.



Example 1 with $v_1 = 1/(x + x^{-2})$, w as in the text (the dashed curve is ϕ)



Example 2 with $v_1 = -x - 1/3x^2$, $w = -x$

Note that one can also use the sequence of solutions $z = z_n$ defined by

$$z' = f(x, z), \quad z(n) = v(n) \quad (n = 1, 2, 3, \ldots) \tag{B$_n$}$$

in the construction of a global solution. One obtains $v \le z_n \le z_{n+1} \le y_{n+1} \le y_n \le w$ in $[0, n]$, $n = 1, 2, 3, \ldots$

We apply this theorem to the two examples given at the beginning of this supplement.

Example 1. The inequality $Pv \ge 0$ holds for $v = \mathrm{e}^{-x}$, and the inequality $Pw \le 0$ is satisfied by the function

$$w(x) = \begin{cases} 2 - x & \text{for} \quad 0 \le x \le 1, \\ 1/x & \text{for} \quad x > 1. \end{cases}$$

Thus there exists a global solution ϕ with $\mathrm{e}^{-x} < \phi(x) < 1/x$. The reader should show that $v_1 = 1/(x + x^{-2})$ is also a lower bound.

Example 2. One can chose $v = -(x + 1)$, $w = -x$, as can be easily seen. Thus there exists a global solution ϕ that satisfies the inequality $-(x + 1) < \phi(x) < -x$. The reader should show that $v_1 = -x - 1/3x^2$ is a better lower bound.

XIV. Uniqueness. The characterization of the distinctive global solution differs from case to case, and it seems impossible to give a sufficiently general uniqueness theorem. Instead, we illustrate how to proceed in certain specific cases. Suppose that f_y is continuous in D. If ϕ and ψ are two global solutions with $\phi < \psi$, then by the mean value theorem, the difference $u = \psi - \phi > 0$ satisfies

$$u' = \psi' - \phi' = f(x, \psi) - f(x, \phi) = f_y(x, y^*)u(x), \tag{14}$$

where $\phi(x) < y^* < \psi(x)$. We consider again the two examples and show how one can use (14) to derive a uniqueness result.

Example 1. Here the global solutions ϕ, ψ are bounded; that is, there exists $L > 0$ such that $0 < \phi < \psi < L$ holds in $[0, \infty)$ and hence $f_y(x, y) = 1/y^2 > 1/L^2 := \alpha$. Thus

$$u' \geq \alpha u, \quad \text{which implies that} \quad u(x) \geq u(0)\mathrm{e}^{\alpha x}.$$

But u is bounded by assumption. This contradiction proves the assertion made at the beginning that there is only one bounded global solution. The estimate $v_1 < \phi < 1/x$ (see XIII) shows that ϕ behaves like $1/x$ as $x \to \infty$ and that

$$\frac{1}{x} - \frac{1}{x^4} < \phi(x) < \frac{1}{x}.$$

Example 2. Let y be a solution and $y(a) \geq -a$ for some $a \geq 0$. It is easy to see from the differential equation that there exists $b > a$ with $y(b) > 0$. Since the solution of the initial value problem $v' = v^3$, $v(b) = y(b)$ is a lower solution to y and since there exists $c < \infty$ such that $v(x) \to \infty$ as $x \to c$, it follows that $y(x)$ is not a global solution.

If ϕ and ψ are global solutions with $\phi < \psi$, then accordingly, $\psi(x) < -x$. Thus in (14) we have $f_y(x, y^*) = 3{y^*}^2 > 3x^2$, and hence $u = \psi - \phi \geq \delta \exp(x^3)$, where $\delta = u(0) > 0$.

In particular, $z = -\phi \geq \delta \exp(x^3)$. Since z satisfies $z' = z^3 - x^3$, we have $z' > \frac{1}{2} z^3$ for large x. This implies in a manner similar to the case above that $z = -\phi$ exists only in a finite interval $[a, b)$ and tends to ∞ as $x \to b$. Therefore, there exists only one global solution. This proves the assertion made at the beginning for the second example.

Remark. In the two examples, each local solution can be extended to the left to $-\infty$ (proof follows by considering the sign of f). Thus the distinguished solution exists in $\mathbb{R}$, and the separatrix $C = \operatorname{graph} \phi$ divides the xy-plane into two domains G_1 (above C) and G_2 (below C). If y is a solution of the differential equation with the initial value $y(\xi) = \eta$, then the location of (ξ, η) instantly gives information about the qualitative behavior of the solution y for increasing positive x.

XV. Exercises. (a) Show that the qualitative behavior from Example 1 also holds for the differential equation

$$y' = x^\alpha - y^{-\beta} \quad (y > 0) \quad \text{with} \quad \alpha > 0,\ \beta > 0.$$

(b) Show that the same qualitative statements hold for the differential equation

$$y' = x^\alpha + |y|^\beta \operatorname{sgn} y \quad \text{with} \quad \alpha > 0,\ \beta > 1$$

as in Example 2 (which is the case $\alpha = \beta = 3$).

Hint. One can take $w = -x^{\alpha/\beta}$ and $v - -ae^x$ with appropriate $a > 0$. A better choice is $v = -(a + x^{1+\alpha})$.

(c) Denote the solution of the differential equation $y' = f(x, y)$ with initial value $y(0) = a$ by $y(x; a)$, and denote by $[0, b_a)$ with $0 < b_a \leq \infty$ the maximal interval of existence to the right. Prove the following

Theorem. *Let the function f be continuous and satisfy a local Lipschitz condition in y. Let A be the set of initial values a such that $y(x; a) \to \infty$ as $x \to b_a$. Assume that for every $b > 0$ there exist initial values $a \in A$ and $a' \notin A$ such that b_a, $b_{a'} > b$. Then*

$$\phi(x) = y(x; a_0) \quad \textit{with} \quad a_0 = \inf A$$

is a global solution of (13).

(d) Show that all solutions $y(x; a)$ of the differential equation $y' = \frac{\pi}{2}(1 + y^2)\cos x$ tend to $+\infty$ or $-\infty$ as $x \to b_a$, where $b_a < \frac{3}{2}\pi$, and compare this result with the assumptions on b_a in the preceding theorem.

(e) In the differential equation

$$y' = h(x) + g(y)$$

let h be continuous in $[0, \infty)$ and let g be locally Lipschitz continuous in $\mathbb{R}$. Further, let $g(y) \to \pm\infty$ as $y \to \pm\infty$, and let the integrals

$$\int_\alpha^\infty \frac{dy}{g(y)} \quad \text{and} \quad \int_{-\infty}^{-\alpha} \frac{dy}{g(y)}$$

be convergent (we assume $|g(y)| > 0$ for $|y| \geq \alpha$). Show that there exists a global solution.

Remark. In the *American Mathematical Monthly* 94 (1987), p. 694, one finds Example 1 as Problem 6551: "Prove that the differential equation $y' = x - 1/y$ has a unique solution in $[0, \infty)$ which is positive throughout and tends to zero at $+\infty$." In Vol. 96 (1989) three different solutions are given on pages 631–635 and 657–659, but no general method is suggested.

XVI. Computing the Separatrix. Theorem XIII not only establishes the existence of a separatrix ϕ, it can also be used to determine it numerically. By solving the initial value problems (A_n) and (B_n), one obtains upper and lower bounds for ϕ. R. Lohner (1988) developed an algorithm that gives *exact* upper and lower bounds for the solution of initial value problems (for systems). It applies to a large class of functions f and uses the programming languages PASCAL-XSC or ACRITH-XSL, which reflect advances in interval arithmetic. I owe Dr. Lohner many thanks for carrying out the calculations in the two examples. They lead to the following surprisingly good (and, as said, exact) bounds for the initial value $\phi(0)$:

Example 1: $v(x) = 1/(x + 1/x^2)$, $w(x) = 1/x$, (A_n), (B_n) with $n = 6$ gives

$$\phi(0) \in 1.28359\,87104\,63599\,52345\,264\,{}^{44}_{30}.$$

Example 2: $v(x) = -x - 1/3x^2$, $w(x) = -x$, (A_n), (B_n) with $n = 5$ gives

$$\phi(0) \in -0.66727\,09125\,44323\,65855\,63\,{}^{57}_{61}.$$

XVII. Exercise. Driver's Equation. The initial value problem

$$y'(x) = 1 + f(y) - f(x), \qquad y(0) = 0,$$

where $f : \mathbb{R} \to \mathbb{R}$ is continuous, has $y(x) = x$ as a solution. R.D. Driver posed the problem (*American Mathematical Monthly* 73 (1966), 783, advanced problem 5415) of determining whether this is, in general (i.e., for all f), the only solution. Some results on uniqueness were given in the solution section of the *Monthly* (76 (1969), 948–949), but much was left to be desired. Our treatment relies heavily on the results of this section on differential inequalities.

If y is a solution, then $z(x) := -y(-x)$ is a solution to the corresponding problem with $f_1(s) := f(-s)$. Therefore, we consider only solutions for $x > 0$. Prove the following:

(a) Replacing $f(s)$ by $f(s) + \text{const.}$ does not change the problem. Hence one may assume $f(0) = 0$.

(b) If y is a solution and $y(c) = c$ $(c > 0)$, then $z(x) = y(x + c) - c$ is a solution of the problem with f replaced by $f_c(s) = f(s + c) - c$.

(c) $y^*(x) = x$ is the maximal solution to the right, i.e., $y(x) \le x$ for $x \ge 0$ and every solution.

(d) Uniqueness to the right holds if f satisfies locally (i.e., in compact intervals) a one-sided Lipschitz condition $f(y) - f(z) \le L|y - z|$ for $y > z$.

(e) Uniqueness holds if $f = g - h$, where g and h are (weakly) increasing.

Hints: (c) Let $F(s) = \exp f(s)$. The differential equation is equivalent to $\exp(y' - 1) = F(y)/F(x)$. Using $\mathrm{e}^{s-1} \ge s$, one obtains $y' \le F(y)/F(x)$. Hence y is a subsolution to the problem with separated variables $z' = F(z)/F(x)$, $z(0) = 0$, which gives $y(x) \le z(x) = x$.

(e) Let f be increasing. Then by (c), $y' \le 1$. Consider the function

$$k(x) = \int_{y(x)}^{x} (1 - f(s))\, ds \Longrightarrow k'(x) = f(y)(y' - 1).$$

Assume $0 = f(0) \le f(x) < 1$ and $y(x) > 0$ for $0 \le x < b$. Since $k' \le 0$ and $k \ge 0$ in $[0, b]$, it follows that $k(x) = 0$ and hence $y(x) = x$ in $[0, b]$. Use (b) to show that $y(x) = x$ in $[0, c]$ implies $y(x) = x$ in $[0, c + \varepsilon]$, $\varepsilon > 0$.

In the general case $f = g - h$ we have $y' \ge 1 + g(y) - g(x)$. Since $z' = 1 + g(z) - g(x)$, $z(0) = 0$ implies $z(x) = x$, Theorem VIII shows that $y \ge z$. Use (c).

Our treatment follows that of G. Herzog and R. Lemmert, *Remarks on Driver's equation*, *Ann. Polon. Math.* LIX.2 (1994), 197–202. The authors also construct an example of nonuniqueness.

XVIII. Exercise. Prove the following:

Theorem. *Let v, w be two solutions of the differential equation $y'(x) = f(x,y)$ in the interval $J = [\xi, \xi+a]$. Assume that $v(\xi) < w(\xi)$ and that $f(x,y)$ is (weakly) increasing or decreasing in y. Then* (i) *$v \leq w$ in J and* (ii) *$w(x) - v(x)$ is increasing or decreasing, respectively.*

Chapter III
First Order Systems. Equations of Higher Order

§ 10. The Initial Value Problem for a System of First Order

I. Systems of Differential Equations. Direction Fields. By a first order system of differential equations (in explicit form) we mean a set of simultaneous equations of the form

$$\begin{aligned} y_1' &= f_1(x, y_1, \ldots, y_n) \\ \vdots \quad & \qquad \vdots \\ y_n' &= f_n(x, y_1, \ldots, y_n). \end{aligned} \tag{1}$$

Here the n functions $f_1(x, y_1, \ldots, y_n), \ldots, f_n(x, y_1, \ldots, y_n)$ are defined on a set D of $(n+1)$-dimensional $(x, y_1, \ldots, y_n)$-space $\mathbb{R}^{n+1}$. A vector function $(y_1(x), \ldots, y_n(x))$ is a *solution* (or an *integral*) of (1) in the interval J if the functions $y_\nu(x)$ are differentiable in J and if (1) is satisfied identically when they are substituted into the equation. Naturally, we require $(x, y_1(x), \ldots, y_n(x)) \in D$ for $x \in J$. Vector notation will be used whenever possible. We denote n-dimensional column vectors with boldface letters, as shown in the following:

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \quad \mathbf{y}(x) = \begin{pmatrix} y_1(x) \\ \vdots \\ y_n(x) \end{pmatrix}, \quad \mathbf{f}(x, \mathbf{y}) = \begin{pmatrix} f_1(x, \mathbf{y}) \\ \vdots \\ f_n(x, \mathbf{y}) \end{pmatrix}.$$

In this notation, statements of the type "$\mathbf{y}(x)$ is continuous, differentiable, ..." mean that each component y_ν is continuous, differentiable, ... $(\nu = 1, \ldots, n)$. Derivatives and integrals of a vector function $\mathbf{y}(x)$ are also defined componentwise:

$$\mathbf{y}'(x) = \begin{pmatrix} y_1'(x) \\ \vdots \\ y_n'(x) \end{pmatrix}, \quad \int_a^b \mathbf{y}(x)\,dx = \begin{pmatrix} \displaystyle\int_a^b y_1(x)\,dx \\ \vdots \\ \displaystyle\int_a^b y_n(x)\,dx \end{pmatrix}.$$

Written in vector notation, system (1) reads

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}). \tag{1'}$$

As in the case $n = 1$, equation (1′) has a geometric interpretation. The graph of a function $\mathbf{y}(x)$ represents a curve in $\mathbb{R}^{n+1}$. The vector function $\mathbf{f}(x, \mathbf{y})$ determines a direction field in D, which is defined as follows: To each point $(\bar{x}, \bar{\mathbf{y}}) \in D$ is associated a direction, given by the $(n+1)$-dimensional vector $(1, \mathbf{a})$ where $\mathbf{a} = \mathbf{f}(\bar{x}, \bar{\mathbf{y}})$, or equivalently, by the line $\mathbf{y} = \bar{\mathbf{y}} + (x - \bar{x})\mathbf{a}$. Solutions of the differential equation (1′) have the property that their graphs "fit on the direction field."

II. Initial Value Problem. An initial value problem for (1′) asks for a solution that passes through a given point $(\xi, \boldsymbol{\eta}) \in D$, that is, one that satisfies the initial conditions

$$y_\nu(\xi) = \eta_\nu \quad (\nu = 1, \ldots, n) \quad \text{or} \quad \mathbf{y}(\xi) = \boldsymbol{\eta} \quad \text{(in vector form).} \tag{2}$$

Initial value problem (1′), (2) is equivalent to the system of integral equations

$$\mathbf{y}(x) = \boldsymbol{\eta} + \int_\xi^x \mathbf{f}(t, \mathbf{y}(t))\,dt. \tag{3}$$

More precisely: Let $\mathbf{f}$ be continuous in D and $(x, \mathbf{y}(x)) \in D$ for $x \in J$. Suppose $\mathbf{y}(x)$ is differentiable in J and satisfies equation (1′) and the initial conditions (2). Then $\mathbf{y}'$ is continuous and $\mathbf{y}(x)$ satisfies the integral equation (3) for $x \in J$. Conversely, if $\mathbf{y}(x)$ is a *continuous* solution of (3) in J, then $\mathbf{f}(x, \mathbf{y}(x))$ is continuous in J. Hence $\mathbf{y}(x)$ is continuously differentiable and satisfies (1′) and (2). The notation $|\mathbf{a}|$ for the norm of a vector $\mathbf{a} \in \mathbb{R}^n$ was introduced in 5.III.(a). A special case is the Euclidean norm $|\mathbf{a}|_e = \sqrt{a_1^2 + \cdots + a_n^2}$.

III. Equivalence of Norms. Lemma. *All norms in $\mathbb{R}^n$ are equivalent; i.e., if $|\mathbf{a}|$, $|\mathbf{a}|^*$ are any two norms in $\mathbb{R}^n$, then there exist constants $\alpha > 0$, $\beta > 0$ such that*

$$\alpha|\mathbf{a}|^* \le |\mathbf{a}| \le \beta|\mathbf{a}|^* \quad \text{for all} \quad \mathbf{a} \in \mathbb{R}^n.$$

Proof. Clearly, it is sufficient to prove the theorem for the special case $|\mathbf{a}|^* = |\mathbf{a}|_e$. If $\mathbf{e}_\nu$ is the νth unit vector (i.e., $\mathbf{e}_\nu$ is 1 in the νth position, all other components are zero), then it follows from the expansion $\mathbf{a} = \sum a_\nu \mathbf{e}_\nu$ that

$$|\mathbf{a}| \le \sum |a_\nu \mathbf{e}_\nu| = \sum |a_\nu||\mathbf{e}_\nu| \le |\mathbf{a}|_e \sum |\mathbf{e}_\nu| = \beta |\mathbf{a}|_e.$$

This proves the second inequality in the lemma. It follows at once from this inequality that the function $\phi(\mathbf{x}) = |\mathbf{x}|$ is continuous (in the sense of the Euclidean norm). Indeed, by (5.2), we have

$$|\phi(\mathbf{x}) - \phi(\mathbf{y})| = \big||\mathbf{x}| - |\mathbf{y}|\big| \leq |\mathbf{x} - \mathbf{y}| \leq \beta|\mathbf{x} - \mathbf{y}|_e.$$

To prove the first inequality, let S be the unit sphere in $\mathbb{R}^n$; that is, S is the set of vectors with $|\mathbf{a}|_e = 1$. Since S is compact, the continuous function $\phi(\mathbf{x}) = |\mathbf{x}|$ assumes its minimum on S at a point $\mathbf{a}_0 \in S$. Because $\mathbf{a}_0 \neq 0$, we have

$$|\mathbf{a}| \geq |\mathbf{a}_0| = \alpha > 0 \quad \text{for all} \quad \mathbf{a} \in S.$$

Since an arbitrary vector $\mathbf{b} \neq 0$ can be written in the form $\mathbf{b} = c \cdot \mathbf{a}$ with $c = |\mathbf{b}|_e$ and $\mathbf{a} \in S$, the first inequality in the lemma follows:

$$|\mathbf{b}| = c|\mathbf{a}| \geq \alpha c = \alpha|\mathbf{b}|_e.$$

∎

IV. Lipschitz Condition. A vector function $\mathbf{f}(x, \mathbf{y})$ satisfies a *Lipschitz condition* with respect to $\mathbf{y}$ in D (with Lipschitz constant L) if

$$|\mathbf{f}(x, \mathbf{y}) - \mathbf{f}(x, \bar{\mathbf{y}})| \leq L|\mathbf{y} - \bar{\mathbf{y}}| \quad \text{for} \quad (x, \bar{\mathbf{y}}),\ (x, \bar{\mathbf{y}}) \in D. \tag{4}$$

It follows from Lemma III that the question whether $\mathbf{f}$ satisfies a Lipschitz condition is independent of the chosen norm. However, the magnitude of the Lipschitz constant L depends on the choice of the norm.

A function $\mathbf{f}$ is said to satisfy in D a *local Lipschitz condition* with respect to $\mathbf{y}$ if for every point $(x, \mathbf{y}) \in D$, there exists a neighborhood $U : |x - \bar{x}| < \delta$, $|\mathbf{y} - \bar{\mathbf{y}}| < \delta$ $(\delta > 0)$ such that $\mathbf{f}$ satisfies a Lipschitz condition in $D \cap U$. In general, the Lipschitz constant may vary from neighborhood to neighborhood.

V. Lemma. (a) *If D is convex and if $\mathbf{f}$ and all components of the Jacobian $\partial\mathbf{f}/\partial\mathbf{y} = (\partial f_\nu/\partial y_\mu)^n_{\mu,\nu=1}$ are continuous and bounded in D $(\mu, \nu = 1, \ldots, n)$, then $\mathbf{f}$ satisfies a Lipschitz condition with respect to $\mathbf{y}$ in D.*

(b) *If D is a domain and if $\mathbf{f}$ and $\partial\mathbf{f}/\partial\mathbf{y}$ are continuous in D, then $\mathbf{f}$ satisfies in D a local Lipschitz condition with respect to $\mathbf{y}$.*

(c) *If $\mathbf{f} \in C(D)$ satisfies in D a local Lipschitz condition in $\mathbf{y}$, then $\mathbf{f}$ satisfies a Lipschitz condition in $\mathbf{y}$ on compact subsets of D.*

Proof. (a) Applying the mean value theorem to $f_\nu(x, \mathbf{y})$, we obtain

$$f_\nu(x, \mathbf{y}) - f_\nu(x, \bar{\mathbf{y}}) = \sum_{\mu=1}^{n} \frac{\partial f_\nu(x, \mathbf{y}^*)}{\partial y_\mu}(y_\mu - \bar{y}_\mu),$$

where $(x, \mathbf{y}^*)$ is a point on the line segment connecting $(x, \mathbf{y})$ and $(x, \bar{\mathbf{y}})$. It follows that there exists a constant K such that

$$|f_\nu(x, \mathbf{y}) - f_\nu(x, \bar{\mathbf{y}})| \leq K \max_\mu |y_\mu - \bar{y}_\mu|, \tag{4'}$$

and hence $\mathbf{f}$ satisfies a Lipschitz condition with respect to the maximum norm $|\mathbf{a}| = \max_\nu |a_\nu|$. Part (b) is an immediate consequence of part (a), since a function from $C(D)$ is bounded in compact subsets of D. (c) If $K \subset D$ is compact and the proposition with respect to K is false, then there exist sequences $(x_k, \mathbf{y}_k)$, $(x_k, \mathbf{z}_k)$ in K with

$$|\mathbf{f}(x_k, \mathbf{y}_k) - \mathbf{f}(x_k, \mathbf{z}_k)| \geq k|\mathbf{y}_k - \mathbf{z}_k| \qquad (k = 1, 2, \ldots). \tag{$*$}$$

Because K is compact, we may assume that $(x_k, \mathbf{y}_k) \to (x_0, \mathbf{y}_0) \in K$, and since f is bounded in K, it follows from $(*)$ that $(x_k, \mathbf{z}_k)$ tends to the same point. This leads to a contradiction, since for large k, the points $(x_k, \mathbf{y}_k)$, $(x_k, \mathbf{z}_k)$ belong to a neighborhood of $(x_0, \mathbf{y}_0)$, where $\mathbf{f}$ satisfies a Lipschitz condition in $\mathbf{y}$. ∎

We now state and prove the basic

VI. Existence and Uniqueness Theorem. *Let $\mathbf{f}(x, \mathbf{y})$ be continuous in a domain $D \subset \mathbb{R}^{n+1}$ and satisfy a local Lipschitz condition with respect to $\mathbf{y}$ in D (this hypothesis is satisfied, for instance, if $\partial\mathbf{f}/\partial\mathbf{y} \in C(D)$). If $(\xi, \boldsymbol{\eta}) \in D$, then the initial value problem*

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(\xi) = \boldsymbol{\eta} \tag{5}$$

has exactly one solution. The solution can be extended to the left and right up to the boundary of D.

The line of reasoning used for the case $n = 1$ in §6 carries over. The following special case is proved first (compare with Theorem 6.I).

VII. Theorem. *Let $\mathbf{f}(x, \mathbf{y})$ be continuous and satisfy the Lipschitz condition (4) in $J \times \mathbb{R}^n$, $J = [\xi, \xi + a]$. Then there is exactly one solution to the initial value problem*

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(\xi) = \boldsymbol{\eta}. \tag{6}$$

The solution exists in of J.

Proof. The initial value problem (6) is equivalent to the integral equation

$$\mathbf{y}(x) = \boldsymbol{\eta} + \int_\xi^x \mathbf{f}(t, \mathbf{y}(t))\, dt \quad \text{in} \quad J, \tag{7}$$

which can be written, using the more concise operator notation, in the form

$$\mathbf{y} = T\mathbf{y} \quad \text{where} \quad (T\mathbf{z})(x) = \boldsymbol{\eta} + \int_\xi^x \mathbf{f}(t, \mathbf{z}(t))\, dt. \tag{7'}$$

The set of continuous vector-valued functions defined on J with the norm

$$\|\mathbf{z}\| = \max_J |\mathbf{z}(x)| \mathrm{e}^{-2Lx}$$

is a Banach space ($|\cdot|$ is the norm in $\mathbb{R}^n$ that appears in (4)). The operator T defined by (7′) maps this space into itself. The proof that T satisfies a Lipschitz condition with Lipschitz constant $\frac{1}{2}$ is the same as the one given in 6.I. In this proof, $y(x)$ is now a vector function, and the two simple facts in the following lemma are needed.

VIII. Lemma. *If $\mathbf{z}(x)$ is continuous in an interval $[a,b]$ and $|\cdot|$ is a norm in $\mathbb{R}^n$, then the scalar function $\phi(x) = |\mathbf{z}(x)|$ is continuous in $[a,b]$ and the inequality*

$$\left| \int_a^b \mathbf{z}(x)\, dx \right| \leq \int_a^b |\mathbf{z}(x)|\, dx$$

holds.

Proof. It follows from inequality (5.2) that

$$|\phi(x_k) - \phi(x)| \leq |\mathbf{z}(x_k) - \mathbf{z}(x)| \to 0 \quad \text{as} \quad x_k \to x;$$

therefore $\phi(x)$ is continuous.

Let us denote the integrals in the lemma by I_1 and I_2. Then the inequality $|I_1| \leq I_2$ must be proved. We consider a partition $P : a = x_0 < x_1 < \cdots < x_p = b$ and its measure of fineness $|P| = \max\{(x_i - x_{i-1}) : i = 1, \ldots, p\}$. Corresponding Riemann sums are given by

$$\sigma(P;\mathbf{z}) = \sum_{i=1}^{p} (x_i - x_{i-1})\mathbf{z}(x_i) \quad \text{for} \quad I_1 = \int_a^b \mathbf{z}(x)\, dx,$$

$$\sigma(P;|\mathbf{z}|) = \sum_{i=1}^{p} (x_i - x_{i-1})|\mathbf{z}(x_i)| \quad \text{for} \quad I_2 = \int_a^b |\mathbf{z}(x)|\, dx.$$

The triangle inequality implies

$$|\sigma(P;\mathbf{z})| \leq \sigma(P;|\mathbf{z}|). \tag{$\star$}$$

Now consider a sequence (P_k) of partitions with $\lim |P_k| = 0$. Then, by the Riemann definition of the integral,

$$\sigma(P_k;\mathbf{z}) \to I_1 \quad \text{and} \quad \sigma(P_k;|\mathbf{z}|) \to I_2$$

as $k \to \infty$. The inequality $|I_1| \leq I_2$ now follows from $\star$). ∎

The solutions of (6) are the fixed points of T. Since T is a contraction, Theorem VII follows from the Contraction Principle 5.IX. The solution is the limit of a uniformly convergent sequence of successive approximations

$$\mathbf{y}_{k+1} = (T\mathbf{y}_k)(x) = \boldsymbol{\eta} + \int_\xi^x \mathbf{f}(t, \mathbf{y}_k(t))\, dt \quad (k = 0, 1, 2, \ldots). \tag{8}$$

The first term $\mathbf{y}_0(x) \in C(J)$ can be arbitrarily chosen. ∎

The general theorem is now derived from this special case in a series of steps that are completely analogous to those in §6 for the one-dimensional case.

In §7 and §8 two additional existence theorems were proved. The extension of the proofs of these theorems to the n-dimensional case is also straightforward. Consequently, we will state these theorems without proof.

IX. Peano Existence Theorem. *If* $\mathbf{f}(x, \mathbf{y})$ *is continuous in the domain* D *and* $(\xi, \boldsymbol{\eta}) \in D$*, then the initial value problem* (5) *has at least one solution. Every solution can be extended to the left and right up to the boundary of* D.

Remark. The results on upper and lower solutions and on maximal and minimal solutions obtained in §9 do not extend to general systems, but only to systems that have a certain monotonicity property. We shall treat this important question in Supplement I below.

X. Existence Theorem for Complex Differential Equations. *Let the vector function* $\mathbf{f}(z, \mathbf{w})$ *of* $n+1$ *complex variables* $(z, \mathbf{w}) = (z, w_1, \ldots, w_n)$ *with values in* $\mathbb{C}^n$ *be holomorphic in a domain* $D \subset \mathbb{C}^{n+1}$ *(i.e., each component is continuously differentiable with respect to all* $n+1$ *complex variables), and let* $(z_0, \mathbf{w}_0) \in D$.

Then the initial value problem

$$\mathbf{w}' = \mathbf{f}(z, \mathbf{w}), \quad \mathbf{w}(z_0) = \mathbf{w}_0 \tag{9}$$

has exactly one holomorphic solution $\mathbf{w}(z)$*. The solution exists (at least) in the disk* $K : |z - z_0| < \alpha$*, where* $\alpha > 0$ *is determined as in* 8.II.

The solution $\mathbf{w}$ (i.e., each of the components $w_\nu(z)$) can be expanded in a power series about the point z_0 with a radius of convergence $\geq \alpha$.

XI. Autonomous Systems. In this section, we develop a general framework for problems of the type introduced in 3.V. We call the system of differential equations (1) *autonomous* if the right-hand side $\mathbf{f}(x, \mathbf{y})$ does not depend explicitly on x. Thus an autonomous equation has the form

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}). \tag{10}$$

Autonomous equations frequently arise in applications where the independent variable is time. With these problems in mind, we denote the independent variable by t and write $\mathbf{y} = \mathbf{y}(t)$. In the results that follow, $\mathbf{f}$ is assumed to be locally Lipschitz continuous in an open set $G \subset \mathbb{R}^n$. Thus Theorem VI applies to (10) and it follows that initial value problems are uniquely solvable and the solution can be extended to the boundary of $D = \mathbb{R} \times G$ (cf. the definition in 6.VII). This leads us to a number of conclusions about solutions to (10):

(a) A solution $\mathbf{y}$ exists in a maximal open interval $J = (a, b)$.

(b) If $\mathbf{y}$ is a solution of (10) in the interval $J = (\alpha, \beta)$, then $\mathbf{z}(t) := \mathbf{y}(t + c)$ is a solution in the interval $J_{-c} = (\alpha - c, \beta - c)$.

(c) *Phase Space and Trajectories (Orbits).* For autonomous systems, the space $\mathbb{R}^n$ is called the *phase space*. The curve $C = \mathbf{y}(J) := \{\mathbf{y}(t) : t \in J\} \subset G$ in the phase space generated by a solution $\mathbf{y}$ on a maximal interval of existence $J = (a, b)$ is called the *trajectory* or the *orbit* of $\mathbf{y}$; cf. A.I for the definition of curves. If $\mathbf{z}$ is another solution with $\mathbf{z}(t_0) \in C$ and $\mathbf{z}(t_0) = \mathbf{y}(t_1)$ for some $t_1 \in J$, then $\bar{\mathbf{y}}(t) = \mathbf{y}(t_1 - t_0 + t)$ is also a solution by (b), and the relation $\mathbf{z}(t_0) = \bar{\mathbf{y}}(t_0)$ implies $\mathbf{z} = \bar{\mathbf{y}}$. Therefore, the trajectories of $\mathbf{y}$ and $\mathbf{z}$ coincide.

If the "half trajectory" $\mathbf{y}([c, b))$ with $c \in J$ is contained in a compact subset of G, then $b = \infty$; a corresponding statement holds for the interval $(a, c]$.

(d) *The Phase Portrait.* Two trajectories are either disjoint or identical. Each point of G belongs to exactly one trajectory. The collection of all trajectories is called the *phase portrait* of the differential equation; cf. 3.V.

(e) The two differential equations $\mathbf{y}' = \mathbf{f}(\mathbf{y})$ and $\mathbf{y}' = \lambda \mathbf{f}(\mathbf{y})$ (with $\lambda \neq 0$) generate the same trajectories, hence the same phase portraits (with the same orientation if $\lambda > 0$).

(f) *Periodic Solutions.* If $\mathbf{y}$ is a solution and $\mathbf{y}(t_0) = \mathbf{y}(t_1)$ for some $t_0 \neq t_1$, then $\mathbf{y}$ is periodic with period $p = t_0 - t_1$. This follows from (c), with $\mathbf{z} = \mathbf{y}$. The maximal interval of existence is $\mathbb{R}$. A nonconstant, continuous, periodic function has a smallest period $T > 0$, also known as the *minimal period.*

(g) *Critical Points.* A point $\mathbf{a} \in G$ is called a *critical point* (also a *stationary point* or *equilibrium point*) of $\mathbf{f}$ if $\mathbf{f}(\mathbf{a}) = \mathbf{0}$. If $\mathbf{a}$ is a critical point of $\mathbf{f}$, then $\mathbf{y}(t) \equiv \mathbf{a}$ is a solution in $\mathbb{R}$. The corresponding orbit is the singleton $\{\mathbf{a}\}$.

(h) If the solution $\mathbf{y}$ exists for $t \geq t_0$ and if $\mathbf{a} = \lim\limits_{t \to \infty} \mathbf{y}(t)$ exists and belongs to G, then $\mathbf{a}$ is a critical point; i.e. $\mathbf{f}(\mathbf{a}) = \mathbf{0}$.

Proof of (h). Suppose ϕ is a real-valued C^1-function and $\lim\limits_{t \to \infty} \phi'(t) = \alpha \neq 0$. If $\alpha > 0$, then for large t, $\phi'(t) > \alpha/2$. It follows that $\lim\limits_{t \to \infty} \phi(t) = \infty$. Similarly, $\alpha < 0$ implies $\lim \phi(t) = -\infty$. Now the hypotheses imply that $\lim \mathbf{y}'(t) = \mathbf{f}(\mathbf{a})$. The preceding argument applied to the components $y_k(t)$ of $\mathbf{y}$ shows that $\lim \mathbf{y}'(t) = \mathbf{0}$, i.e., $\mathbf{f}(\mathbf{a}) = \mathbf{0}$. ■

Supplement I: Differential Inequalities and Invariance

Does the comparison theorem 9.III carry over to systems when the natural (componentwise) ordering of points is introduced in $\mathbb{R}^n$? Not surprisingly, the answer is negative in general. In the next section, we treat those systems for which such a theorem holds; later we show how to obtain bounds for solutions of (1) in the general case.

For $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, inequalities are defined componentwise:

$$\mathbf{y} \leq \mathbf{z} \Longleftrightarrow y_i \leq z_i \quad \text{for} \quad i = 1, \ldots, n,$$

$$\mathbf{y} < \mathbf{z} \Longleftrightarrow y_i < z_i \quad \text{for} \quad i = 1, \ldots, n.$$

XII. Monotonicity and Quasimonotonicity. The function $\mathbf{f}(x, \mathbf{z}) : D \subset \mathbb{R}^{n+1} \to \mathbb{R}^n$ is said to be *increasing* in $\mathbf{y}$ if $\mathbf{y} \le \mathbf{z}$ implies $\mathbf{f}(x, \mathbf{y}) \le \mathbf{f}(x, \mathbf{z})$, and *quasimonotone increasing* in $\mathbf{y}$ if f_i is increasing in y_j for $i \neq j$, more exactly, if for $i = 1, \ldots, n$,

$$\mathbf{y} \le \mathbf{z},\ y_i = z_i,\ (x, \mathbf{y}), (x, \mathbf{z}) \in D \quad \text{implies} \quad f_i(x, \mathbf{y}) \le f_i(x, \mathbf{z}).$$

An $n \times n$ matrix $C = (c_{ij})$ is said to be *positive* if $c_{ij} \ge 0$ for all components, and *essentially positive* if $c_{ij} \ge 0$ for $i \neq j$. The same terminology is used for matrices $C(x) = (c_{ij}(x))$. In connection with matrix products, vectors $\mathbf{y}, \mathbf{u}, \ldots$ are always assumed to be column vectors, $\mathbf{y} = (y_1, \ldots, y_n)^\top$.

It is easily seen that a linear function $\mathbf{f}(x, \mathbf{y}) = C(x)\mathbf{y}$ is quasimonotone if and only if $C(x)$ is essentially positive. If $D \subset \mathbb{R}^{n+1}$ is open and convex and $\mathbf{f}$ and $\partial \mathbf{f}/\partial \mathbf{y}$ belong to $C(D)$, then $\mathbf{f}$ is quasimonotone increasing if and only if the Jacobian $\partial \mathbf{f}/\partial \mathbf{y}$ is essentially positive. Without convexity, this is not true in general.

If the matrix $C(x) + \lambda I$ is positive for some $\lambda > 0$, then $C(x)$ is obviously essentially positive. Conversely, if $C(x)$ is essentially positive and if the diagonal elements of C are bounded below, then $C + \lambda I$ is positive for large λ. There is a similar relation between monotone and quasimonotone functions. Again it is obvious that $\mathbf{f}(x, \mathbf{z})$ is quasimonotone increasing when $\mathbf{f}(x, \mathbf{z}) + \lambda \mathbf{z}$ is increasing for some $\lambda > 0$. Conversely, if $\mathbf{f}$ is quasimonotone increasing and satisfies a Lipschitz condition in $\mathbf{y}$, then $\mathbf{f}(x, \mathbf{y}) + \lambda \mathbf{y}$ is increasing in $\mathbf{y}$ for large λ. In short, a smooth function is quasimonotone increasing if it becomes monotone increasing when a large multiple of the identity is added.

The propositions that follow can be summarized in a general

Principle. *The theorems in § 9 for a single equation carry over to systems if and only if the right-hand side* $\mathbf{f}(x, \mathbf{y})$ *is quasimonotone increasing in* $\mathbf{y}$.

We use the notation $P\mathbf{v} = \mathbf{v}'(x) - \mathbf{f}(x, \mathbf{v})$ for the defect.

Comparison Theorem. *Assume that* $\mathbf{f} : D \to \mathbb{R}^n$ *is quasimonotone increasing and that* $\mathbf{v}$, $\mathbf{w}$ *are differentiable in* $J = [\xi, \xi + a]$. *Then*

(a) $\mathbf{v}(\xi) < \mathbf{w}(\xi)$, $P\mathbf{v} < P\mathbf{w}$ *in* J *implies* $\mathbf{v} < \mathbf{w}$ *in* J.

(b) *If* $\mathbf{f}(x, \mathbf{y})$ *satisfies a local Lipschitz condition in* $\mathbf{y}$, *then* $\mathbf{v}(\xi) \le \mathbf{w}(\xi)$, $P\mathbf{v} \le P\mathbf{w}$ *in* J *implies* $\mathbf{v} \le \mathbf{w}$ *in* J; *moreover, the index set splits into two subsets* α *and* β, *such that for*

$$i \in \alpha : \quad v_i \; < \; w_i \;\; in \;\; (\xi, \xi + a] \tag{11}$$

$$j \in \beta : \quad v_j \; = \; w_j \;\; in \;\; [\xi, \xi + \delta_j) \;\; and \;\; v_j < w_j \;\; \text{in} \;\; (\delta_j, \xi + a), \tag{12}$$

where $\delta_j > 0$.

A simple consequence is

(c) **M. Hirsch's Theorem.** *Assume that* $\mathbf{f}(\mathbf{y}) \in C^1$ *has an essentially positive and irreducible Jacobian* $\partial \mathbf{f}(\mathbf{y})/\partial \mathbf{y}$. *Then two solutions* $\mathbf{v}$, $\mathbf{w}$ *of* $\mathbf{y}' = \mathbf{f}(\mathbf{y})$ *with* $\mathbf{v}(\xi) \le \mathbf{w}(\xi)$, $\mathbf{v}(\xi) \neq \mathbf{w}(\xi)$ *satisfy* $\mathbf{v}(x) < \mathbf{w}(x)$ *for* $x > \xi$.

Proof. If (a) is false, then there is a first point $x_0 > \xi$ such that $\mathbf{v}(x) \leq \mathbf{w}(x)$ for $x \leq x_0$ and $v_i(x_0) = w_i(x_0)$ for some index i. It follows easily that $v_i'(x_0) \geq w_i'(x_0)$ (apply the reasoning of 9.I to v_i and w_i). On the other hand, $P\mathbf{v} < P\mathbf{w}$ and $\mathbf{v} \leq \mathbf{w}$, $v_i = w_i$ at $x = x_0$, implies by quasimonotonicity

$$w_i' - v_i' > f_i(\mathbf{w}) - f_i(\mathbf{v}) \geq 0 \quad \text{at} \quad x = x_0,$$

which is a contradiction. Here and below, the argument x or x_0 is suppressed.

(b) Assume that $\mathbf{f}$ satisfies a Lipschitz condition $|\mathbf{f}(x, \mathbf{w}+\mathbf{z}) - \mathbf{f}(x, \mathbf{w})| \leq L|\mathbf{z}|$ (maximum norm) and let $\rho = \mathrm{e}^{2Lx}$, $\mathbf{h} = \varepsilon(\rho, \ldots, \rho)$, $\varepsilon > 0$. Then

$$\mathbf{h}' = 2L\mathbf{h} > \mathbf{f}(\mathbf{w} + \mathbf{h}) - \mathbf{f}(\mathbf{w}) \Rightarrow P(\mathbf{w} + \mathbf{h}) > P\mathbf{w}.$$

Hence, by (a), $\mathbf{v} < \mathbf{w} + \mathbf{h}$, which gives $\mathbf{v} \leq \mathbf{w}$, since $\varepsilon > 0$ is arbitrary.

Last part of (b): It follows from the local Lipschitz condition that for some $L > 0$, $\mathbf{f}(x, \mathbf{z}) + L\mathbf{z}$ is (componentwise) increasing in $\mathbf{z}$ as long as $\mathbf{v}(t) \leq \mathbf{z} \leq \mathbf{w}(t)$. From (b) we get $\mathbf{u} = \mathbf{w} - \mathbf{v} \geq 0$, and furthermore,

$$\begin{aligned} \mathrm{e}^{-Lx}(\mathrm{e}^{Lx}\mathbf{u})' &= \mathbf{u}' + L\mathbf{u} = \mathbf{f}(\mathbf{w}) + P\mathbf{w} - \mathbf{f}(\mathbf{v}) - P\mathbf{v} + L(\mathbf{w} - \mathbf{v}) \\ &\geq \mathbf{f}(\mathbf{w}) + L\mathbf{w} - (\mathbf{f}(\mathbf{v}) + L\mathbf{v}) \geq 0; \end{aligned}$$

i.e., each component of $\mathrm{e}^{Lx}\mathbf{u}$ is increasing. This proves the last part. ∎

The *proof* of (c) is an exercise. *Hints:* $\mathbf{f}$ is quasimonotone, hence $\mathbf{v} \leq \mathbf{w}$. Assume β is nonempty, and write $\mathbf{v} = (v_\alpha, v_\beta), \ldots$. Then $v_\alpha < w_\alpha$, $v_\beta = w_\beta$ implies $f_\beta(v_\alpha, v_\beta) \leq f_\beta(w_\alpha, v_\beta)$, and equality is excluded if $\partial f_\beta(y_\alpha, y_\beta)/\partial y_\alpha \geq 0$ is nonzero which is a contradiction.

Sub- and Supersolutions. Maximal and Minimal Solutions. On the basis of the preceding theorem, sub- and supersolutions for the initial value problem (6) can be defined by $\mathbf{v}' < \mathbf{f}(x, \mathbf{v})$, $\mathbf{v}(\xi) < \boldsymbol{\eta}$ and $\mathbf{w}' > \mathbf{f}(x, \mathbf{w})$, $\mathbf{w}(\xi) > \boldsymbol{\eta}$ (or the same with $\leq$ in case (b)) exactly as in 9.IV.

Furthermore, in the case where $\mathbf{f}$ is continuous and quasimonotone increasing, one can construct a maximal solution $\mathbf{y}^*$ and a minimal solution $\mathbf{y}_*$ of (6). For those solutions the propositions

(d) $\mathbf{v}(\xi) \leq \boldsymbol{\eta}$, $\mathbf{v}' \leq \mathbf{f}(x, \mathbf{v})$ in $J \Longrightarrow \mathbf{v} \leq \mathbf{y}^*$ in J;

(e) $\mathbf{w}(\xi) \geq \boldsymbol{\eta}$, $\mathbf{w}' \geq \mathbf{f}(x, \mathbf{w})$ in $J \Longrightarrow \mathbf{w} \geq \mathbf{y}_*$ in J

hold. The proofs of 9.VI and 9.VIII carry over.

Our program announced in the general "principle" has been carried out now, except for the "only if" part. The fact that quasimonotonicity is a necessary prerequisite for the validity of the comparison theorem is proved in Redheffer and Walter (1975). The theory that has been developed above goes back to M. Müller (1926, 1927) and E. Kamke (1932). In the next section we prove a theorem on differential inequalities for general systems, that also is due to M. Müller (1927). It is of a different nature inasmuch as it requires an upper bound and a lower bound simultaneously.

XIII. M. Müller's Theorem for Arbitrary Systems. *Let* $\mathbf{f}(x, \mathbf{y}) : D \to \mathbb{R}^n$ *satisfy a local Lipschitz condition in* $\mathbf{y}$. *Let* $\mathbf{y}, \mathbf{v}, \mathbf{w} : J = [\xi, \xi + a] \to \mathbb{R}^n$ *be differentiable,* $\mathbf{v} \le \mathbf{w}$ *in* J, $\mathbf{v}(\xi) \le \mathbf{y}(\xi) \le \mathbf{w}(\xi)$, $\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$, *and*

$$v_i' \le f_i(x, \mathbf{z}) \text{ for all } \mathbf{z} \in \mathbb{R}^n \text{ such that } \mathbf{v}(x) \le \mathbf{z} \le \mathbf{w}(x),\ z_i = v_i(x),$$

$$w_i' \le f_i(x, \mathbf{z}) \text{ for all } \mathbf{z} \in \mathbb{R}^n \text{ such that } \mathbf{v}(x) \le \mathbf{z} \le \mathbf{w}(x),\ z_i = w_i(x),$$

for $i = 1, \ldots, n$. *Then*

$$\mathbf{v} \le \mathbf{y} \le \mathbf{w} \quad \text{in} \quad J.$$

Remark. The differential inequalities, which look quite complicated at first sight, have a simple geometric meaning. We use the following notation for intervals in $\mathbb{R}^n$:

$$I = [\mathbf{a}, \mathbf{b}] = \{\mathbf{z} \in \mathbb{R}^n : \mathbf{a} \le \mathbf{z} \le \mathbf{b}\}, \quad \text{where } \mathbf{a}, \mathbf{b} \in \mathbb{R}^n \text{ and } \mathbf{a} \le \mathbf{b}.$$

The boundary of this interval consists of $2n$ "faces" I_i and I^i defined by $\mathbf{z} \in I$, $z_i = a_i$ or $z_i = b_i$ $(i = 1, \ldots, n)$. In the two-dimensional case I is a rectangle with sides parallel to the axes, $\mathbf{a}$ is the left lower and $\mathbf{b}$ the right upper corner point, I_1 and I^1 are the left and right vertical sides, while I_2 and I^2 are the lower and upper horizontal sides.

Using this notation, the differential inequality for v_i can be written in the form $v_i' \le f_i(x, \mathbf{z})$ for $\mathbf{z} \in I_i(x)$, where $I(x) = [\mathbf{v}(x), \mathbf{w}(x)]$. One may carry the abbreviation a step further by (i) defining inequalities between a real number s and a set $A \subset \mathbb{R}$,

$$s \le A \iff s \le a \quad \text{for all} \quad a \in A;$$

and (ii) using a familiar notation $f_i(x, B) = \{f_i(x, \mathbf{z}) : \mathbf{z} \in B\}$. Then the differential inequalities in Müller's theorem XIII can be written as

$$\boxed{v_i' \le f_i(x, I(x)_i) \text{ and } w_i' \ge f_i(x, I(x)^i),\ I(x) = [\mathbf{v}(x), \mathbf{w}(x)]\ .}$$

Proof. We again use a two-step method. Assume first that $\mathbf{v}$, $\mathbf{w}$ satisfy strict inequalities. If the conclusion is false, there is a maximal $x_0 > a$ such that $\mathbf{v} < \mathbf{y} < \mathbf{w}$ in $[a, x_0)$ and, e.g., $v_i(x_0) = y_i(x_0)$. Hence $v_i' \ge y_i'$ at x_0. On the other hand, $\mathbf{z} = \mathbf{y}(x_0)$ satisfies $\mathbf{v}(x_0) \le \mathbf{z} \le \mathbf{w}(x_0)$, $v_i(x_0) = z_i$, and hence $v_i'(x_0) < f_i(x_0, \mathbf{y}(x_0)) = y_i'(x_0)$, which is a contradiction. In the second step one may assume that $\mathbf{f}$ satisfies (for the arguments involved) a Lipschitz condition $|\mathbf{f}(x, \mathbf{y}) - \mathbf{f}(x, \mathbf{z})| \le L|\mathbf{y} - \mathbf{z}|$, where $|\cdot|$ is the maximum norm. Then the case of weak inequalities is reduced to the case treated above by considering $\mathbf{w}_\varepsilon = \mathbf{w} + \varepsilon\mathbf{h}$, $\mathbf{v}_\varepsilon = \mathbf{v} - \varepsilon\mathbf{h}$, where $\mathbf{h} = (\rho, \ldots, \rho)$, $\rho = e^{2Lx}$. ∎

Invariant Intervals. Let us consider the important special case where the functions $\mathbf{v}$, $\mathbf{w}$ are constant. Then the assumptions in Müller's theorem are

$$\boxed{\mathbf{y}(\xi) \in I,\ f_i(x, I_i) \geq 0 \ \text{ and } \ f_i(x, I^i) \leq 0, \ \text{ where } \ I = [\mathbf{v}, \mathbf{w}]}$$

$(i = 1, \ldots, n)$, and the conclusion is $\mathbf{y}(x) \in I$ for $x \in J$.

These inequalities signify that on the boundary of I the vector field $\mathbf{f}$ points into I. For example, in the case $n = 2$ we have $f_1 \geq 0$ on the left side and $f_1 \leq 0$ on the right side, and similarly $f_2 \geq 0$ on the lower side and ≤ 0 on the upper side. When this is true, then $\mathbf{y}(\xi) \in I$ implies $\mathbf{y}(x) \in I$ for $x > \xi$. This fact is expressed by saying that I is an *invariant interval* (*invariant rectangle* in the case $n = 2$). Thus a theorem on invariant intervals is a very special case of Müller's theorem.

XIV. The Case $n = 2$. Consider problem (1) for $\mathbf{y} = (y_1, y_2)$, i.e.,

$$y_1' = f_1(x, y_1, y_2),\ y_2' = f_2(x, y_1, y_2),\ y_1(\xi) = \eta_1,\ y_2(\xi) = \eta_2, \tag{13}$$

and assume for simplicity that $\mathbf{f}(x, \mathbf{y}) = (f_1, f_2)$ is locally Lipschitz continuous in y. We discuss several monotonicity assumptions with respect to the theorems in the two preceding sections. The behavior of f_1 in the variable y_2 and of f_2 in the variable y_1 is crucial. Let $\mathbf{y}$ be the (unique) solution of (13).

(a) **Case (I, I).** If $f_1(t, y_1, y_2)$ is increasing in y_2 and $f_2(t, y_1, y_2)$ is increasing in y_1, then f is quasimonotone increasing. In this case,

$$v_1' \leq f_1(x, v_1, v_2),\ v_2' \leq f_2(x, v_1, v_2) \quad \text{and} \quad v_1(\xi) \leq \eta_1,\ v_2(\xi) \leq \eta_2$$

implies $v_1 \leq y_1$ and $v_2 \leq y_2$ in J.

(b) **Case (D, D).** Let f_1 be decreasing in y_2 and f_2 be decreasing in y_1. Then $\mathbf{f}$ is not quasimonotone, but the equivalent system for $\bar{y}_1 = -y_1$ and y_2,

$$\bar{y}_1' = -f_1(x, -\bar{y}_1, y_2), \quad y_2' = f_2(x, -\bar{y}_1, y_2),$$

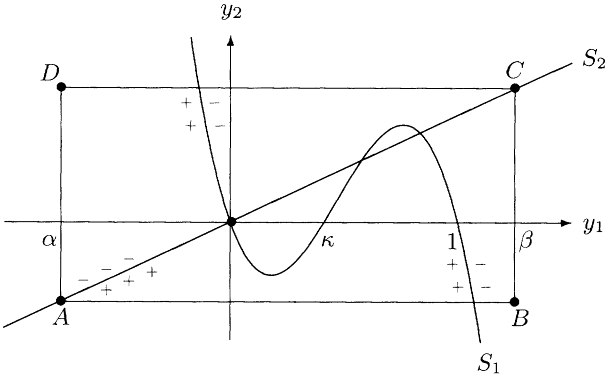
is quasimonotone. Therefore, the statement given in (a) holds for this system. In terms of the original system it reads as follows:

$$v_1' \geq f_1(x, v_1, v_2), \quad v_2' \leq f_2(x, v_1, v_2), \quad v_1(\xi) \geq \eta_1, \quad v_2(\xi) \leq \eta_2$$

implies $v_1 \geq y_1$ and $v_2 \leq y_2$ in J. In such a case $\mathbf{v} = (v_1, v_2)$ is sometimes called an upper–lower solution.

(c) **Case (D, I).** If f_1 is decreasing in y_2 and f_2 is increasing in y_1, it is impossible to enforce quasimonotonicity by a simple transformation. We employ Müller's theorem XIII. The conditions for $\mathbf{v}$, $\mathbf{w}$ are $\mathbf{v}(\xi) \leq \boldsymbol{\eta} \leq \mathbf{w}(\xi)$, $\mathbf{v} \leq \mathbf{w}$ and

$$v_1' \leq f_1(x, v_1, w_2), \quad w_1' \geq f_1(x, w_1, v_2),$$



$S_1 : y_2 = f(y_1) = -y_1(y_1 - \kappa)(y_1 - 1)$
$S_2 : y_2 = by_1,\ b = \sigma/\gamma > 0$

FitzHugh–Nagumo equations

$$v_2' \le f_2(x, v_1, v_2), \quad w_2' \ge f_2(x, w_1, w_2),$$

and the conclusion is $\mathbf{v} \le \mathbf{y} \le \mathbf{w}$.

Example. The FitzHugh–Nagumo Equations. These equations represent a simple model for the mechanism underlying signal transmission in neurons. Usually, these equations are given in the form

$$\boxed{u' = \sigma v - \gamma u, \ v' = f(v) - u}$$

where σ and γ are positive constants and $f(v) = -v(v-\kappa)(v-1)$ with $0 < \kappa < 1$ (typically). Here $u(t)$ represents the density of a chemical substance and $v(t)$ an electric potential depending on time t. For further information, see Jones and Sleeman (1983). With the notation $v = y_1$, $u = y_2$ we have

$$\text{(d)} \quad y_1' = f_1(y_1, y_2) = f(y_1) - y_2, \quad y_2' = f_2(y_1, y_2) = \sigma y_1 - \gamma y_2.$$

Since f_1 is decreasing in y_2 and f_2 increasing in y_1, the conditions from (c) for a subfunction $\mathbf{v} = (v_1, v_2)$ and a superfunction $\mathbf{w} = (w_1, w_2)$ are

$$v_1' \le f(v_1) - w_2, \quad w_1' \ge f(w_1) - v_2,$$
$$v_2' \le \sigma v_1 - \gamma v_2, \quad w_2' \ge \sigma w_1 - \gamma w_2.$$

In the figure the curves $S_1 : f_1 = 0$ and $S_2 : f_2 = 0$ are drawn, and the side of the curve where the component is positive or negative is indicated. A rectangle $[A, C]$ $(A < C)$ with corners A, B, C, D is invariant if $\mathbf{v}(t) \equiv A$ and $\mathbf{w}(t) \equiv C$ satisfy the preceding inequalities; i.e. if

$$A \text{ is below } S_2, \quad C \text{ is above } S_2,$$
$$D \text{ is below } S_1, \quad B \text{ is above } S_1;$$

for example, on the side AB we must have $f_2 \geq 0$, which is guaranteed if A is below S_2.

The straight line S_2 is given by $y_2 = by_1$, $b = \sigma/\gamma$. If we choose A and C on S_2, i.e., $A = (\alpha, b\alpha)$, $C = (\beta, b\beta)$ with $\alpha < 0 < \beta$, then $B = (\beta, b\alpha)$, $D = (\alpha, b\beta)$, and the two remaining conditions regarding S_1 read

$$f(\alpha) \geq b\beta, \quad f(\beta) \leq b\alpha.$$

If $-\alpha = \beta$ is a large positive number, then these inequalities are satisfied:

There are arbitrarily large invariant rectangles, which implies that all solutions of the FitzHugh–Nagumo equations exist for all $t \geq 0$.

(e) *Exercise.* Find sub- and superfunctions depending on t (shrinking rectangles). Assume that A and C, that is, $\mathbf{v}(t)$ and $\mathbf{w}(t)$, are on the line

$$y_2 = ay_1, \quad \text{where} \quad a = b + \varepsilon/\gamma \quad (\varepsilon > 0)$$

(this line is slightly steeper than S_2). Use the ansatz $\mathbf{v}(t) = (1, a)\alpha e^{-\delta(t-c)}$, $\mathbf{w}(t) = (1, a)\beta e^{-\delta(t-c)}$ $(\alpha < 0 < \beta,\ \delta > 0)$ and give conditions on f such that $\mathbf{v}$, $\mathbf{w}$ satisfy the inequalities in (d) for $t \leq c$.

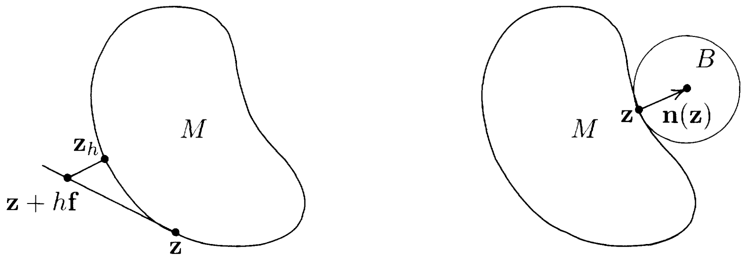
(f) Show that there exist arbitrarily large invariant rectangles if the function f in the FitzHugh–Nagumo system satisfies $f(s)/s \to -\infty$ as $s \to \pm\infty$.

XV. Invariant Sets in $\mathbb{R}^n$. Tangent Condition. A set $M \subset \mathbb{R}^n$ is said to be (*positively*) *invariant* or *flow-invariant* with respect to the system $\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$ if for any solution $\mathbf{y}$, $\mathbf{y}(a) \in M$ implies $\mathbf{y}(x) \in M$ for $x > a$ (as long as the solution exists). The basic hypothesis for invariance is a tangent condition that roughly states that at a boundary point $\mathbf{z} \in \partial M$ the vector $\mathbf{f}(x, \mathbf{z})$ is either tangent to M or points into the interior of M. The problem is to find a formulation that applies to arbitrary sets M and does not use any smoothness of the boundary of M. The tangent condition is given in two forms. It is assumed that $\mathbf{f}$ is defined in $G \supset J \times \overline{M}$, $J = [a, b]$:

$$\lim_{h \to 0+} \frac{1}{h} \text{dist}\,(\mathbf{z} + h\mathbf{f}(x, \mathbf{z}), M) = 0 \quad \text{for} \quad \mathbf{z} \in \overline{M}, \quad x \in J, \tag{T$_d$}$$

$$\begin{array}{l}\langle \mathbf{n}(\mathbf{z}), \mathbf{f}(x, \mathbf{z}) \rangle \leq 0 \quad \text{for} \quad x \in J, \quad \mathbf{z} \in \partial M, \text{ where} \\ \qquad \mathbf{n}(\mathbf{z}) \text{ is the outer normal to } M \text{ at } \mathbf{z}.\end{array} \tag{T$_i$}$$

Here, dist $(\mathbf{z}, M)$ denotes the distance from $\mathbf{z}$ to M, and $\langle \mathbf{y}, \mathbf{z} \rangle$ is the familiar inner product $y_1 z_1 + \cdots + y_n z_n$. The vector $\mathbf{n}(\mathbf{z}) \neq 0$ is said to be an outer normal to M at $\mathbf{z} \in \partial M$ if the *open* ball B with center at $\mathbf{z} + \mathbf{n}(\mathbf{z})$ and radius $|\mathbf{n}(\mathbf{z})|$ has no point in common with M. Geometrically speaking, the ball B touches M at $\mathbf{z}$, $\mathbf{z} \in \partial M \cap \overline{B}$. The index d or i in the tangent condition indicates that the definition employs the distance or the inner product in $\mathbb{R}^n$, respectively. Naturally, (T$_d$) is true for $\mathbf{z} \in$ int $\overline{M}$. First we give without proof some relations between these two conditions.



Condition (T_d) Condition (T_i)

Tangent conditions

(a) *(T_d) implies (T_i).*

(b) *If $\mathbf{f}$ is continuous in $J \times \overline{M}$, then (T_i) implies (T_d), and (T_d) even holds uniformly in compact subsets of $J \times \overline{M}$.*

XVI. Invariance Theorem (with Uniqueness Condition). *The closed set $M \subset \mathbb{R}^n$ is flow-invariant with respect to the differential equation $\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$ if $\mathbf{f}$ satisfies the tangent condition* (T_i) *or* (T_d) *and a condition of Lipschitz type*

$$\langle \mathbf{y}_1 - \mathbf{y}_2, \mathbf{f}(x, \mathbf{y}_1) - \mathbf{f}(x, \mathbf{y}_2) \rangle \le L|\mathbf{y}_1 - \mathbf{y}|_e^2. \tag{12}$$

Proof. Let $\mathbf{y}$ be a solution of $\mathbf{y}' = f(x, \mathbf{y})$, let $\mathbf{y}(a) \in M$ and $\rho(x) = \operatorname{dist}(\mathbf{y}(x), M)$. Assume that $\mathbf{y}(s) \notin M$, i.e., $\rho(s) > 0$ for some $s > a$, and let $\mathbf{z} \in \partial M$ be such that $\rho(s) = |\mathbf{y}(s) - \mathbf{z}|$. The function $\sigma(x) = |\mathbf{y}(x) - \mathbf{z}|$ satisfies $\rho(x) \le \sigma(x)$, $\rho(s) = \sigma(s)$, and hence $D^+\rho(s) \le \sigma'(s)$ ($D^+\rho$ is a Dini derivative). Note that σ is differentiable near s and

$$\frac{1}{2}(\sigma^2)' = \sigma\sigma' = \langle \mathbf{y}(x) - \mathbf{z}, \mathbf{y}'(x) \rangle = \langle \mathbf{y}(x) - \mathbf{z}, \mathbf{f}(x, \mathbf{y}(x)) \rangle.$$

Now, $\mathbf{n}(\mathbf{z}) = \mathbf{y}(s) - \mathbf{z}$ is an outer normal to M at $\mathbf{z}$, and (T_i) implies $\langle \mathbf{n}(\mathbf{z}), \mathbf{f}(s, \mathbf{z}) \rangle \le 0$. Using this inequality, we obtain for $t = s$,

$$\sigma\sigma' = \langle \mathbf{n}(\mathbf{z}), \mathbf{f}(s, \mathbf{y}(s)) \rangle \le \langle \mathbf{n}(\mathbf{z}), \mathbf{f}(s, \mathbf{y}) - \mathbf{f}(s, \mathbf{z}) \rangle \le L\rho^2.$$

Hence $D^+\rho(s) \le L\rho(s)$. Since s was arbitrary, we have $D^+\rho \le L\rho$ whenever $\rho > 0$. There exists an interval $[b, c]$ such that $\rho(b) = 0$ and $\rho > 0$ in $(b, c]$. It follows from Theorem 9.VIII that $\rho = 0$ in $[b, c]$, which is a contradiction. ∎

In conclusion we state without proof another

Invariance Theorem (Existence). *Let $M \subset \mathbb{R}^n$ be closed, $\mathbf{f}(x, \mathbf{y}) : [\xi, \xi + a] \times M \to \mathbb{R}^n$ bounded and continuous, and assume that a tangent condition* (T_i) *or* (T_d) *holds. Then, for any $\boldsymbol{\eta} \in M$, the initial value problem* (6) *has a solution* $\mathbf{y}$ *such that $\mathbf{y}(x) \in M$ for $\xi \le x \le \xi + a$.*

In particular, if M is compact and if $\mathbf{f}$ *is continuous in $[\xi, \infty) \times M$, then there exists a global solution satisfying $\mathbf{y}(x) \in M$ for all $x \ge \xi$.*

Remarks. 1. Nagumo (1942) formulated condition (T_d) and proved the last theorem. The invariance problem was revitalized in the late sixties. Condition (T_i) goes back to Bony (1969), who proved Theorem XVI under the assumption that $\mathbf{f} = \mathbf{f}(\mathbf{y})$ is locally Lipschitz continuous; Brezis (1970) showed the same under condition (T_d). The propositions XV.(a) and (b) are due to Redheffer (1972) and Crandall (1972), respectively. The paper by Hartman (1972) and other papers cited here give the impression that the authors were not aware of Nagumo's theorem. Obviously, invariance follows from this theorem if $\mathbf{f}$ is continuous and if the assumptions regarding $\mathbf{f}$ guarantee uniqueness.

2. The invariance theorem XVI and its proof carry over to differential equations in a Hilbert space H (where $\mathbf{f}$ and the solution have values in H and the inner product in H appears in (12)).

XVII. An Example from Ecology: Competing Species. We consider nonnegative solutions $(u(t), v(t))$ of the autonomous system

$$u' = u(3 - u - 2v), \quad v' = v(4 - 3u - v). \tag{13}$$

In the biological model, $u(t)$ and $v(t)$ are the numbers of individuals in two competing populations that feed on the same limited food source, and t is time. If v is absent, $u(t)$ is governed by the logistic equation $u' = u(3-u)$ with growth rate $3 - u$, which diminishes to $3 - u - 2v$ in the presence of v. The same applies to v.

We discuss the global behavior of solutions as $t \to \infty$ using the phase plane. Since both u and v are nonnegative, only the first quadrant $Q = [0, \infty)^2$ in the uv-plane is considered.

Writing the system in the form $(u', v') = F(u, v) = (f(u, v), g(u, v))$, we find that $f = 0$ on the line $\overline{BD}$ and on the v-axis, while $g = 0$ on the line $\overline{AC}$ and the u-axis; in the figure, the sign of f or g on the two sides of the zero line is indicated. This figure shows that there are

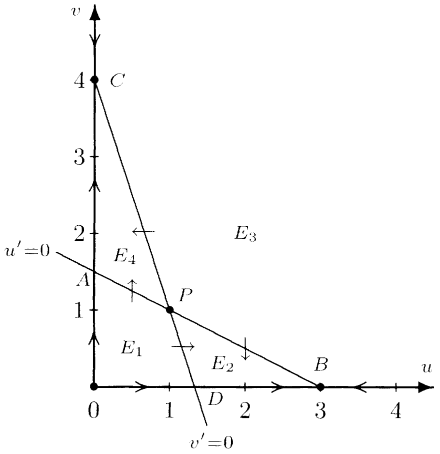
four stationary points, $\quad (0, 0),\ B = (3, 0),\ P = (1, 1),\ C = (0, 4)$

and

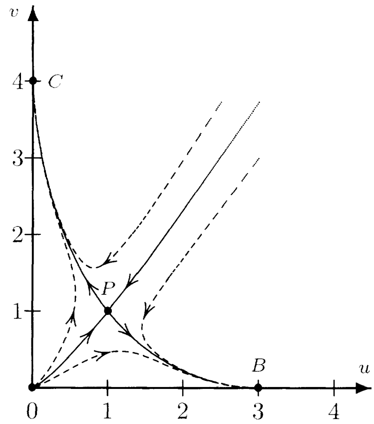
four regions, $E_1 - E_4$, where the signs of u' and v' do not change;

the arrows show the direction of F on the boundary of these regions. First observation: the regions E_2 and E_4 are positively invariant, and so is Q.

Consider a solution (u, v) starting (say, at $t = 0$) in region E_3. Both components u and v decrease as t increases. The solution either stays in E_3 for all



Nullclines: $u' = f = 0$, $v' = g = 0$



Phase Portrait

$t > 0$ or enters one of E_2 (through $\overline{BP}$) or E_4 (through $\overline{PC}$) and remains there. Similarly, a solution starting in E_1 stays there or enters E_2 or E_4, remaining there. Because $u(t)$ and $v(t)$ are eventually monotone in all cases, the limit

$$\lim_{t\to\infty}(u(t), v(t)) = (u_\infty, v_\infty) = P_\infty \in Q$$

exists. According to XI.(g), $F(P_\infty) = 0$; i.e., every solution $\neq 0$ converges to B, C, or P as $t \to \infty$. In the first case, the v-population dies out; in the second case it is the u-population which becomes extinct. In the third case the solution converges to a state of coexistence P. In each of E_1 and E_3 there is a unique solution (modulo time shift, see XI.(b), (c)) converging to P. The corresponding orbits combine to form a curve from 0 through P to infinity, which divides the first quadrant in two regions. A solution starting (at any time) in the upper region converges to C, and one starting in the lower region converges to B. Hence this curve, also called a *separatrix*, categorizes the asymptotic behavior of all solutions as $t \to \infty$. A proof (not simple) is indicated in (d) below.

Exercise. (a) Show that the regions E_1 and E_3 and Q are negatively invariant and that every solution starting in E_1 tends to 0 as $t \to -\infty$.

(b) The diagonal $u = v$ cuts the first quadrant into a lower part Q_l and an upper part Q_u. Show that the regions $(E_1 \cup E_4) \cap Q_u$ and $(E_2 \cup E_3) \cap Q_l$ are positively invariant and that a solution starting in one of these regions does not converge to P as $t \to \infty$.

(c) Let F be the set of points between the straight lines $v = u$ and $v = 2u-2$. Show that for $i = 1$ and $i = 3$, the sets $E_i' = E_i \cap F$ are negatively invariant and that a solution starting in $E_i \setminus E_i'$ does not converge to P.

(d) Show that there exists a unique solution $(u^*(t), v^*(t))$ (modulo time shift) starting in E_3' and converging to P (same for E_1').

Hint: Make P the origin of coordinates $(\bar{u}, \bar{v})$, i.e., $u = 1+\bar{u}$, $v = 1+\bar{v}$. Then (13) becomes $(\bar{u}', \bar{v}') = -((1+\bar{u})(\bar{u}+2\bar{v}), (1+\bar{v})(3\bar{u}+\bar{v})) = (\bar{f}, \bar{g})$. Consider the differential equation for the trajectories in explicit form $\bar{v} = \bar{v}(\bar{u})$ (see equation (2c) in § 3), which reads

$$\frac{d\bar{v}}{d\bar{u}} = \frac{\bar{g}}{\bar{f}} = \frac{(1+\bar{v})(3\bar{u}+\bar{v})}{(1+\bar{u})(\bar{u}+2\bar{v})} =: \bar{h}(\bar{u}, \bar{v}).$$

The solutions $\bar{v}_n$ with initial value $\bar{v}_n(1/n) = 2/n$ converge monotonically to a solution $\bar{v}^*(\bar{u})$ that describes the trajectory (lying in E_3') of a solution (u^*, v^*) of (13) converging to P. There is only one such solution in E_3', since $\bar{h}(\bar{u}, \bar{v})$ is decreasing in $\bar{v}$ near 0 $(= P)$, which implies that the difference of two solutions is increasing as $\bar{u} \to 0+$.

Supplement II: Differential Equations in the Sense of Carathéodory

Solutions in the Sense of Carathéodory. Some facts from the theory of Lebesgue measure and Lebesgue integral for real functions are required in the following. We denote by $L(J)$ the class of functions that are measurable and integrable over J and by $AC(J)$ the class of absolutely continuous functions on J. In this section, we consider a generalization of the notion of a solution to a differential equation, introduced by Carathéodory (1918). A function $\mathbf{y}(x)$ is a solution of the differential equation (1) in the sense of Carathéodory (abbreviated, a C-solution) if $\mathbf{y}$ is absolutely continuous in the interval J and satisfies the differential equation (1) almost everywhere in J. If $\mathbf{y}$ also satisfies the initial conditions, then it is called a C-solution of the initial value problem (6). The right-hand side $\mathbf{f}(x, \mathbf{y}) : D \to \mathbb{R}^n$ is required to satisfy in D a

Carathéodory Condition: $\mathbf{f}(x, \mathbf{y})$ *is continuous as a function of* $\mathbf{y}$ *for fixed* x *and measurable as a function of* x *for fixed* $\mathbf{y}$.

XVIII. Existence and Uniqueness Theorem. *Let* $J = [\xi, \xi + a]$ *and* $S = J \times \mathbb{R}^n$ *and assume that the function* $\mathbf{f} : S \to \mathbb{R}^n$ *satisfies a Carathéodory condition in* S.

(a) *If there exists* $h \in L(J)$ *with* $|\mathbf{f}(x, \mathbf{y})| \le h(x)$ *in* S, *then the initial value problem* (6) *has at least one* C*-solution in* J.

(b) *If* $\mathbf{f}$ *satisfies* $\mathbf{f}(x, \mathbf{y}) \in L(J)$ *for fixed* $\mathbf{y}$ *and a generalized Lipschitz condition*

$$|\mathbf{f}(x, \mathbf{y}) - \mathbf{f}(x, \bar{\mathbf{y}})| \le l(x)|\mathbf{y} - \bar{\mathbf{y}}| \quad \text{in} \quad S, \quad \text{where} \quad l(x) \in L(J), \tag{14}$$

then there exists a unique C*-solution of* (6) *in* J.

Sketch of the Proof. Once again the proofs from §6 and §7 carry over. First one needs the following lemma.

XIX. Lemma. *If* $\mathbf{f}$ *satisfies the hypotheses of* XVIII *and if* $\mathbf{u}(x) \in C(J)$, *then* $\mathbf{f}(x, \mathbf{u}(x)) \in L(J)$.

It is sufficient to show that $\mathbf{f}(x, \mathbf{u}(x))$ is measurable. This is clearly the case if $\mathbf{u}$ is constant, and hence also if $\mathbf{u}$ is a step function (constant on intervals). If $\mathbf{u}$ is continuous and $(\mathbf{u}_k)$ is a sequence of step functions such that $\mathbf{u}_k \to \mathbf{u}$ pointwise, then $\mathbf{f}(x, \mathbf{u}_k) \to \mathbf{f}(x, \mathbf{u})$. This implies the measurability of $\mathbf{f}(x, \mathbf{u}(x))$. ∎

The fundamental theorem of calculus for the Lebesgue integral states that

$$\psi \in L(J),\ \phi(x) = \phi(\xi) + \int_\xi^x \psi(t)\,dt \Longleftrightarrow \phi \in AC(J), \quad \phi' = \psi \quad \text{a.e.}$$

We can now proceed along the lines of the earlier proofs, since solving the initial value problem is equivalent to finding a continuous solution of the integral equation (7). In the case (b) one can show that the operator T is a contraction in $C(J)$ with Lipschitz constant $\frac{1}{2}$. In the proof we use the norm

$$\|\mathbf{z}\| = \max_J |\mathbf{z}(x)| e^{-2L(x)} \quad \text{with} \quad L(x) = \int_\xi^x l(t)\,dt.$$

In the case (a), one can use the approximating functions $\mathbf{z}_\alpha$ introduced in §7. Equicontinuity of the family $(\mathbf{z}_\alpha)$ follows from the inequality

$$|\mathbf{z}_\alpha(x_1) - \mathbf{z}_\alpha(x_0)| \le \int_{x_0}^{x_1} h(t)\,dt = H(x_1) - H(x_0) \quad (x_0 < x_1),$$

where $H(x) = \displaystyle\int_\xi^x h(t)\,dt$ is (uniformly) continuous in J.

Thus we easily obtain the following theorem.

XX. Existence and Uniqueness Theorem. *Let* $D \subset \mathbb{R}^{n+1}$ *be open and suppose that the assumptions of* XVIII *hold on every set of the form* $S = J \times K$ *contained in* D, *where* $J \subset \mathbb{R}$ *is a compact interval and* $K \subset \mathbb{R}^n$ *is a closed ball. Then the initial value problem* (5) *has a solution for any* $(\xi, \boldsymbol{\eta}) \in D$. *Every solution can be extended to the left and right to the boundary of* D. *In the case* (b), *the solution is unique.*

The next two theorems have important applications. Note that all functions are real-valued.

XXI. Theorem on Differential Inequalities. *Let the function* $f(x, y) : D \subset \mathbb{R}^2 \to \mathbb{R}$ *satisfy a local generalized Lipschitz condition in* y, *i.e., on every compact rectangle* $R = J_x \times J_y \subset D$ (J_x, J_y *compact intervals*), *let there exist a function* $l \in L(J_x)$ *such that*

$$|f(x, y) - f(x, \bar{y})| \le l(x)|y - \bar{y}| \quad \text{for} \quad x \in J_x, \quad y, \bar{y} \in J_y. \tag{15}$$

Suppose that the functions ϕ, $\psi \in AC(J)$, $J = [\xi, \xi + a]$, *satisfy the conditions*

(a) $\phi(\xi) \leq \psi(\xi)$,

(b) $P\phi \leq P\psi$ *a.e. in* J *with* $P\phi = \phi' - f(x, \phi)$.

Then either $\phi < \psi$ *in* J, *or there exists* $c \in [\xi, \xi + a]$ *such that* $\phi = \psi$ *in* $[\xi, c]$ *and* $\phi < \psi$ *in* $(c, \xi + a]$.

A corresponding statement holds for the interval $J_- = [\xi - a, \xi]$ *with the assumption* (b′) $P\phi \geq P\psi$ *instead of* (b).

This extends Theorem 9.IX to differential equations in the sense of Carathéodory. The proof from 9.IX carries over. One simply replaces Lx with the function $L(x) = \int l(x)\,dx$. Theorem XXI provides a basis for introducing *upper* and *lower solutions* exactly as in 9.IX.

For general f, an analogue of Theorem 9.III with strict inequality does not hold for C-solutions. It fails even for classical solutions if the strict inequality in (b) is violated at just one point; take, e.g., $y' = \sqrt{|y|}$, $\phi = x^3$, $\psi = 0$ in $|x| < 1/10$. Nevertheless, we have

XXII. Maximal and Minimal Solutions. Theorem. *Let* $J = [\xi, \xi + a]$ *and* $S = J \times \mathbb{R}$. *Assume that* $f : S \to \mathbb{R}$ *satisfies Carathéodory's condition and* $|f(x, y)| \leq h(x) \in L(J)$ *in* S. *Then the initial value problem* $y' = f(x, y)$, $y(\xi) = \eta$ *has a maximal solution* y^* *and a minimal solution* y_* *existing in* J, *and for* $\phi, \psi \in AC(J)$,

$$\phi' \leq f(x, \phi) \ \text{ a.e. in } J, \quad \phi(\xi) \leq \eta \quad \text{implies} \quad \phi \leq y^* \ \text{ in } J,$$

$$\phi' \geq f(x, \psi) \ \text{ a.e. in } J, \quad \psi(\xi) \geq \eta \quad \text{implies} \quad \psi \geq y^* \ \text{ in } J.$$

In particular, $y_* \leq y \leq y^*$ *in* J *for every solution* y.

Proof. Obviously, $|y(x) - \eta| \leq \int_\xi^x h(t)\,dt =: H(x)$ for every solution $y(x)$. Let R be the rectangle $R = J \times [\eta - C, \eta + C]$, where $C = 3H(\xi + a) + 2$, and consider for $0 < \alpha < 1$ the "modulus of continuity"

$$\delta_\alpha(x) = \sup\{|f(x, y)| - f(x, z)| : (x, y), (x, z) \in R,\ |y - z| \leq \alpha\}.$$

It satisfies $0 \leq \delta_\alpha(x) \leq 2h(x)$ and and $\delta_\alpha(x) \to 0$ as $\alpha \to 0+$ (pointwise), because f is uniformly continuous in y in the compact interval $[\eta - C, \eta + C]$. Let w_α be a solution of

$$w'_\alpha = f(x, w_\alpha) + \delta_\alpha(x) \ \text{ in } J, \quad w_\alpha(\xi) = \eta + \alpha. \tag{$\star$}$$

The estimates given above show that $\eta - C + 2 \leq w_\alpha(x) \leq \eta + C - 1$ in J. We show first that $\phi < w_\alpha$ in J. Let $u = w_\alpha - \phi$. Then $|u(x)| < \alpha$ implies $(x, w_\alpha(x)), (x, \phi(x)) \in R$ and

$$u' = w'_\alpha - \phi' \geq f(x, w_\alpha) - f(x, \phi) + \delta_\alpha(x) \geq -\delta_\alpha(x) + \delta_\alpha(x) = 0.$$

It follows easily that $u \geq 0$, i.e., $\phi \leq w_\alpha$ in J.

Since $|w'_\alpha(x)| \le 3h(x)$, the family $\{w_\alpha\}$ is equicontinuous. Let (α_n) be a null sequence and (β_n) a subsequence such that the sequence (w_{β_n}) converges uniformly in J; cf. Theorem 7.IV. If $y^*(x) = \lim w_{\beta_n}(x)$, then clearly $\phi \le y^*$ in J. Now consider the integral equation for w_α equivalent to $(\star)$ and take $\alpha = \beta_n$. For $n \to \infty$, we obtain an integral equation for y^* which shows that y^* is a solution of the initial value problem for $y' = f(x,y)$. Here we have used that $\int_\xi^x \delta_{\beta_n}(t)\,dt \to 0$. This follows from Lebesgue's dominated convergence theorem since $\lim \delta_{\beta_n}(t) = 0$ (pointwise). ∎

Remarks. 1. It is now a matter of routine to show that under the assumptions of Theorem XX, case $n = 1$, there exists a maximal solution y^* and a minimal solution y_* and that both solutions approach the boundary of D to the left and to the right.

2. Theorem XXII extends to systems of differential equations which are quasimonotone in $\mathbf{y}$. However, the above proof needs modification. If the inequality $\boldsymbol{\phi} < \mathbf{w}_\alpha$ does not hold in J, then there is a first point $c > \xi$ such that $\boldsymbol{\phi} < \mathbf{w}_\alpha$ for $x < c$ and $\phi_i(c) = w_{\alpha i}(c)$. One considers $u = w_{\alpha i} - \phi_i$ in an interval $[c - \varepsilon, c]$, where $0 \le u \le \alpha$, and obtains $u' \ge 0$ in this interval using quasimonotonicity. This implies $u(c) > 0$, a contradiction.

3. Exercise. Let $\xi = \eta = 0$, $f(x,y) = y/x$ for $0 < x \le 1$, $x \log x \le y \le 0$ and $f(x,y) = 0$ for $y \ge 0$, $f(x,y) = \log x$ for $y \le x \log x$. Find the maximal and minimal solution and a negative function ψ satisfying $\psi' = f(x,\psi) + 1$, $\psi(0) = 0$.

XXIII. Solution Estimates. (a) *If $\mathbf{y}(x)$ is absolutely continuous in J, then so is $\phi(x) = |\mathbf{y}(x)|$, and*

$$|\phi'(x)| \le |\mathbf{y}'(x)| \quad \text{a.e. in} \quad J.$$

The absolute continuity follows from the inequality

$$|\phi(x_2) - \phi(x_1)| \le |\mathbf{y}(x_2) - \mathbf{y}(x_1)|$$

obtained using (5.2). Dividing both sides now by $x_2 - x_1 > 0$ and taking the limit as $x_2 \to x_1+$ or $x_1 \to x_2-$, leads to the inequality $|\phi'| \le |\mathbf{y}'|$. ∎

Theorem. *Let $\mathbf{y}$ be a C-solution of* (1) *in the interval $J = [\xi, \xi + a]$, and let the inequality*

$$|\mathbf{f}(x,\mathbf{y})| \le \omega(x, |\mathbf{y}|) \tag{16}$$

hold, where $\omega(x,r)$ is defined in $J \times [0,\infty)$ and satisfies a generalized local Lipschitz condition in r. Then any function $\rho \in AC(J)$ with the properties

$$\rho' \ge \omega(x,\rho) \quad \text{a.e. in} \quad J, \quad \rho(\xi) \ge |\mathbf{y}(\xi)|,$$

is an upper bound for the solution,

$$|\mathbf{y}(x)| \le \rho(x) \quad \text{in} \quad J. \tag{17}$$

The estimate (17) *holds in an interval* $J_- = [\xi - a, \xi]$ *to the left of* ξ *if*

$$\rho' \leq -\omega(x, \rho) \quad \text{a.e. in } J_-, \quad \rho(\xi) \geq |\mathbf{y}(\xi)|.$$

Proof. Let $\phi(x) = |\mathbf{y}(x)|$. From proposition (a) and (16), it follows that

$$\phi' \leq |\mathbf{f}(x, \mathbf{y})| \leq \omega(x, \phi).$$

The conclusion now follows from Theorem XXII with f replaced by ω. ∎

Exercise. Show that, with the Euclidean norm, the estimate (17) also holds under the weaker assumption $\langle \mathbf{y}, (\mathbf{f}(x, \mathbf{y})\rangle \leq |\mathbf{y}|_e \omega(x, |\mathbf{y}|_e)$. Here $\langle \cdot, \cdot \rangle$ is the scalar product in $\mathbb{R}^n$.

XXIV. Exercise. Separated Variables. Show that the Theorems 1.VII–VIII hold for the differential equation with separated variables

$$y' = f(x)g(y) \quad \text{under the assumption} \quad f \in L(J_x),\ g \in C(J_y)$$

and that the representation formula (1.8) remains valid.

XXV. Exercise. Solve the initial value problem ($n = 3$)

$$\mathbf{y}' = \begin{pmatrix} y_2 y_3 \\ -y_1 y_3 \\ 2 \end{pmatrix} \quad \text{with} \quad \mathbf{y}(0) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

by the method of successive approximation. What is the kth approximation if $\mathbf{y}_0(x) = \mathbf{y}(0)$?

§ 11. Initial Value Problems for Equations of Higher Order

I. Transformation to an Equivalent First Order System. Consider the nth order scalar differential equation in explicit form

$$y^{(n)} = f(x, y, y', \ldots, y^{(n-1)}) \quad (n \geq 1). \tag{1}$$

A function $y(x)$ is a solution of this equation in an interval J if it is n-times differentiable in J and satisfies equation (1) identically for $x \in J$. Equation (1) can be transformed into a system of n first order differential equations for n functions $y_1(x), \ldots, y_n(x)$:

$$\begin{aligned} y_1' &= y_2 \\ y_2' &= y_3 \\ \vdots\ &\quad \vdots \\ y_{n-1}' &= y_n \\ y_n' &= f(x, y_1, y_2, \ldots, y_n). \end{aligned} \tag{2}$$

Equation (1) and the system of equations (2) are equivalent in the following sense: If $y(x)$ is a solution of (1), then the vector function $\mathbf{y} = (y_1, y_2, \ldots, y_n) := (y, y', \ldots, y^{(n-1)})$ is a solution of (2). Conversely, if $\mathbf{y}$ is a (differentiable) solution of (2) and one sets $y_1(x) := y(x)$, then $y(x)$ is n-times differentiable, $y_2(x) = y'(x), \ldots, y_n(x) = y^{(n-1)}(x)$, and equation (1) holds.

In a similar manner, systems of nth order equations can be transformed into systems of first order equations by introducing new functions for the lower order derivatives. For example, the equation of motion of a point mass of mass 1 in three-dimensional space in the presence of a force field $\mathbf{k}(t, \mathbf{x})$ is given by

$$\ddot{\mathbf{x}} = \mathbf{k}(t, \mathbf{x}) \quad \text{for} \quad \mathbf{x} = \mathbf{x}(t),$$

or, written in expanded form with $\mathbf{x} = (x, y, z)$, $\mathbf{k} = (f, g, h)$,

$$\begin{aligned} \ddot{x} &= f(t, x, y, z) \\ \ddot{y} &= g(t, x, y, z) \\ \ddot{z} &= h(t, x, y, z). \end{aligned}$$

This system is equivalent to the following system of six first order differential equations for six unknown functions x, y, z, u, v, w.

$$\begin{aligned} \dot{x} &= u, & \dot{u} &= f(t, x, y, z) \\ \dot{y} &= v, & \dot{v} &= g(t, x, y, z) \\ \dot{z} &= w, & \dot{w} &= h(t, x, y, z). \end{aligned}$$

We return to equation (1). An initial value problem for the equivalent system (2) prescribes the values of the functions $y_1, \ldots, y_n$ at a point ξ. Corresponding initial conditions for (1) are given by

$$y(\xi) = \eta_0,\ y'(\xi) = \eta_1,\ \ldots,\ y^{(n-1)}(\xi) = \eta_{n-1}. \tag{3}$$

For example, the initial value problem for a second order differential equation reads

$$y'' = f(x, y, y'), \quad y(\xi) = \eta_0, \quad y'(\xi) = \eta_1.$$

Because of the equivalence of the two initial value problems, all of the theorems derived earlier for the system (2) can be applied to (1), (3). Here, the right-hand side of (2), $\mathbf{f} = (f_1, \ldots, f_n)$, takes the special form

$$\begin{aligned} f_i(x, y_1, \ldots, y_n) &= y_{i+1} \quad (i = 1, \ldots, n-1) \\ f_n(x, y_1, \ldots, y_n) &= f(x, y_1, \ldots, y_n). \end{aligned}$$

Thus $\mathbf{f}$ is continuous if f is continuous, holomorphic if f is holomorphic, and satisfies a Lipschitz condition if f satisfies a Lipschtiz condition

$$|f(x, \mathbf{y}) - f(x, \bar{\mathbf{y}})| \le L|\mathbf{y} - \bar{\mathbf{y}}|. \tag{4}$$

In particular, if $f(x, y_1, \ldots, y_n)$ is continuous in a domain $D \subset \mathbb{R}^{n+1}$ and if the partial derivatives $\partial f / \partial y_i$ are continuous in D, then both f and $\mathbf{f}$ satisfy a local Lipschitz condition in $\mathbf{y}$. We summarize.

II. Peano Existence Theorem. *Let $f(x, \mathbf{y})$ be continuous in a domain $D \subset \mathbb{R}^{n+1}$, and let $(\xi, \eta_0, \ldots, \eta_{n-1}) \in D$. Then the initial value problem* (1), (3) *has at least one solution. Every solution can be extended up to the boundary of D.*

III. Uniqueness Theorem. *Let f satisfy the hypotheses of Theorem* II. *If in addition, f satisfies a local Lipschitz condition* (4), *then there exists exactly one solution. This is the case, in particular, if $\partial f/\partial \mathbf{y} \in C(D)$.*

IV. Complex Existence Theorem. *Let $f(z, w_1, \ldots, w_n)$ be a function of $n+1$ complex variables z, $w_1, \ldots, w_n$. If f is defined and holomorphic in $D \subset \mathbb{C}^{n+1}$, then there exists exactly one function $w(z)$, defined and holomorphic in a neighborhood of the point $(z_0, \zeta_0, \zeta_1, \ldots, \zeta_{n-1}) \in D$, that satisfies the differential equation*

$$w^{(n)} = f(z, w, w', \ldots, w^{(n-1)})$$

and the initial conditions

$$w(z_0) = \zeta_0, \; w'(z_0) = \zeta_1, \; \ldots, \; w^{(n-1)}(z_0) = \zeta_{n-1}.$$

The solution can be expanded in a power series

$$w(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k.$$

For *example*, the ansatz $y(x) = a_0 + a_1 x + a_2 x^2 + \cdots$ in the initial value problem

$$y'' + y, \quad y(0) = \eta_0, \quad y'(0) = \eta_1 \tag{$\star$}$$

leads to the recursion formulas

$$a_0 = \eta_0, \quad a_1 = \eta_1, \quad k(k-1)a_k + a_{k-2} = 0 \quad (k = 2, 3, \ldots),$$

from which it is easy to see that

$$\eta_0 = 1, \quad \eta_1 = 0 \quad \Longrightarrow \quad y = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - + \cdots = \cos x,$$

$$\eta_0 = 0, \quad \eta_1 = 1 \quad \Longrightarrow \quad y = x - \frac{x^3}{3!} + \frac{x^5}{5!} - + \cdots = \sin x,$$

and hence the solution of problem ($\star$) is $y = \eta_0 \cos x + \eta_1 \sin x$.

In the next three parts we consider some differential equations of second order that either can be integrated directly or reduced to simpler differential equations.

V. $$\boxed{y'' = f(x, y')}$$

This equation is equivalent to a first order differential equation for $z(x) = y'(x)$,

$$z' = f(x, z).$$

If the initial condition is $y(\xi) = \eta_0$, $y'(\xi) = \eta_1$, one first looks for the solution z with $z(\xi) = \eta_1$. Then y is found by quadrature:

$$y(x) = \eta_0 + \int_\xi^x z(t)\,dt.$$

VI. $$\boxed{y'' = f(y, y')}$$

Suppose $y(x)$ is a solution and $x(y)$ its inverse. We introduce the function $p(y)$ defined by

$$p(y) = y'(x(y)).$$

Note that $dx(y)/dy = 1/p(y)$. The expression for the derivative of p, $dp/dy = y''(x(y))/p(y)$, leads to the first order differential equation

$$\frac{dp}{dy} = \frac{1}{p} f(y, p) \quad \text{for} \quad p = p(y).$$

Suppose $p(y)$ is a solution of this equation. Then we may calculate

$$x(y) = \int \frac{1}{p(y)}\,dy.$$

The solution $y(x)$ is the inverse of this function.

Example.

$$y'' = y'^2 \cdot \sin y, \quad y(0) = 0, \quad y'(0) = 1.$$

The equation for $p(y)$ is

$$dp/dy = p \cdot \sin y \quad \text{with} \quad p(0) = y'(0) = 1.$$

Therefore

$$p(y) = \mathrm{e}^{1-\cos y} \quad \Rightarrow \quad x(y) = \int_0^y \mathrm{e}^{\cos s - 1}\,ds.$$

VII. $\boxed{y'' = f(y)}$

This is a special case of VI. Suppose f is continuous, $y(x)$ is a solution, and $F(y) = \int f(y)\,dy$ is a primitive of $f(y)$.

Multiplying the differential equation by $2y'$, we obtain

$$(y'^2)' = 2y'y'' = 2y'f(y) = 2\frac{d}{dy}F(y(x)),$$

which implies

$$y'^2 = 2F(y) + C.$$

Solving for y' gives a differential equation

$$y' = \pm\sqrt{2F(y) + C},$$

which does not depend explicitly on the independent variable.

If $\eta_1 \neq 0$, then the value of the constant C and the sign are uniquely determined by the initial condition $y(0) = \eta_0$, $y'(0) = \eta_1$, and the differential equation has a unique solution $y(x)$ with $y(0) = \eta_0$ (Theorem 1.VII). In the case where $\eta_1 = 0$, the situation is somewhat more complicated, and there may be more than one solution.

In the remainder of this section, we discuss some examples from physics.

VIII. The Catenary. A flexible chain or cable with no stiffness, suspended from two points, hangs under the influence of gravity in a shape called a *catenary* (from Latin *catena*, chain). Let (a, y_a), (b, y_b), $a < b$, be the two suspension points in the xy-plane, ρ the density (mass per unit length) of the chain, which is assumed to be constant, and g the constant of gravitational acceleration. We choose a reference frame such that the gravitational force operates in the direction of the negative y-axis. In order to determine the function $y(x)$ that describes the shape of the curve, we conduct a thought experiment: We cut the chain at a point $(x, y(x))$ and remove and replace the right-hand part by a force $\mathbf{k}_x = (H_x, V_x)$ in such a way that the left-hand part of the chain remains at rest. Then

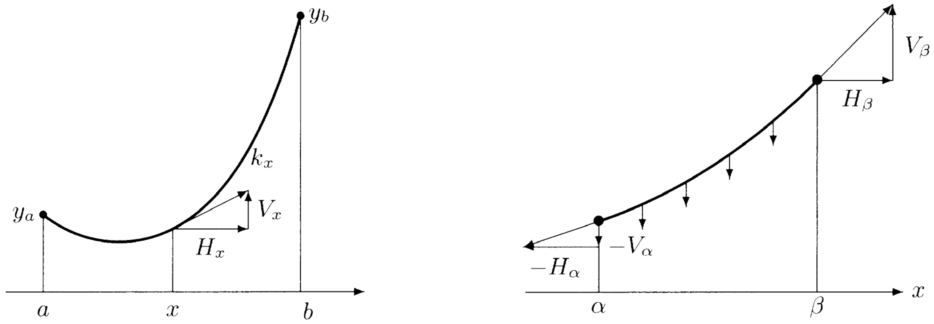
$$\mathbf{k}_x = (H_x, H_y) \quad \text{with} \quad y'(x) = \frac{V_x}{H_x}$$

(this relation reflects the assumption that the chain has no stiffness). If, on the other hand, the left-hand part is removed and the right retained, then an opposite force $-\mathbf{k}_x$ is required.

Now consider the segment of the chain between α and β. It is kept at rest by forces $-\mathbf{k}_\alpha = -(H_\alpha, V_\alpha)$ on the left and $\mathbf{k}_\beta = (H_\beta, V_\beta)$ on the right. Since the system is at rest, the sum of the forces is zero. Separating into horizontal and vertical components, one is led to a pair of equations

$$H_\beta - H_\alpha = 0 \implies H_x = \text{const.} = H,$$

$$V_\beta - V_\alpha - g\int_\alpha^\beta \rho\sqrt{1 + y'^2}\,dx = 0.$$



The integral in the second equation measures the length L of the curve segment (cf. A.I), and the term $\rho g L$ gives the downward pull due to gravity. The second equation can be rewritten in the form

$$\int_\alpha^\beta \left(Hy'' - \rho g \sqrt{1+y'^2} \right) dx = 0,$$

since $V_\beta - V_\alpha = \int_\alpha^\beta V'(x)\,dx$ and $V = Hy'$. Now we use a well-known argument: Since α, β can be chosen arbitrarily, the integrand must vanish, and it follows that

$$y'' = c\sqrt{1+y'^2} \quad \text{with} \quad c = \frac{\rho g}{H} > 0 \qquad \textit{Catenary equation.}$$

This equation belongs to the type studied in V. From the differential equation $z' = c\sqrt{1+z^2}$ for $z = y'$ one easily obtains $\operatorname{arcsinh} z = c(x+A)$ and hence $z = y' = \sinh c(x+A)$. The general solution is then given by

$$y = B + \frac{1}{c}\cosh c(x+A) \quad (A,\ B \text{ arbitrary}).$$

Remark. The equation of the catenary (and the above derivation) remains valid if the density ρ is variable.

The Boundary Value Problem. Suppose the distance $b-a$ and $y_b - y_a$ are given, together with the length of the cable L; naturally, we require that

$$L^2 > (b-a)^2 + (y_b - y_a)^2. \tag{5}$$

Without loss of generality, let $A = 0$, i.e., a reference frame is chosen such that the minimum of y is at $x = 0$. Then we have

$$\begin{aligned} L &= \int_a^b \sqrt{1+y'^2} = \frac{1}{c}\int_a^b y''\,dx = \frac{1}{c}[y'(b) - y'(a)] \\ &= \frac{1}{c}[\sinh cb - \sinh ca] = \frac{2}{c}\sinh\frac{c(b-a)}{2}\cdot\cosh\frac{c(a+b)}{2} \end{aligned} \tag{6}$$

and

$$y_b - y_a = \frac{1}{c}[\cosh cb - \cosh ca] = \frac{2}{c}\sinh\frac{c(b-a)}{2} \cdot \sinh\frac{c(a+b)}{2}. \tag{7}$$

From this pair of equations, it follows that

$$L' := \sqrt{L^2 - (y_b - y_a)^2} = \frac{2}{c}\sinh\frac{c(b-a)}{2}. \tag{8}$$

The value of c can be determined from this equation. The equation is first written in the form of a fixed point equation. $\xi = \phi(\xi)$ can then be solved by iteration. Setting $\xi = c/2$ one obtains $\xi = (\sinh(b-a)\xi)/L'$, or, in terms of the inverse function,

$$\xi = \frac{1}{b-a}\sinh^{-1}(L'\xi). \tag{9}$$

If $\phi(\xi)$ denotes the right-hand side of (9), then $L' > b - a$ implies that $\phi'(0) > 1$. Since arcsinh (t) is concave for $t \geq 0$ and grows like $\ln(t)$ for large t, there exists exactly one positive fixed point ξ. The reader should verify using a sketch that the sequence (x_n) obtained by iteration $x_{n+1} = \phi(x_n)$ converges to ξ for every starting value $x_0 > 0$. In the special case $y_a = y_b$, we have $a = -b < 0$ and $L = L'$. In the general case, once c has been determined, a can be determined from (7), (8), and $a + b = (b-a) + 2a$:

$$a = \frac{1}{c}\operatorname{arcsinh}\frac{y_b - y_a}{L'} - \frac{b-a}{2}. \tag{10}$$

Historical Remark. In the second discussion in the *Discorsi* (1638), Galileo expressed the opinion that the catenary was a parabola. The true form was found, independently, in 1691 by Leibniz, Johann Bernoulli, and Huygens. Leibniz originated the name *catenary.*

Example. Let $b-a = 200$, $y_b - y_a = 100$, $L = 240$, hence $L' = 20\sqrt{144 - 25} = 218.1742$. By iteration in (9), one obtains $c = 2\xi = 0.007287$ and $a = -39.114$. Thus the chain is described by the function

$$y(x) = 137.237 \cdot \cosh(0.007287x) + B,$$

its sag $S = \min(y_a, y_b) - \min y(x)$ is given by

$$S = y(a) - y(0) = \frac{1}{c}(\cosh ca - 1) = 5.612.$$

IX. Catenary Problems. (a) For fixed $b - a$ and $y_b - y_a$, the horizontal component of the force H is a function of L. Show: $H = H(L)$ is monotone decreasing and tends to ∞ as $L \to \sqrt{(b-a)^2 + (y_b - y_a)^2}$ and toward 0 as $L \to \infty$.

(b) Let $b - a = 200$, $y_a = y_b$, $L = 240$. How large is the sag?

(c) Let $b - a = 200$, $y_a = y_b$, $S = 20$. How large is L?

(d) Let $y_a = y_b$ and span = sag = 1 m. How long is the chain?

(e) Prove: If $y(x) = (1/c)\cosh cx$ describes a catenary and $z(x) = \alpha + \beta x^2$ is a parabola with $y(0) = z(0)$, $y(b) = z(b)$, then $y < z$ in $(0, b)$.

(f) (*Galileo Vindicated?*) For which density function $\rho = \rho(x)$ is the parabola $y = \alpha + \beta x^2$ a solution of the catenary equation? Interpret this result in terms of (e).

X. Nonlinear Oscillations. We consider the differential equation

$$\ddot{x} + h(x) = 0 \quad \text{for} \quad x = x(t). \tag{11}$$

Here h is locally Lipschitz continuous in $\mathbb{R}$ and satisfies the conditions $h(0) = 0$ and $x \cdot h(x) > 0$ for $x \neq 0$. This equation is of type VII.

(a) If $x(t)$ is a solution of (11), then the functions $x(t+c)$ and $x(-t)$ are also solutions.

By VII, the function

$$E(x, \dot{x}) = \frac{1}{2}\dot{x}^2 + H(x) \quad \text{with} \quad H(x) = \int_0^x h(s)\,ds \tag{12}$$

is constant for every solution of (11), and for the initial conditions $x(0) = x_0$, $\dot{x}(0) = v_0$, $E(x, \dot{x}) = E(x_0, v_0) =: \alpha$. For $v_0 > 0$, we have $\dot{x} = \sqrt{2(\alpha - H(x))}$, and using 1.V, we obtain the solution in the form

$$\sqrt{2}t = \int_{x_0}^{x(t)} \frac{ds}{\sqrt{\alpha - H(s)}} \quad \text{with} \quad \alpha = E(x_0, v_0) > 0. \tag{13}$$

In applications to mechanics, $x(t)$ describes the motion of a point mass of mass 1, $x = 0$ corresponds to the equilibrium state, and $-h(x)$ gives the magnitude of a "restoring force" (its sign is opposite to the displacement x, whence its name). The function E in (12) is the *total energy*, the sum of the kinetic energy $\frac{1}{2}\dot{x}^2$ and the potential energy $H(x)$ (the work done in moving the mass from x to $x + dx$ is $h(x)\,dx$). The equation $E(x(t), \dot{x}(t)) = \text{const}$ has the following interpretation: In the oscillations of a frictionless system, like the one under consideration here, there is a continual exchange between kinetic and potential energy, while the total energy remains constant.

Equation (11) can be written as an autonomous system

$$\dot{x} = y, \quad \dot{y} = -h(x). \tag{11$'$}$$

In 3.V, we described how to construct a phase portrait, from which the qualitative behavior of the solutions of such systems can be read. The trajectories in the xy-plane are level sets of $E(x, y)$. Clearly, $H(x) > 0$ for $x \neq 0$. Therefore, $0 = E(0, 0) < E(x, y)$ for $(x, y) \neq 0$.

Theorem. *Suppose $H(x) \to \infty$ as $x \to \pm\infty$. Then every solution to the differential equation* (11) *is periodic. A solution $x(t)$ takes on its extreme values when $\dot{x}(t) = 0$; the extreme values of $\dot{x}(t)$ occur when $x(t) = 0$.*

Proof. By (11′), $x(t)$ is strongly monotone increasing for $y > 0$ and decreasing for $y < 0$; a similar statement holds for $y(t)$. This implies the statement about the extreme values. The remaining conclusions follow from A.VIII–IX. We indicate a direct proof.

The equation $E(x,y) = \alpha > 0$ describes a closed Jordan curve K_α that surrounds the origin and is symmetric to the x-axis and that is represented by

$$y = \pm\sqrt{2(\alpha - H(x))} \quad \text{for} \quad r_1 \le x \le r_2,$$

where $r_1 < 0 < r_2$ and $H(r_1) = H(r_2) = \alpha$. These values are uniquely determined because of the strong monotonicity of H in the intervals $(-\infty, 0]$ and $[0,\infty)$.

If $(x(t), y(t))$ is a solution of (11′) with the initial values $x(0) = x_0 > 0$, $y(0) = y_0 = 0$, then since the curve K_α is bounded, the solution exists in $\mathbb{R}$; cf. 10.XI.(c). Using an argument similar to the one used for the predator–prey model in 3.VI, one shows that the solution runs over the entire curve K_α and is periodic. (Initially $\dot{y} < 0$; $y(t)$ is strongly monotone decreasing, as long as $x(t) > 0$; from $y \le -\varepsilon$ it follows that $\dot{x} \le -\varepsilon$, hence there is a $t_1 > 0$ with $x(t_1) = 0$; etc.) ∎

We discuss some examples.

(b) *The Harmonic Oscillator.* In the linear case $h(x) = \omega^2 x$ with $\omega > 0$, the equation is

$$\ddot{x} + \omega^2 x = 0 \qquad \textit{harmonic oscillator equation.}$$

The general solution is $x(t) = r \sin \omega(t+c)$ $(r \ge 0)$, and the total energy is given by $2E(x,y) = y^2 + \omega^2 x^2$. Thus the trajectories are ellipses $y^2 + \omega^2 x^2 = 2\alpha$. The (minimal) period $T = 2\pi/\omega$ is the same for all solutions. The number $\nu = 1/T = \omega/(2\pi)$ is called the *frequency* (the number of oscillations per second), ω the *circular frequency*, and r the *amplitude* of the oscillation.

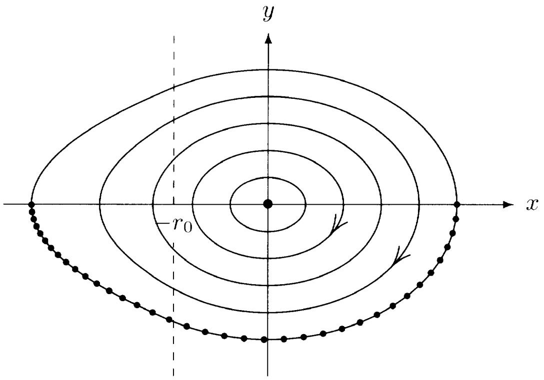
In mathematical models of elastic objects, a linear relationship between displacement and tension is known as Hooke's law. The classical example is a mass hanging on a spiral spring; here $k = \omega^2$ is the spring constant.

(c) *Mass on a Rubber Band.* Suppose a rubber band is fixed at the upper end A and hangs vertically downward. Coordinates are chosen such that the x-axis is oriented downward and the position of the lower end B, when the system is at rest, is at the origin. If B is pulled downward, then the restoring force is given by Hooke's law $h(x) = kx$ with $k > 0$. On the other hand, if B is pushed upward, then, in contrast to the case of the spiral spring, there is no restoring force. A mass m is now attached at B. The vertical motion, described by $\xi(t)$, satisfies the equation

$$m\ddot{\xi} = mg - k\xi_+ \quad \text{with} \quad \xi_+ = \max\{\xi, 0\}$$

(g is the constant of gravitational force). The equilibrium position r_0 is given by $r_0 = mg/k$. Writing $\xi(t) = r_0 + x(t)$ and setting $b = k/m > 0$, we obtain the differential equation

$$\ddot{x} + h(x) = 0 \quad \text{with} \quad h(x) = (g + bx)_+ - g$$



Mass on a rubber band

(the reader should make a sketch of h). In this example, the potential is

$$H(x) = \begin{cases} \frac{1}{2}bx^2 & \text{for} \quad x \geq -r_0, \\ \frac{1}{2}g(r_0 + 2x) & \text{for} \quad x < -r_0. \end{cases}$$

For small perturbations, the orbits

$$K_\alpha : \quad E(x, y) = \frac{1}{2}y^2 + H(x) = \alpha \quad \text{with} \quad \alpha > 0$$

are ellipses, and the system behaves like a harmonic oscillator. However, if $\alpha > H(r_0)$, then $\xi(t) = r_0 + x(t)$ also takes on negative values, and the orbit is a combination of an ellipse and a parabola.

The above example is typical of mechanical systems that are not symmetric with respect to their rest state. An interesting system of this type is a suspension bridge, where a linear theory is appropriate for small (vertical, torsional, ...) oscillations but fails for large amplitudes. A nonlinear theory of the dynamics of suspension bridges is currently under investigation; cf. in this regard McKenna and Walter (1990) and Lazer and McKenna (1990).

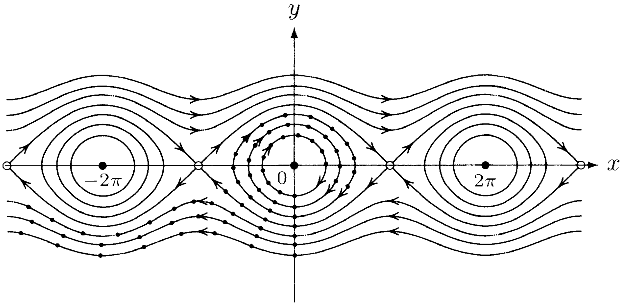
(d) *The Mathematical Pendulum.* One end A of a (weightless) rod of length l is attached to a pivot, and a mass m is attached to the other end B. The system moves in a plane under the influence of the gravitational force of magnitude mg, which acts vertically downward. If ϕ denotes the angle between the vertical and the rod, then the tangential component of the downward force $-mg \sin\phi$ acts on the mass point at B. If the motion is described in terms of the angle $\phi(t)$, then $s(t) = l\phi(t)$ gives the distance traveled by the mass (measured from the lowest point), and from the equation of motion $m\ddot{s} = -mg\sin\phi$, we obtain

$$\ddot{\phi} + a\sin\phi = 0 \quad \text{with} \quad a = g/l \qquad \textit{mathematical pendulum.}$$

For small values of $|\phi|$, one may replace $\sin\phi$ by $\phi \approx \sin\phi$, which leads to

$$\ddot{\phi} + a\phi = 0 \quad \text{with} \quad a = g/l \qquad \textit{linearized pendulum}$$

as a usable approximation. Thus, for small displacements, the pendulum behaves like a harmonic oscillator with circular frequency $\omega = \sqrt{g/l}$.



Mathematical pendulum

The potential energy for the mathematical pendulum is given by $H(\phi) = a(1 - \cos\phi)$, $0 \leq H(\phi) \leq 2a$, and the total energy is $E(\phi, \dot{\phi}) = \frac{1}{2}\dot{\phi}^2 + H(\phi) \geq 0$. The assumption $\phi \cdot h(\phi) > 0$, mentioned in earlier results, is violated here. It holds only for $|\phi| < \pi$, and accordingly, the level sets

$$K_\alpha : \quad \dot{\phi}^2 = 2(\alpha - a(1 - \cos\phi))$$

in the $(\phi, \dot{\phi})$-plane are closed curves only for $0 < \alpha < 2a =: \alpha_0$. These closed curves correspond to periodic oscillations with a maximum angular displacement $< \pi$. For $\alpha = 0$, one obtains stationary solutions $\phi(t) \equiv 2k\pi$. If $\alpha > \alpha_0$, the orbits K_α are unbounded wavy lines, corresponding to continual rotations about the pivot point A. Such are obtained from the initial conditions $\phi(0) = 0$, $\dot{\phi}(0) = v_0$ with $\frac{1}{2}v_0^2 > \alpha_0$ (the upper wavy lines correspond to $v_0 > 0$, the lower wavy lines to $v_0 < 0$.) Solutions in the two cases $0 < \alpha < \alpha_0$ and $\alpha > \alpha_0$ have completely different behavior and are divided by the set K_{α_0}, which is called the *separatrix*. The solution $\phi(t)$ that corresponds to the separatrix, for instance the one with initial values $\phi(0) = 0$, $\dot{\phi}(0) = \sqrt{2\alpha_0}$, is monotone increasing and tends toward π as $t \to \infty$ (the reader should give a description of the pendulum motion for this case).

Historical Remark. Galileo writes in the first discussion of the *Discorsi* (1638) that the frequency of a pendulum does not depend on the maximal angular diflection and is proportional to $\sqrt{1/l}$. He is correct in the second assertion; cf. XI.(j) below.

XI. Exercises on Nonlinear Oscillations. Suppose h is continuous in $\mathbb{R}$ and satisfies $x \cdot h(x) > 0$ for $x \neq 0$. Let H and E be defined as in X.

(a) Prove: If h is continuous, then every initial value problem for (11) is uniquely solvable (cf. Exercise XIII). Theorem X remains true.

(b) Let $H(x)$ tend toward A as $x \to -\infty$ and toward B as $x \to \infty$, with $0 < A \leq B < \infty$. Describe the global behavior (periodicity, behavior for large $|t|$) of the solution $x(t)$ with initial values $x(0) = \xi$, $\dot{x}(0) = \eta$ in each of the following cases: (i) $0 < E(\xi, \eta) < A$. (ii) $A \leq E(\xi, \eta) < B$. (iii) $E(\xi, \eta) \geq B$. Make a sketch of the phase portrait.

(c) For which values of A, B (cf. (b)) is the following statement true? Every solution of the differential equation that has a zero and whose first derivative has a zero is periodic.

(d) Suppose, additionally, that h is odd. Prove: If $x(t)$ is a solution to the differential equation, then $v(t) := x(c-t)$ and $w(t) := -x(t)$ are also solutions. Further, the relations

$$x(c) = 0 \Longrightarrow x(c+t) = -x(c-t),$$
$$\dot{x}(d) = 0 \Longrightarrow x(d+t) = x(d-t)$$

hold for every solution x. Thus the solutions have the same symmetry properties as the sine function.

(e) *The Period of an Oscillation.* Let V denote the duration of the positive quarter oscillation starting at the point $(0, v_0)$ and ending at $(r, 0)$ ($v_0 > 0$, $r > 0$ is the largest swing). Using (13) with $\alpha = H(r) = \frac{1}{2}v_0^2$, we get

$$V = V(r) = \frac{1}{\sqrt{2}} \int_0^r \frac{ds}{\sqrt{H(r) - H(s)}}.$$

(f) Prove: If h is a symmetric forcing term (h odd), then $T = 4V$ is the period of an oscillation.

(g) Prove: If $h(x) \le h^*(x)$ for $x > 0$, then $V(r) \ge V^*(r)$. If, in addition, $h(c) < h^*(c)$ for some $c > 0$, then $V(r) > V^*(r)$ holds for $r \ge c$.

(h) Prove: If $h(x)/x$ is weakly or strongly increasing [decreasing], then $V(r)$ is weakly or strongly decreasing [increasing]. As an illustration, calculate $V(r)$ for the case $h(x) = x^\alpha$ ($\alpha > 0$).

(i) Prove: If $h'_+(0) = \omega^2$ exists, then $\lim_{r\to 0} V(r) = \pi/(2\omega)$, i.e., for small displacements the system is approximately a harmonic oscillator.

(j) *The Mathematical Pendulum.* Show that for the mathematical pendulum (cf. X.(d)) (using $1 - \cos\alpha = 2\sin^2\dfrac{\alpha}{2}$),

$$V(r) = \frac{1}{2\sqrt{a_0}} \int_0^r \frac{ds}{\sqrt{\sin^2\frac{r}{2} - \sin^2\frac{s}{2}}} = \frac{1}{a} \int_0^{\pi/2} \frac{du}{\sqrt{1 - k^2\sin^2 u}},$$

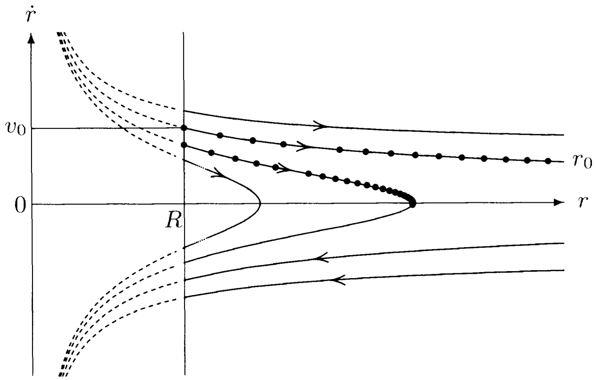
where $k = \sin r/2$ (substitution $\sin s/2 = k \sin u$). This function is an elliptic integral of the first kind. From the binomial expansion for $(1-x)^{-1/2}$, it follows that

$$V(r) = \frac{\pi}{2\sqrt{a_0}}(1 + a_1k^2 + a_2k^4 + \cdots);$$

in particular, $V(0) = \pi/(2\sqrt{a})$, in agreement with (i) and X.(b) ($a = \omega^2$ here). Calculate a_1, a_2. By what percentage is the period of an oscillation of the mathematical pendulum larger than that of the linearized pendulum equation if the maximum displacement is 5° (10°; 15°; 20°)?

Hints: (g) Consider the difference $H(r) - H(s)$.

(h) Consider $V(r)$ and $V(qr)$ with $q > 1$ and write $V(qr)$ as an integral from 0 to r. The quotient in the corresponding H-differences has the form



Free fall

$[f(r) - f(s)]/[g(r) - g(s)]$ with $f(x) = H(qx)$, $g(x) = q^2 H(x)$. The generalized mean value theorem of differential calculus can be used.

(i) Use (g).

Remark. The dependence of the period of the oscillation on the function h was investigated thoroughly in 1961 by Z. Opial (*Ann. Polon. Math.* 10, 42–72); (g) and (h) can be found there. The book by Reissig, Sansone, and Conti (1963) contains these results and a number of others.

XII. Free Fall. In part II of the Introduction, we presented the example of free fall from a great height. The initial value problem

$$m\ddot r = -\gamma \frac{Mm}{r^2}, \quad r(0) = R, \quad \dot r(0) = v_0$$

describes the vertical motion of a body with mass m starting at a distance R from the center of the earth with initial velocity v_0, where $r(t)$ is the distance from the center of the earth at time t. From the equation

$$\frac{d}{dt}\dot r^2 = -2\gamma M \frac{\dot r}{r^2}$$

it follows that

$$E(r, \dot r) = \frac{1}{2}\dot r^2 - \gamma \frac{M}{r} \qquad \textit{total energy function}$$

is constant. Here $\frac{1}{2}\dot r^2$ is the kinetic energy term and $-\gamma M/r$ the potential energy. The latter is normalized in such a way that it vanishes at infinity and hence is negative. The trajectories satisfy the equation $E(r, \dot r) = \alpha$. If $\alpha \geq 0$, the trajectories run to infinity; for $\alpha < 0$, they are return curves (describing a body that falls back to the earth after being thrown vertically upward).

The smallest total energy for a motion without return is $\alpha = 0$. The corresponding differential equation reads

$$\dot r = \sqrt{2\gamma M/r}. \tag{14}$$

It has the solutions

$$r_0 = a(t+c)^{2/3}, \quad a = \sqrt[3]{9\gamma M/2}.$$

If R denotes the radius of the earth, then it follows from (14) that the velocity of the corresponding motion at $r = R$ is given by

$$v_0 = \sqrt{2\gamma M/R} \approx 11.2\,\mathrm{km/sec}.$$

This is the so-called 'escape velocity', the minimum velocity that a rocket requires in order to escape the gravitational field of the earth.

XIII. Exercise. Let $f \in C(J)$. Show: The initial value problem

$$y'' = f(y), \quad y(0) = \eta_0 \in J, \quad y'(0) = \eta_1$$

is locally uniquely solvable in each of the following three cases.

(a) $\eta_1 \neq 0$.

(b) $\eta_1 = 0$, $f(\eta_0) \neq 0$.

(c) $\eta_1 = 0$, $f(\eta_0) = 0$, $(y - \eta_0)(f(y) - f(\eta_0)) \leq 0$ for y close to η_0.

(d) Give all of the solutions for the initial value problem $y'' = \sqrt{|y|}$, $y(0) = y'(0) = 0$.

XIV. Exercise. (a) Let the function $f(x,y)$ be continuous in the strip $J \times \mathbb{R}$ (J an open interval), and let $f(x,y) \cdot y > 0$ for $y \neq 0$. Let N be the number of zeros and E the number of local extrema of a solution $y(x)$ of the differential equation

$$y'' = f(x,y).$$

Prove: If solutions to initial value problems with the initial conditions $y(\xi) = y'(\xi) = 0$ ($\xi \in J$) are unique, then $N + E \leq 1$ for every solution $y \not\equiv 0$ of the differential equation. Find and sketch all solutions of the initial value problem

$$y'' = (\operatorname{sgn} y)\sqrt{|y|}, \quad y(0) = y'(0) = 0$$

(to the left and to the right). This example shows that without the uniqueness assumption the inequality $N + E \leq 1$ is not always true.

(b) Investigate the differential equation

$$y'' = -f(x,y)$$

under the same assumptions and show that between two successive zeros of a solution there is exactly one extremum, that at each zero there is a point of inflection, and that the set of zeros does not have a point of accumulation in J. Which of these conclusions is false without the uniqueness assumption introduced in (a)?

XV. Exercise. Let $g \in C(J)$, $J = [0, b]$, and

$$u(x) = \int_0^x k(x-t)g(t)\,dt, \quad x \in J,$$

where $k = k(x;\lambda)$ $(\lambda \in \mathbb{R})$ is the solution of

$$k'' - \lambda k = 0, \quad k(0) = 0, \quad k'(0) = 1.$$

(a) Find $k(x;\lambda)$ explicitly and show that u is a solution of

$$u'' - \lambda u = g(x) \quad \text{in} \quad J, \quad u(0) = u'(0) = 0.$$

(b) Let $f(x, u)$ be continuous in $J \times \mathbb{R}$. Show that u is a continuous solution of

$$u(x) = \int_0^x k(x-t)f(t, u(t))\,dt \quad \text{in} \quad J$$

if and only if it is a C^2-solution of

$$u'' - \lambda u = f(x, u) \quad \text{in} \quad J, \quad u(0) = u'(0) = 0.$$

Supplement: Second Order Differential Inequalities

We consider comparison theorems related to the initial value problem

$$y'' = f(x, y, y') \quad \text{in} \quad J = [\xi, b], \quad y(\xi) = \eta_0, \quad y'(\xi) = \eta_1. \tag{15}$$

It is assumed for simplicity that f is defined in $J \times \mathbb{R}^2$. The differential equation is equivalent to a system for $\mathbf{y} = (y_1, y_2) = (y, y')$,

$$y_1' = y_2, \quad y_2' = f(x, y_1, y_2),$$

which is quasimonotone increasing if $f = f(x, y, p)$ is increasing in y.

XVI. Comparison Theorem. *Assume that $f(x, y, p)$ is (weakly) increasing in y and that $v, w \in C^2(J)$, $J = [\xi, b]$. Then the inequalities*

(a) *$Pv < Pw$ in J, where $Pv = v'' - f(x, v, v')$,*

(b) *$v(\xi) \le w(\xi)$, $v'(\xi) \le w'(\xi)$*

imply that $v < w$ and $v' < w'$ in $(\xi, b]$.

If f satisfies a local Lipschitz condition in y and p, then the inequalities (a′) *$Pv \le Pw$ in J (instead of* (a)*) and* (b) *imply that $v \le w$ and $v' \le w'$ in J.*

Sketch of the *proof*: Assume that the conclusion is false. Since $v'(\xi) = w'(\xi)$ implies $v''(\xi) < w''(\xi)$ by (a), it follows that $v' < w'$ in a largest interval (ξ, c) with $\xi < c \le b$ and $v'(c) = w'(c)$. Therefore, $v''(c) \ge w''(c)$ (see proof in 9.I). This together with the monotonicity of f and $v(c) < w(c)$ leads to $(Pv)(c) \ge (Pw)(c)$, contradictory to (a). ∎

Applications. 1. *Upper and Lower Solutions* for problem (15) are now introduced in the usual way. If $f(x,y,p)$ is increasing in y and satisfies a local Lipschitz condition in y and p, then

$$w'' \geq f(x,w,w') \ \text{ in } \ J, \quad w(\xi) \geq \eta_0,\ w'(\xi) \geq \eta_1 \quad \textit{(upper solution)}$$

implies $w \geq y$ and $w' \geq y'$ in J, where y is the solution of problem (15).

2. *Maximal and Minimal Solutions.* If $f(x,y,p)$ is continuous and increasing in y, then the initial value problem (15) has a maximal solution y^* and a minimal solution y_* with the property that $y_* \leq y \leq y^*$ and $y_*' \leq y' \leq y^{*\prime}$ for any other solution y. The maximal solution is constructed as a limit of solutions w_n of problem (15) in which f is replaced by $f(x,y,y') + 1/n$. The sequence (w_n) is decreasing and converges to a solution y^* with the properties just stated. Furthermore, the following comparison theorem holds:

$$v'' \leq f(x,v,v'),\ v(\xi) \leq \eta_0,\ v'(\xi) \leq \eta_1 \Longrightarrow v \leq y^*,\ v' \leq y^{*\prime} \quad \text{in} \quad J.$$

Exercise. Prove these statements; formulate and prove similar propositions for lower solutions.

3. The comparison theorem holds in an interval $[a,\xi]$ to the left of ξ if (b) is replaced by (b$_-$) $v(\xi) \leq w(\xi)$, $v'(\xi) \geq w'(\xi)$.

4. These results extend to the case where f is continuous in an open set $D \subset \mathbb{R}^3$ and increasing in y. In particular, the maximal and minimal solutions can be extended to the right and left up to the boundary of D.

5. *Exercise.* Show using Exercise XV that for $u \in C^2(J)$, $g \in C(J)$, $J = [0,v]$, $\lambda > 0$,

$$u'' - \lambda u \leq g(x) \quad \text{in} \quad J, \quad u(0) \leq 0, \quad u'(0) \leq 0$$

implies

$$u(x) \leq \frac{1}{\sqrt{\lambda}} \int_0^x \sinh \sqrt{\lambda}\,(x-t) g(t)\, dt.$$

XVII. Nonlinear Differential Operators. The Δ_p Operator. We introduce instead of y'' a nonlinear operator $Ly = (\phi(x,y'))'$ and consider a corresponding equation $Ly = f(x,y)$ with the defect operator $Pv = Lv - f(x,v)$.

Comparison Theorem. *Let $J = [\xi,b]$, $J_0 = (\xi,b]$, and assume that $f(x,y)$ is increasing in y and $\phi(x,p)$ is strictly increasing in p. If $v,w \in C^1(J)$ with $\phi(x,v')$, $\phi(x,w') \in C^0(J) \cap C^1(J_0)$ satisfy*

$$Pv \leq Pw \quad \textit{in } J_0, \quad \textit{and} \quad v(\xi) < w(\xi), \quad v'(\xi) \leq w'(\xi),$$

then $v < w$ and $v' \leq w'$ in J.

Exercise. Make a proof along the lines $v < w \Longrightarrow Lv \leq Lw \Longrightarrow \phi(x,v') \leq \phi(x,w') \Longrightarrow v' \leq w' \Longrightarrow w - v$ increasing.

As an application, we consider the Δ_p operator (p-Laplacian) in $\mathbb{R}^n$,

$$\Delta_p u = \operatorname{div}\left(|\nabla u|^{p-2}\nabla u\right) \quad \text{for} \quad p > 1 \quad (\nabla u = \operatorname{grad} u)$$

(note that Δ_2 is the classical Δ operator), which for radial functions $u = u(r)$ is given by

$$\Delta_p u = r^{1-n}(r^{n-1}|u'|^{p-2}u')', \quad r = |x| \quad (x \in \mathbb{R}^n).$$

Using the operator $L_\alpha^p y = r^{-\alpha}(r^\alpha |y'|^{p-2}y')'$ (note that L_α^2 is the operator L_α introduced in 6.XII), one can deal with the Δ_p operator for radial solutions as was done in 6.XIV with equations involving Δu.

Theorem. *Let f be continuous and bounded in $J \times \mathbb{R}$, $J = [0, b]$, $J_0 = (0, b]$, and $\alpha > 0$, $p > 1$. Then the problem*

$$L_\alpha^p y = f(x, y) \quad \text{in } J_0, \quad y(0) = \eta, \quad y'(0) = 0$$

has a solution $y \in C^1(J)$ with $L_\alpha^p y \in C(J)$. If f is increasing in y, then the comparison theorem (with $v'(0) = w'(0) = 0$) holds for this equation.

The existence proof uses Schauder's fixed point theorem 7.XII. Uniqueness is more difficult. Show: The problem $L_\alpha^4 u = u$, $u(0) = u'(0) = 0$ has three solutions of the form $u = c \cdot x^2$.

§ 12. Continuous Dependence of Solutions

The problem discussed here arises naturally in the modeling of physical processes using differential equations. Numerical values, representing physical quantities, enter into the differential equation and the initial conditions (initial position, initial velocity, mass, gravitational constant, ...). These quantities are obtained from measurements and, consequently, are not precisely known. One would require, based upon experience with the physical problems, that the solutions (say, to an initial value problem that models the motion of an object) are "insensitive" to small changes in these numerical values. This idea is given a more precise formulation and investigated in this section.

I. Well-Posed Problems. A mathematical problem used to model a well-defined physical process that proceeds in a unique manner should satisfy three general requirements.

(a) *Existence.* The problem has at least one solution.

(b) *Uniqueness.* The problem has not more than one solution.

(c) *Continuous Dependence.* The solution depends continuously on the data that are present. For instance, if the model results in an initial value problem, then the solution depends continuously on the right side of the differential equation and on the initial values. Or to put it another way: If a "small" change is made on the right side or in the initial values, then the solution changes only a "little."

A problem is called *well-posed* if it has the properties (a), (b), and (c). Our aim is to prove well-posedness for the initial value problem $\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$, $\mathbf{y}(\xi) = \boldsymbol{\eta}$. Since the requirements (a), (b) have already been investigated, we deal primarily with requirement (c) in this section. Let us remark in closing that there are instances where a small change in a parameter introduces important new physical phenomena (like resonance), so that (c) is not always a physically meaningful requirement.

Historical Remark. Existence and uniqueness for the initial value problem were rigorously treated by Cauchy as early as around 1820. Continuous dependence on the data as an equally important requirement was emphasized about a century later by the French mathematician Jacques Hadamard (1865–1963); cf. Hadamard (1923, Chap. II). In Courant and Hilbert's classic *Methoden der Mathematischen Physik* (Vol. II, 1937, p. 176) a "*sachgemäßes Problem*" is described, perhaps for the first time, in much the same way as in (a)–(c) above. In the English edition (1962) this is translated as a "properly posed problem."

II. Differential Equations for Complex-Valued Functions of a Real Variable. We first extend the notion of an initial value problem by allowing the functions that appear to be complex-valued. The independent variable x, however, remains a real variable as before. Since $\mathbb{C}$ and $\mathbb{R}^2$ (and likewise $\mathbb{C}^n$ and $\mathbb{R}^{2n}$) are equivalent as sets, as metric spaces, and with respect to the additive structure, we can represent the complex-valued function $y : J \to \mathbb{C}$ as a pair of real functions

$$y(x) = (u(x), v(x)) = u(x) + iv(x) \quad \text{with} \quad u = \operatorname{Re} y, \quad v = \operatorname{Im} y.$$

Likewise, we write $\mathbf{y} = (\mathbf{u}, \mathbf{v})$ for $\mathbf{y} \in \mathbb{C}^n$ $(\mathbf{u}, \mathbf{v} \in \mathbb{R}^n)$. Then $\mathbf{f}(x, \mathbf{y}) : J \times \mathbb{C}^n \to \mathbb{C}^n$ can be written in the form

$$\mathbf{f}(x, \mathbf{y}) = (\mathbf{g}(x, \mathbf{u}, \mathbf{v}), \mathbf{h}(x, \mathbf{u}, \mathbf{v})),$$

and hence the "real–complex" system of n differential equations

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}) \tag{1}$$

is equivalent to a real system of $2n$ differential equations

$$\begin{aligned} \mathbf{u}' &= \mathbf{g}(x, \mathbf{u}, \mathbf{v}), \\ \mathbf{v}' &= \mathbf{h}(x, \mathbf{u}, \mathbf{v}). \end{aligned} \tag{2}$$

Further, continuity, respectively Lipschitz continuity, with respect to $\mathbf{y}$ for $\mathbf{f}$ is equivalent to continuity, respectively Lipschitz continuity, with respect to $(\mathbf{u}, \mathbf{v})$ for $\mathbf{g}$ and $\mathbf{h}$.

It follows that the earlier theorems for real systems remain valid for systems with complex-valued functions. This statement can also be verified directly, since the earlier proofs remain valid without changes if $C(J)$ is understood to be a Banach space of complex-valued functions.

Example. In the differential equation

$$y' = \lambda y + g(x)$$

let $\lambda = \mu + i\nu$ and $g(x) = h(x) + ik(x) \in C(\mathbb{R})$. The equivalent real system is

$$\begin{aligned} u' &= \mu u - \nu v + h(x), \\ v' &= \nu u + \mu v + k(x). \end{aligned}$$

As in the real case, the general solution is

$$y(x; C) = C\mathrm{e}^{\lambda x} + \int_0^x \mathrm{e}^{-\lambda(x-t)} g(t)\, dt \quad (C \in \mathbb{C}!).$$

The proof is left as an exercise for the reader.

It is frequently more convenient to work with (1) instead of (2) from a practical point of view (cf. the above example). There is also an important theoretical reason for preferring (1). In the example given above, the right-hand side of the differential equation is a *holomorphic function* of the parameter $\lambda \in \mathbb{C}$. This property arises frequently, and an important theorem says that if the right-hand side is holomorphic in λ, then the solutions are also holomorphic in λ (this is evident in the example). This theorem is proved in the next section. It is needed in a later chapter in the investigation of eigenvalue problems, among others.

Notice. The theorems in §12 are true for systems where the right-hand sides and solutions or approximate solutions are real-valued, as well as for those where they are complex-valued; in both cases, however, the independent variable x is always real.

The following estimation theorem, Theorem III, deals with the initial value problem

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}) \quad \text{in } J, \quad \mathbf{y}(\xi) = \boldsymbol{\eta}. \tag{3}$$

It gives an estimate for the difference $\mathbf{z}(x) - \mathbf{y}(x)$, where $\mathbf{y}(x)$ is a solution to (3) and $\mathbf{z}(x)$ is an "approximate solution." The two quantities

$$\mathbf{z}(\xi) - \boldsymbol{\eta} \quad \text{and} \quad P\mathbf{z} = \mathbf{z}' - \mathbf{f}(x, z) \quad (\textit{defect})$$

are used to measure of how "good" $\mathbf{z}(x)$ is as an approximation to $\mathbf{y}(x)$. Theorem III establishes a bound $\rho(x)$ for the difference $|\mathbf{z}(x) - \mathbf{y}(x)|$ that depends on a bound on the initial deviation (a), and a bound on the defect (b), and, most important, a condition (d) on $\mathbf{f}$ that includes the Lipschitz condition as a special case.

III. Estimation Theorem. *Let the vector functions $\mathbf{y}(x)$, $\mathbf{z}(x)$ and the real-valued function $\rho(x)$ be defined and differentiable in the interval $J : \xi \le x \le \xi + a$. Let the real-valued functions $\delta(x)$ and $\omega(x, z)$ be defined in J and $J \times \mathbb{R}$, resp., and suppose the following conditions are satisfied:*

(a) $|\mathbf{z}(\xi) - \mathbf{y}(\xi)| < \rho(\xi)$,

(b) $\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$, $\quad |\mathbf{z} - \mathbf{f}(x, \mathbf{z})| \le \delta(x)$ *in* J,

(c) $\rho' > \delta(x) + \omega(x, \rho(x))$ *in* J, *and*

(d) $|\mathbf{f}(x, \mathbf{y}) - \mathbf{f}(x, \mathbf{z})| \le \omega(x, |\mathbf{y} - \mathbf{z}|)$ *in* J.

Then

$$|\mathbf{z}(x) - \mathbf{y}(x)| < \rho(x) \quad \textit{in} \quad J.$$

If $\omega(x, y)$ *is continuous and locally Lipschitz continuous in* y, *then the theorem holds with* $\le$ *in all places. Naturally, we assume that* $\mathbf{f}$ *is defined in a set* D *containing* graph $\mathbf{y}$ *and* graph $\mathbf{z}$.

In the proof, we need the following Lemma.

IV. Lemma. *If the vector function* $\mathbf{g}(x)$ *is differentiable at* x_0, *then the scalar function* $\phi(x) = |\mathbf{g}(x)|$ *satisfies*

$$D^-\phi(x_0) \le |\mathbf{g}'(x_0)| \quad \textit{and} \quad D^+\phi(x_0) \le |\mathbf{g}'(x_0)|.$$

As a matter of fact, the one-sided derivatives ϕ'_+ and ϕ'_- exist at x_0; cf. B.IV.

The *proof* proceeds by passing to the limit as $h \to 0+$ in the inequality (here $h > 0$),

$$\frac{\phi(x_0) - \phi(x_0 - h)}{h} = \frac{|\mathbf{g}(x_0)| - |\mathbf{g}(x_0 - h)|}{h} \le \left| \frac{\mathbf{g}(x_0) - \mathbf{g}(x_0 - h)}{h} \right|,$$

and similarly for the second inequality. ∎

Proof of the Estimation Theorem III. We apply Theorem 9.III with $\phi(x) = |\mathbf{z}(x) - \mathbf{y}(x)|$, $\psi(x) = \rho(x)$, and ω instead of f. By hypothesis (a), the requirement $\phi(\xi) < \psi(\xi)$ is satisfied. Since $P\rho = \rho' - \omega(x, \rho) > \delta(x)$ by (c), it remains to prove that $P\phi \le \delta(x)$. Indeed, it follows from the lemma and assumptions (b), (d) that

$$\begin{aligned} D^-\phi(x) &\le |\mathbf{z}'(x) - \mathbf{y}'(x)| \\ &= |\mathbf{z}' - \mathbf{f}(x, \mathbf{z}) + \mathbf{f}(x, \mathbf{z}) - \mathbf{f}(x, \mathbf{y})| \\ &\le \delta(x) + \omega(x, |\mathbf{z} - \mathbf{y}|) = \delta(x) + \omega(x, \rho(x)). \end{aligned}$$

In the theorem with $\le$ we apply Theorem 9.VIII. ∎

The most important special case is the following

V. Lipschitz Condition. Theorem. *If* $\mathbf{f}$ *satisfies a Lipschitz condition in* D,

$$|\mathbf{f}(x, \mathbf{y}_1) - \mathbf{f}(x, \mathbf{y}_2)| \le L|\mathbf{y}_1 - \mathbf{y}_2|, \tag{4}$$

and if $\mathbf{y}(x)$ *is a solution and* $\mathbf{z}(x)$ *an approximate solution to the initial value problem* (3) *in* J *such that*

$$|\mathbf{z}(\xi) - \mathbf{y}(\xi)| \le \gamma, \quad |\mathbf{z}'(x) - \mathbf{f}(x, \mathbf{z})| \le \delta \tag{5}$$

(γ, δ are constants), then the estimate

$$|\mathbf{y}(x) - \mathbf{z}(x)| \le \gamma e^{L|x-\xi|} + \frac{\delta}{L}(e^{L|x-\xi|} - 1) \tag{6}$$

holds in J. *Here* J *is an arbitrary interval with* $\xi \in J$.

In the estimation theorem, we set $\omega(x, \mathbf{z}) = L\mathbf{z}$ and use the second version. All four assumptions are satisfied if $\rho(x)$ is the solution of

$$\rho' = \delta + L\rho \quad \text{in} \quad J, \quad \rho(\xi) = \gamma.$$

This leads to the bound in (6) for $x > \xi$. The case $x < \xi$ can be reduced to the estimation theorem in exactly the same way by a reflection about the point ξ. ∎

If $\delta = \gamma = 0$ in (5), then the estimate (6) implies that $\mathbf{y}(x) = \mathbf{z}(x)$. Thus (6) contains the uniqueness result proved earlier in the case where the right-hand side satisfies a Lipschitz condition. However, it also includes significantly more, namely a

VI. Theorem on Continuous Dependence. *Let* J *be a compact interval with* $\xi \in J$ *and let the function* $\mathbf{y} = \mathbf{y}_0(x)$ *be a solution of the initial value problem*

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}) \quad \text{in} \quad J, \quad \mathbf{y}(\xi) = \boldsymbol{\eta}. \tag{3}$$

The α*-neighborhood* ($\alpha > 0$) *of* graph $\mathbf{y}_0$ *(definition: the set of all points* $(x, \mathbf{y})$ *with* $x \in J$, $|\mathbf{y} - \mathbf{y}_0(x)| \le \alpha$*) will be denoted by* S_α. *Suppose there exists* $\alpha > 0$ *such that* $\mathbf{f}(x, \mathbf{y})$ *is continuous and satisfies the Lipschitz condition* (4) *in* S_α.

Then the solution $\mathbf{y}_0(x)$ *depends continuously on the initial values and on the right-hand side* $\mathbf{f}$. *In other words: For every* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that if* $\mathbf{g}$ *is continuous in* S_α *and the inequalities*

$$|\mathbf{g}(x, \mathbf{y}) - \mathbf{f}(x, \mathbf{y})| < \delta \quad \text{in} \quad S_\alpha, \quad |\boldsymbol{\zeta} - \boldsymbol{\eta}| < \delta \tag{7}$$

are satisfied, then every solution $\mathbf{z}(x)$ *of the "perturbed" initial value problem*

$$\mathbf{z}' = \mathbf{g}(x, \mathbf{z}), \quad \mathbf{z}(\xi) = \boldsymbol{\zeta} \tag{8}$$

exists in all of J *and satisfies the inequality*

$$|\mathbf{z}(x) - \mathbf{y}_0(x)| < \varepsilon \quad \text{in} \quad J. \tag{9}$$

Proof. Let $\mathbf{z}(x)$ satisfy (7) and (8). As long as the curve $\mathbf{z}(x)$ remains in S_α, (5) is satisfied for $\mathbf{y}_0(x)$ and $\mathbf{z}(x)$ with $\gamma = \delta$. Thus (6) holds with $\gamma = \delta$. If $\gamma = \delta$ is chosen sufficiently small in (6), then it is easy to see that the right side of (6) is $\leq \alpha/2$. As long as $\mathbf{z}(x)$ remains in S_α, i.e., as long as $|\mathbf{y}_0(x) - \mathbf{z}(x)| < \alpha$, the estimate (6) holds and hence, as a matter of fact, $|\mathbf{y}_0(x) - \mathbf{z}(x)| \leq \alpha/2$. From here one sees immediately that the curve $\mathbf{z}(x)$ cannot leave the neighborhood S_α. Thus the estimate (6) with $\gamma = \delta$ holds in all of J. Therefore condition (9) is easily satisfied (for an arbitrarily given $\varepsilon > 0$); one has simply to take $\gamma = \delta$ so small that the right-hand side of (6) is $< \varepsilon$. ∎

Remarks. 1. The theorem applies, in particular, if D is open and $\mathbf{f}$ and $\partial \mathbf{f}/\partial \mathbf{y}$ are in $C^0(D)$. Indeed, if $\mathbf{y}_0$ is a solution on a compact interval J, then there exists an $\alpha > 0$ such that $S_\alpha \subset D$, and furthermore, a Lipschitz condition holds because of the continuity of the derivatives of $\mathbf{f}$.

2. The theorem applies to the case where $|P\mathbf{z}| = |\mathbf{z}' - \mathbf{f}(x, \mathbf{z})| < \delta$ because $\mathbf{z}$ is a solution of (8) with $\mathbf{g}(x, \mathbf{y}) := \mathbf{f}(x, \mathbf{y}) + (P\mathbf{z})(x)$.

Supplement: General Uniqueness and Dependence Theorems

The Lipschitz condition in Theorem VI can be replaced with a significantly weaker uniqueness condition.

VII. Dependence and Uniqueness Theorem. *Let the real-valued function $\omega(x, z)$ be defined for $x \in J := [\xi, \xi + a]$ $(a > 0)$, $z \geq 0$, and have the property:*

(U) *For every $\varepsilon > 0$, there exist $\delta > 0$ and a function $\rho(x)$ such that*

$$\rho' > \delta + \omega(x, \rho) \quad \text{and} \quad 0 < \delta < \rho(x) < \varepsilon \quad \text{in} \quad J.$$

If $\mathbf{f}$ *satisfies the estimate*

$$|\mathbf{f}(x, \mathbf{y}_1) - \mathbf{f}(x, \mathbf{y}_2)| \leq \omega(x, |\mathbf{y}_1 - \mathbf{y}_2|) \tag{10}$$

in $D \subset J \times \mathbb{R}^n$, respectively $J \times \mathbb{C}^n$, then the initial value problem (3) *has at most one solution. The solution depends continuously on $\boldsymbol{\eta}$ and $\mathbf{f}$ in the sense described in Theorem* VI.

Proof. If $\mathbf{y}$ is a solution of (3) and $\varepsilon > 0$ is given, we determine $\rho(x)$ and δ according to (U). If $\mathbf{z}$ satisfies (7), (8), then it follows immediately from Theorem III that $|\mathbf{y}(x) - \mathbf{z}(x)| < \rho(x) < \varepsilon$. ∎

VIII. Examples of Well-Posedness. The theorem states in particular that if $\mathbf{f}$ is continuous and the function ω satisfies (U), then an estimate of the form (10) gives a condition for well-posedness of problem (3). Some examples of functions ω that satisfy (U) are:

(a) *The Lipschitz Condition* (R. Lipschitz 1876): $\omega(x, z) = Lz$.

(b) *Osgood's Condition* (1898): $\omega(x,z) = q(z)$, where $q \in C[0,\infty)$, $q(0) = 0$, $q(z) > 0$ for $z > 0$, and

$$\int_0^1 \frac{dz}{q(z)} = \infty.$$

(c) *Bompiani's Condition* (1925): Let the function $\omega(x,z)$ be continuous and ≥ 0 for $x \in J$, $z \geq 0$. Let $\omega(x,0) = 0$ and suppose the following condition is satisfied:

If $\phi(x) \geq 0$ is a solution of the initial value problem

$$\phi' = \omega(x,\phi) \quad \text{in} \quad J_1 := [\xi, \xi + \alpha), \quad \phi(\xi) = 0,$$

then $\phi = 0$ in J_1.

(d) *Krasnosel'skii–Krein Condition* (1956):

$$\omega(x,z) = \min\left(Cz^\alpha, \frac{kz}{x-\xi}\right) \quad \text{for} \quad x > \xi$$

with $0 < \alpha < 1$, $0 < k(1-\alpha) < 1$, $C > 0$.

Example (a) is clearly a special case of (b). By Theorem 1.VIII, (b) is a special case of (c). To show that a function ω that satisfies (c) also satisfies (U), we modify ω for $z \geq 1$ by setting $\omega(x,z) = \omega(x,1)$ for $z \geq 1$. Then ω is bounded. Let ρ_n be a solution of the initial value problem

$$\rho_n' = \omega(x,\rho_n) + \frac{1}{n} \quad \text{in} \quad J, \quad \rho_n(0) = \frac{1}{n}.$$

Since ω is bounded, ρ_n exists in all J. By Theorem 9.III, the sequence (ρ_n) is monotone decreasing. Therefore $\phi(x) = \lim \rho_n(x)$ exists. Furthermore, the sequence (ρ_n) is equicontinuous, and hence the convergence is uniform (follows from the boundedness of ω and 7.III). Representing the initial value problem for ρ_n as an integral equation and taking the limit as $n \to \infty$ gives

$$\phi(x) = \int_\xi^x \omega(t,\phi(t))\,dt.$$

From (c) it follows that $\phi = 0$ in every interval $J_1 = [\xi, \xi + \alpha]$ in which $\phi \leq 1$ (note that ω was modified, but only for $z \geq 1$). It follows easily that $\phi = 0$ in J. Thus (ρ_n) converges uniformly to 0 in J. Therefore, for every $\varepsilon > 0$, there exists $\delta = 1/(2n)$ and $\rho = \rho_n$ such that (U) is satisfied.

Example (d) is also a special case of (c). However, the verification is somewhat more difficult; cf. Walter (1970; p. 108).

Remark. More general uniqueness conditions of Nagumo, Kamke, and others can be found in the literature. Their importance, however, is limited by the fact, proved first by Olech (1960), that a *continuous* function $\mathbf{f}$ that satisfies such a general condition also satisfies the condition in Theorem VII. References to the literature and historical remarks are contained in Walter (1970; see, in particular, §14).

§ 13. Dependence of Solutions on Initial Values and Parameters

In this section the problem of dependence of the solution of an initial value problem on the data is investigated further. At the same time, the structure of the problem is generalized in two directions. First, we consider the case where the right side of the differential equation depends on a parameter λ; that is, $\mathbf{f} = \mathbf{f}(x, \mathbf{y}; \lambda)$. Second, we consider

I. Volterra Integral Equations of the form

$$y(x) = g(x) + \int_{\xi}^{x} k(x, t, y(t))\, dt. \tag{1}$$

The integral equation (6.3) for an initial value problem is a special case. There $g(x) = \eta$ is constant and the "kernel" $k(x, t, z)$ is independent of x.

Example. In §11 we showed how to transform an initial value problem

$$y'' = f(x, y) \quad \text{with} \quad y(0) = \eta_0, \quad y'(0) = \eta_1$$

into a system of two differential equations of first order. This problem can also be written as an equivalent Volterra integral equation,

$$y(x) = \eta_0 + \eta_1 x + \int_0^x (x - t) f(t, y(t))\, dt \tag{2}$$

(the proof is an exercise). For the investigation of some types of problems equation (2) is more useful than the first order system.

More generally, we consider a vector integral equation of Volterra type depending on a parameter λ,

$$\mathbf{y}(x; \lambda) = \mathbf{g}(x; \lambda) + \int_{\alpha(\lambda)}^{x} \mathbf{k}(x, t, \mathbf{y}(t; \lambda); \lambda)\, dt \tag{3}$$

in an interval J. Here x and t are always real. However, $\mathbf{g}$, $\mathbf{k}$, and $\mathbf{y}$ may be complex-valued. More than one real or complex parameter is permitted; i.e., $\lambda \in \mathbb{R}^m$ or $\mathbb{C}^m$. Our objective is to study the dependence of $\mathbf{y}(x; \lambda)$ on λ. It is not assumed that $\alpha(\lambda) < x$.

II. Theorem on Continuous Dependence. *Let $J = [a, b]$ and let $K \subset \mathbb{R}^m$ be compact. Let the functions $\mathbf{g} : J \times K \to \mathbb{R}^n$, $\alpha : K \to J$ and $\mathbf{k} : J^2 \times \mathbb{R}^n \times K \to \mathbb{R}^n$ be continuous in their respective domains, and let $\mathbf{k}$ satisfy a Lipschitz condition*

$$|\mathbf{k}(x, t, \mathbf{u}; \lambda) - \mathbf{k}(x, t, \mathbf{v}; \lambda)| \le L|\mathbf{u} - \mathbf{v}|. \tag{4}$$

Then the integral equation (3) has exactly one solution $\mathbf{y}(x; \lambda)$ for every $\lambda \in K$. The solution is continuous as a function of $(x; \lambda)$, that is, $\mathbf{y}(x; \lambda) \in C(J \times K)$.

The theorem is also valid with $\mathbb{R}$ replaced by $\mathbb{C}$ (the complex case).

The *proof* uses an argument similar to the one in 6.I. The space $B = C(J \times K)$ with the norm

$$\|\mathbf{u}\| := \sup\{|\mathbf{u}(x;\lambda)|\mathrm{e}^{-2L|x-\alpha(\lambda)|} : (x;\lambda) \in J \times K\}$$

is a Banach space. For $\mathbf{u} \in B$, we define

$$(T\mathbf{u})(x;\lambda) = \mathbf{g}(x;\lambda) + \int_{\alpha(\lambda)}^{x} \mathbf{k}(x,t,\mathbf{u}(t;\lambda);\lambda)\,dt. \tag{5}$$

Clearly, $T\mathbf{u} \in B$ if $\mathbf{u} \in B$. The Lipschitz condition (4) implies that

$$|(T\mathbf{u} - T\mathbf{v})| \le L\left|\int_{\alpha(\lambda)}^{x} |\mathbf{u}(t;\lambda) - \mathbf{v}(t;\lambda)|\,dt\right|.$$

Since $|\mathbf{u}(t,\lambda)| \le \|\mathbf{u}\| \cdot \mathrm{e}^{2L|t-\alpha(\lambda)|}$, the right-hand side of the inequality is less than or equal to

$$L\|\mathbf{u}-\mathbf{v}\|\left|\int_{\alpha(\lambda)}^{x} \mathrm{e}^{2L|t-\alpha(\lambda)|}dt\right| \le \frac{1}{2}\|\mathbf{u}-\mathbf{v}\|\mathrm{e}^{2L|x-\alpha(\lambda)|},$$

and hence,

$$\|T\mathbf{u} - T\mathbf{v}\| \le \frac{1}{2}\|\mathbf{u}-\mathbf{v}\|. \tag{6}$$

The conclusion follows now from the fixed point theorem 5.IX as does the following corollary. ■

Corollary. *For every $\mathbf{u}_0 \in C(J \times K)$, the sequence of successive approximations $(\mathbf{u}_k)$ with $\mathbf{u}_{k+1} = T\mathbf{u}_k$ $(k = 0, 1, 2, \ldots)$ converges uniformly on $J \times K$ to the solution $\mathbf{y}$.*

III. Theorem on Analyticity in $\boldsymbol{\lambda}$. *Let the assumptions from* II *(complex case) hold, and let K° be the interior of $K \subset \mathbb{C}$. If $\alpha(\lambda)$ is constant, $\mathbf{g}(x;\lambda)$ is holomorphic with respect to λ in K° for fixed $x \in J$, and $\mathbf{k}(x,t,\mathbf{y};\lambda)$ is holomorphic with respect to $(\mathbf{y},\lambda)$ in $\mathbb{C}^n \times K^\circ$ for fixed $(t,x) \in J^2$, then the solution $\mathbf{y}$ is holomorphic with respect to $\lambda \in K^\circ$ for fixed $x \in J$.*

This result follows from the corollary in II and the following fact: If $\mathbf{u} \in C(J \times K)$ is holomorphic with respect to $\lambda \in K^\circ$, then the same holds for $T\mathbf{u}$. Thus, if the successive approximations are started with a function $\mathbf{u}_0$ that is holomorphic with respect to λ, then the whole sequence and, by uniform convergence, its limit are holomorphic with respect to λ. ■

Extension. Clearly, a corresponding theorem holds if both real and complex parameters appear in the right-hand side. Let $\lambda = (\lambda', \lambda'')$ with $\lambda' \in \mathbb{R}^p$, $\lambda'' \in \mathbb{C}^q$. If $\alpha(\lambda) = \alpha(\lambda')$ and if $\mathbf{g}$ and $\mathbf{k}$ are holomorphic in λ'' and $(\mathbf{y}, \lambda'')$, respectively, then the same holds for the solution.

Theorem III implies that the solution $\mathbf{y}(x;\lambda)$ is differentiable with respect to complex parameters. The corresponding proposition with respect to real parameters is more difficult to prove. The following proof can be omitted at first reading. It depends on a theorem about approximate iteration in a Banach space, which is an extension of the contraction principle 5.IX. It was proved in a special form (with $\beta_k = 0$) by Ostrowski (1967).

IV. Ostrowski's Theorem on Approximate Iteration. *Let D be a closed subset of a Banach space B and let $R : D \to D$ be a contraction,*

$$\|Ru - Rv\| \le q\|u - v\| \quad \text{in} \quad D \quad \text{with } q < 1.$$

Suppose the sequence (v_k) in D satisfies

$$v_{k+1} = Rv_k + a_k \quad \textit{where,} \quad \|a_k\| \le \alpha_k + \beta_k\|v_k\|$$

and (α_k), (β_k) are real null sequences. Then the sequence (v_k) converges to the fixed point z of R: $\lim v_k = z$ with $z = Rz$ (the existence of the fixed point is guaranteed by the contraction principle 5.IX*).*

Proof. Because $z = Rz$, we have

$$\|v_{k+1} - z\| \le \|Rv_k - Rz\| + \|a_k\| \le q\|v_k - z\| + \|a_k\|.$$

Using the hypothesis on a_k and the inequality $\|v_k\| \le \|v_k - z\| + \|z\|$, we have

$$\|a_k\| \le \alpha_k + \beta_k\|v_k - z\| + \beta_k\|z\|.$$

Hence the term $\varepsilon_k = \|v_k - z\|$ satisfies

$$\varepsilon_{k+1} \le (q + \beta_k)\varepsilon_k + \gamma_k \quad \text{with} \quad \gamma_k = \alpha_k + \beta_k\|z\|,$$

where the sequence(γ_k) is a null sequence.

From here it easily follows that $\varepsilon_k \to 0$. Let $\varepsilon > 0$ and $r \in (q, 1)$ be given. Then there exists an index p such that $q + \beta_k \le r$ and $\gamma_k \le \varepsilon(1 - r)$ for $k \ge p$, which implies $\varepsilon_{k+1} \le r\varepsilon_k + \varepsilon(1 - r)$ for $k \ge p$. In terms of $\delta_k := \varepsilon_k - \varepsilon$ this reads

$$\delta_{k+1} \le r\delta_k \quad \text{for} \quad k \ge p.$$

If $\delta_k \ge 0$ for all $k \ge p$, then obviously $\delta_k \to 0$ monotonically; if one member δ_n is negative, then $\delta_k < 0$ for all $k > n$. Hence $\limsup \delta_k \le 0$, which implies $\limsup \varepsilon_k \le \varepsilon$; i.e., $\lim \varepsilon_k = 0$, since $\varepsilon_k \ge 0$. ∎

V. Differentiability with Respect to Real Parameters. First we recall the chain rule for vector functions. If $\mathbf{u} : \mathbb{R} \to \mathbb{R}^n$ and $\mathbf{v}(t) := \mathbf{f}(\mathbf{u}(t))$ with $\mathbf{f} \in C^1$, then

$$v_i'(t) = \sum_{j=1}^{n} \frac{\partial f_i}{\partial y_j}(\mathbf{u}(t)) \cdot u_j'(t) \quad (i = 1, \ldots, n).$$

In matrix notation this formula is written

$$\mathbf{v}'(t) = \mathbf{f}_{\mathbf{y}}(\mathbf{u}(t))\mathbf{u}'(t),$$

where $\mathbf{f}_{\mathbf{y}} = \partial\mathbf{f}/\partial\mathbf{y}$ is the $n \times n$ Jacobi matrix of $\mathbf{f}$, and $\mathbf{u}'$, $\mathbf{v}'$ are column vectors. The same formula holds for complex-valued functions ($\mathbb{C}$ instead of $\mathbb{R}$). In matrix products, vectors are always interpreted as column vectors (for example $\mathbf{g}$, $\mathbf{k}$, $\mathbf{y}$).

We show that the solution is differentiable with respect to a real parameter and that (3) can be formally differentiated, provided that all the derivatives of $\mathbf{g}$ and $\mathbf{k}$ that appear exist and are continuous. Let T be given by (5), and let $\lambda = (\lambda', \lambda'')$ with $\lambda' \in \mathbb{R}$; the remaining parameters (real or complex) are lumped together in λ''. Differentiating formally, we obtain

$$\frac{\partial}{\partial\lambda'}(T\mathbf{u}) = S(\mathbf{u}, \mathbf{u}_{\lambda'}), \tag{7}$$

where

$$\begin{aligned} S(\mathbf{u},\mathbf{v})(x;\lambda) \quad &= \quad \mathbf{g}_{\lambda'}(x;\lambda) - \alpha_{\lambda'}(\lambda)\mathbf{k}(x,\alpha(\lambda),\mathbf{u}(\alpha(\lambda);\lambda);\lambda) \\ &\quad + \int_{\alpha(\lambda)}^{x} \big[\mathbf{k}_{\lambda'}(x,t,\mathbf{u}(t;\lambda);\lambda) \\ &\quad + \mathbf{k}_{\mathbf{y}}(x,t,\mathbf{u}(t;\lambda);\lambda)\mathbf{v}(t;\lambda)\big]\, dt. \end{aligned} \tag{8}$$

VI. Theorem. *Let the assumptions from* II *hold; let the derivatives* $\alpha_{\lambda'}$, $\mathbf{g}_{\lambda'}$, $\mathbf{k}_{\lambda'}$ $\mathbf{k}_{\mathbf{y}}$ *be defined and continuous in* K°, *respectively* $J \times K^\circ$, *respectively* $J^2 \times \mathbb{R}^n \times K^\circ$ *(in the complex case* $J^2 \times \mathbb{C}^n \times K^\circ$*), and let* $\mathbf{y}(x;\lambda)$ *be the solution of* (3). *Then the derivative* $\mathbf{y}_{\lambda'}$ *exists and is continuous in* $J \times K^\circ$, *and*

$$\mathbf{y}_{\lambda'} = S(\mathbf{y}, \mathbf{y}_{\lambda'}). \tag{9}$$

Proof. Let C^* be the set of all $\mathbf{u} \in C(J \times K)$ such that $\mathbf{u}_{\lambda'}$ exists and is continuous in $J \times K^\circ$. Let the operators T, S be defined by (5), (8), respectively. Clearly, if $\mathbf{u} \in C^*$, then $T\mathbf{u} \in C^*$ and (7) holds.

Let $\mathbf{u}_0 \in C^*, \mathbf{v}_0 = \partial\mathbf{u}_0/\partial\lambda'$ and define sequences $(\mathbf{u}_k)$, $(\mathbf{v}_k)$ iteratively as follows

$$\mathbf{u}_{k+1} = T\mathbf{u}_k, \quad \mathbf{v}_{k+1} = S(\mathbf{u}_k, \mathbf{v}_k)\ (k \ge 0).$$

Thus $\mathbf{u}_1 = T\mathbf{u}_0$, and using (7),

$$\mathbf{v}_1 = S(\mathbf{u}_0, \mathbf{v}_0) = S(\mathbf{u}_0, \partial\mathbf{u}_0/\partial\lambda') = \partial\mathbf{u}_1/\partial\lambda'.$$

Proceeding inductively in this manner, one shows that in general,

$$\mathbf{u}_k \in C^* \quad \text{and} \quad \mathbf{v}_k = \partial\mathbf{u}_k/\partial\lambda' \quad (k \ge 0).$$

Now choose a compact subset $K_1 \subset K^\circ$ and consider S and T as operators on the Banach space $B = C(J \times K_1)$ with the norm defined in II:

$$\|\mathbf{u}\| := \sup\{|\mathbf{u}(x;\lambda)|\mathrm{e}^{-2L|x-\alpha(\lambda)|} \mid (x;\lambda) \in J \times K_1\}.$$

From Corollary II it follows that $\lim \mathbf{u}_k = \mathbf{y}$, where $\mathbf{y}$ is the solution to (3). We now apply Theorem IV to the operator $R = S(\mathbf{y}, \cdot)$ in $D = B$:

$$R\mathbf{v} = S(\mathbf{y}, \mathbf{v}) = \int_{\alpha(\lambda)}^{x} \mathbf{k}_{\mathbf{y}}(\ldots, \mathbf{y})\mathbf{v}(t;\lambda)\,dt + \text{ terms without } \mathbf{v}.$$

The kernel of this integral,

$$\mathbf{k}^*(x,t,\mathbf{v};\lambda) = \mathbf{k}_{\mathbf{y}}(x,t,\mathbf{y}(t;\lambda);\lambda)\mathbf{v},$$

is linear in $\mathbf{v}$. Thus we have

$$\mathbf{k}^*(x,t,\mathbf{v},\lambda) = \lim_{h\to 0} \frac{1}{h}\left\{\mathbf{k}(x,t,\mathbf{y}(t;\lambda)+h\mathbf{v};\lambda) - \mathbf{k}(x,t,\mathbf{y}(t;\lambda);\lambda)\right\}.$$

Using the Lipschitz condition (4), we see that the difference in braces on the right hand side is $\le L|h\mathbf{v}|$ in absolute value. From here it follows immediately that $|\mathbf{k}^*(x,t,\mathbf{v};\lambda)| \le L|\mathbf{v}|$, i.e., that $\mathbf{k}^*$ satisfies the Lipschitz condition (4) with respect to $\mathbf{v}$. Therefore, exactly as in the proof of II, R satisfies the Lipschitz condition in Theorem IV with $q = \frac{1}{2}$. The fixed point $\mathbf{z}$ of R is a function $\mathbf{z} \in C(J \times K_1)$ that satisfies

$$\mathbf{z} = S(\mathbf{y}, \mathbf{z}). \tag{10}$$

The sequence $(\mathbf{v}_k)$ satisfies

$$\mathbf{v}_{k+1} = S(\mathbf{u}_k, \mathbf{v}_k) = R\mathbf{v}_k + \mathbf{a}_k \quad \text{with} \quad \mathbf{a}_k = S(\mathbf{u}_k, \mathbf{v}_k) - S(\mathbf{y}, \mathbf{v}_k).$$

The term $\mathbf{a}_k$ consists of four differences corresponding to the four summands of S (see (8)). Since $\lim \mathbf{u}_k = \mathbf{v}_k$ (uniformly), the sum of the first three differences, which do not contain $\mathbf{v}_k$, is bounded by a term α_k with $\lim \alpha_k = 0$. The fourth difference is bounded by a term of the form $\beta_k\|\mathbf{v}_k\|$ with $\lim \beta_k = 0$. Thus Theorem IV can be applied, and we obtain

$$\mathbf{u}_k \to \mathbf{y} \quad \text{and} \quad \mathbf{v}_k = \partial\mathbf{u}_k/\partial\lambda' \to \mathbf{z} \quad \text{in} \quad B = C(J \times K_1).$$

It follows now from an elementary theorem in analysis (the statement and proof follow in VII) that $\mathbf{y}$ is differentiable with respect to λ' in $J \times K_1^\circ$ and $\mathbf{z} = \mathbf{y}_{\lambda'}$. Since K_1 was arbitrary, the proof of Theorem IV is complete. ∎

VII. Lemma. *Let $(\phi_n(t))$ be a sequence in $C^1[a,b]$ and let $\phi_n \to \phi$ and $\phi_n' \to \psi$ uniformly in $[a,b]$. Then $\phi \in C^1[a,b]$ and $\phi' = \psi$.*

Proof. Passing to the limit in the formula

$$\phi_n(t) = \phi_n(a) + \int_a^t \phi_n'(s)\,ds,$$

one obtains

$$\phi(t) = \phi(a) + \int_a^t \psi(s)\, ds$$

and hence the conclusion. ∎

VIII. Remarks. (a) If $G \subset \mathbb{R}^p$ is an open set and $u(x_1, \ldots, x_p)$ is a continuous function in $\bar{G}$, then the statement $u_{x_i} \in C(\bar{G})$ means that this derivative exists in G and has a continuous extension to $\bar{G}$. The class $C^k(\bar{G})$ is defined correspondingly. There is a corollary to Theorem VI that is true in this sense. If K is the closed hull of K° and if all of the assumptions of VI hold for K in place of K°, then $\mathbf{y}_{\lambda'} \in C(J \times K)$. In this case, (10) has a solution $\mathbf{z} \in C(J \times K)$.

(b) Statements about higher order derivatives follow immediately from Theorem VI. Since derivative $\mathbf{y}_{\lambda'}$ also satisfies an integral equation (9) of the type of equation (3) (as a matter of fact, it is a linear equation), one can again apply VI. Arguing along these lines, one obtains the following

Theorem. *If the partial derivatives of α and $\mathbf{g}$ with respect to $\lambda_1, \ldots, \lambda_m$ and the partial derivatives of $\mathbf{k}$ with respect to t; $y_1, \ldots, y_n$; $\lambda_1, \ldots, \lambda_m$ up to order p are continuous in their domains (given explicitly in Theorem VI), then the derivatives of $\mathbf{y}$ with respect to $\lambda_1, \ldots, \lambda_m$ up to order p are continuous in $J \times K^\circ$, and they can be obtained by differentiating with respect to λ_i under the integral sign in* (3).

IX. The Initial Value Problem. When an initial value problem

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(\xi) = \boldsymbol{\eta} \tag{11}$$

has exactly one (maximally extended) solution for a given $(\xi, \boldsymbol{\eta})$, then we denote this solution by $\mathbf{y}(x; \xi, \boldsymbol{\eta})$ in order to emphasize the dependence on the initial values.

Example. The general solution of the differential equation $(n = 1)$

$$y' = 2xy + 1 - 2x^2$$

is given by $y = x + Ce^{x^2}$. It follows that

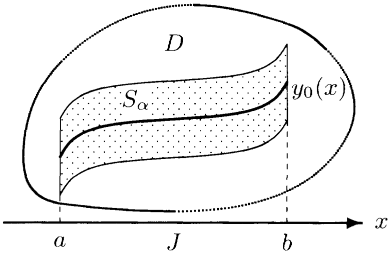
$$y(x; \xi, \eta) = x + (\eta - \xi)e^{x^2 - \xi^2}.$$

The theorems obtained above can be applied to (11), since the initial value problem is equivalent to the integral equation

$$\mathbf{y}(x; \xi, \boldsymbol{\eta}) = \boldsymbol{\eta} + \int_\xi^x \mathbf{f}(t, \mathbf{y}(t; \xi, \boldsymbol{\eta}))\, dt. \tag{12}$$

Here we have $\lambda = (\xi, \boldsymbol{\eta})$, $\mathbf{g}(x; \lambda) = \boldsymbol{\eta}$, $\alpha(\lambda) = \xi$.

However, since $\mathbf{f}$ is frequently not defined on all of the strip $J \times \mathbb{R}^n$, respectively $J \times \mathbb{C}^n$, we need a corresponding local theorem.



X. Theorem. *Let* $\mathbf{f}$ *and* $\partial\mathbf{f}/\partial\mathbf{y}$ *be defined and continuous in a domain* $D \subset \mathbb{R} \times \mathbb{R}^n$. *Let* $(\xi_0, \boldsymbol{\eta}_0) \in D$ *and* $\mathbf{y}_0(x) := \mathbf{y}(x; \xi_0, \boldsymbol{\eta}_0)$. *Suppose* $J = [a, b]$ *is a compact interval such that* $\mathbf{y}_0$ *exists on* J, *and denote by* S_α *the set* $S_\alpha = \{(x, y) : x \in J, |\mathbf{y} - \mathbf{y}_0(x)| \leq \alpha\}$.

Then there exists an $\alpha > 0$ *with* $S_\alpha \subset D$ *such that the function* $\mathbf{y}(x; \xi, \boldsymbol{\eta})$ *is defined in* $J \times S_\alpha$ *(i.e., every solution of the initial value problem with* $(\xi, \boldsymbol{\eta}) \in S_\alpha$ *exists at least in* J*), the functions* $\mathbf{y}$, $\mathbf{y}_\xi$, $\mathbf{y}_{\boldsymbol{\eta}}$ *and their derivatives with respect to* x, *which are denoted by* $\mathbf{y}'$, $\mathbf{y}'_\xi$, $\mathbf{y}'_{\boldsymbol{\eta}}$, *are continuous in* $J \times S_\alpha$, *and*

$$\mathbf{y}_\xi(x; \xi, \boldsymbol{\eta}) = -\mathbf{f}(\xi, \boldsymbol{\eta}) + \int_\xi^x \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}(t; \xi, \boldsymbol{\eta}))\mathbf{y}_\xi(t; \xi, \boldsymbol{\eta})\, dt, \tag{13}$$

$$\mathbf{y}_{\boldsymbol{\eta}}(x; \xi, \boldsymbol{\eta}) = I + \int_\xi^x \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}(t; \xi, \boldsymbol{\eta})) \cdot \mathbf{y}_{\boldsymbol{\eta}}(t; \xi, \boldsymbol{\eta})\, dt, \tag{14}$$

and

$$\mathbf{y}_\xi(x; \xi, \boldsymbol{\eta}) + \mathbf{y}_{\boldsymbol{\eta}}(x; \xi, \boldsymbol{\eta}) \cdot \mathbf{f}(\xi, \boldsymbol{\eta}) = 0. \tag{15}$$

Remarks. 1. Notation: $\mathbf{y}$, $\mathbf{y}'$, $\mathbf{y}_\xi$, $\mathbf{f}$ are column vectors; $\mathbf{f}_{\mathbf{y}}$, $\mathbf{y}_{\boldsymbol{\eta}}$, $\mathbf{y}'_{\boldsymbol{\eta}}$ are $n \times n$ matrices (cf. the remark in V); I is the identity matrix; (14) is a linear matrix-integral equation (it is equivalent to n vector-integral equations for the columns $\mathbf{y}_{\eta_i}$); the product $\mathbf{f}_{\mathbf{y}} \cdot \mathbf{y}_{\boldsymbol{\eta}}$ is a matrix product.

2. The theorem remains valid in the complex case with $\mathbb{C}^n$ in place of $\mathbb{R}^n$.

We simplify the *proof* by extending $\mathbf{f}$ continuously and differentiably to the strip $J \times \mathbb{R}$ and then applying the earlier results (the proof is also valid in the complex case). Let $\beta > 0$ be chosen such that $S_{2\beta} \subset D$. We determine a real function $h(s) \in C^1(\mathbb{R})$ that satisfies

$$h(s) = \begin{cases} 1 & \text{for} \quad x \leq \beta, \\ 0 & \text{for} \quad x \geq 2\beta, \end{cases}$$

and $0 \leq h(x) \leq 1$, and we define

$$\mathbf{f}^*(x, \mathbf{y}) = \mathbf{f}(x, \mathbf{y}_0(x) + (\mathbf{y} - \mathbf{y}_0(x))h(|\mathbf{y} - \mathbf{y}_0(x)|)).$$

Clearly, the expression $\mathbf{y}_0(x) + (\mathbf{y} - \mathbf{y}_0(x))h(|\mathbf{y} - \mathbf{y}_0(x)|)$ that appears on the right-hand side lies in $S_{2\beta}$ for $(x, \mathbf{y}) \in J \times \mathbb{R}^n$. Moreover, $\mathbf{f} = \mathbf{f}^*$ in S_β, and $\mathbf{f}^*$ has continuous and bounded derivatives in $J \times \mathbb{R}^n$ with respect to y_i and hence satisfies a Lipschitz condition (4). By Theorem II, the system (11) with $\mathbf{f}^*$ on the right has a solution $\mathbf{y}^*(x; \xi, \boldsymbol{\eta})$ that exists and is continuous in $J \times J \times \mathbb{R}^n$. If the initial point $(\xi, \boldsymbol{\eta})$ lies on the curve $\mathbf{y}_0(x)$, then $\mathbf{y}^*(x; \xi, \boldsymbol{\eta}) = \mathbf{y}_0(x)$. Thus by the uniform continuity of $\mathbf{y}^*$ in all variables, there is an α, $0 < \alpha < \beta$, such that $\mathbf{y}^*$ remains in S_β for $(\xi, \boldsymbol{\eta}) \in S_\alpha$. For these values of $(\xi, \boldsymbol{\eta})$, the relation $\mathbf{y}(x; \xi, \boldsymbol{\eta}) = \mathbf{y}^*(x; \xi, \boldsymbol{\eta})$ holds. Thus by Theorem VI, (12) can be formally differentiated, and one obtains (13), (14), first with $\mathbf{f}^*$, which, however, can then be replaced by $\mathbf{f}$ for $(\xi, \boldsymbol{\eta}) \in S_\alpha$. Since in (13) and (14) the integrand is a continuous function, the derivatives $\mathbf{y}', \mathbf{y}'_\xi, \ldots$ exist and are continuous.

By (13), (14), the left side of (15), call it $\mathbf{v}$, satisfies a homogeneous linear integral equation

$$\mathbf{v}(x) = \int_\xi^x \mathbf{f}_\mathbf{y} \cdot \mathbf{v}(t)\, dt,$$

which has only the zero solution (Theorem II). This proves equation (15). ∎

XI. Higher Order Derivatives. Differentiation with Respect to Parameters. Now suppose the differential equation depends on a parameter,

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}, \lambda). \tag{16}$$

Instead of (12), one then has

$$\mathbf{y}(x; \xi, \boldsymbol{\eta}, \lambda) = \boldsymbol{\eta} + \int_\xi^x \mathbf{f}(t, \mathbf{y}(t; \xi, \boldsymbol{\eta}, \lambda), \lambda)\, dt. \tag{17}$$

Let $\mathbf{f}$ here be defined in $D \times K^\circ$, where D is a domain and K° is an open set in λ-space. Let $\lambda = (\lambda', \lambda'')$, where λ' is a real or complex parameter and λ'' is a vector parameter. The following theorem extends the results of Theorem X to the present case (S_α is defined as in X):

Theorem. *If $\mathbf{f}$, $\mathbf{f}_\mathbf{y}$, $\mathbf{f}_{\lambda'}$ are continuous in $D \times K^\circ$ and if there exists a solution $\mathbf{y}_0(x) = \mathbf{y}(x; \xi_0, \boldsymbol{\eta}_0, \lambda_0)$ in a compact interval J, where $(\xi_0, \boldsymbol{\eta}_0, \lambda_0) \in D \times K^\circ$, then there exists $\alpha > 0$ such that $\mathbf{y}(x; \xi, \boldsymbol{\eta}, \lambda)$ is defined in $J \times S_\alpha \times U_\alpha$, $U_\alpha = \{\lambda : |\lambda - \lambda_0| \le \alpha\}$. The statements in Theorem* X, *in particular* (13) – (15), *hold in $J \times S_\alpha \times U_\alpha$ (the variable λ needs to be added to the formulas). Further, $\mathbf{y}_{\lambda'}$ is continuous on this set, and*

$$\mathbf{y}_{\lambda'}(x; \xi, \boldsymbol{\eta}, \lambda) = \int_\xi^x \{\mathbf{f}_{\lambda'}(t, \mathbf{y}, \lambda) + \mathbf{f}_\mathbf{y}(t, \mathbf{y}, \lambda)\mathbf{y}_{\lambda'}(t; \xi, \boldsymbol{\eta}, \lambda)\}\, dt. \tag{18}$$

In the complex case, $\mathbf{f}$ is assumed to be holomorphic in $(\mathbf{y}, \lambda')$, and the solution is holomorphic in λ'.

Proof. Real Case. One chooses again $\beta > 0$ with $S_{2\beta} \subset D$ and extends $\mathbf{f}$ continuously and differentiably as in Theorem X. The extension $\mathbf{f}^*$ is defined in $J \times \mathbb{R}^n \times K^\circ$ and agrees with $\mathbf{f}$ in $S_\beta \times K^\circ$. Theorem VI is applied to the equivalent integral equation (17) with $\mathbf{f}^*$ in place of $\mathbf{f}$ (note that if $K_1 \subset K^\circ$ is compact, then since $\mathbf{f}^*_{\mathbf{y}}$ is bounded on $J \times \mathbb{R}^n \times K_1$, it follows that $\mathbf{f}^*$ satisfies a Lipschitz condition with respect to $\mathbf{y}$ in this set). All of the statements now follow for the extension $\mathbf{f}^*$. Further, there exists an α, $0 < \alpha < \beta$, such that if $(\xi, \boldsymbol{\eta}) \in S_\alpha$, $\lambda \in U_\alpha \subset K^\circ$, then the solution $\mathbf{y}$ remains in $S_{\beta/2}$, and therefore all the statements also hold for $\mathbf{f}$.

Complex Case. In this case it is necessary to prove the holomorphy with respect to λ'. By hypothesis, $\mathbf{f}$ is holomorphic in $(\mathbf{y}, \lambda')$. This suggests an application of Theorem III, which comes to nothing, because $\mathbf{f}^*$ is not holomorphic in $\mathbf{y}$. However, the proof of Theorem III can be carried over. Beginning with the term $\mathbf{u}_0 := \mathbf{y}(x; \xi, \boldsymbol{\eta}, \lambda_0)$, a sequence $(\mathbf{u}_k)$ of successive approximations is constructed with respect to the equation for $\mathbf{f}^*$. Since $(\mathbf{u}_k)$ tends uniformly to $\mathbf{y}$ and since $\mathbf{y}$ is in $S_{\beta/2}$ for $(\xi, \boldsymbol{\eta}, \lambda) \in S_\alpha \times U_\alpha$, it follows that $\mathbf{u}_k$ remains in S_β for these parameter values for $k \geq k_0$. Now, $\mathbf{u}_0$ is the solution $\mathbf{y}(\ldots, \lambda_0)$, which implies $\mathbf{u}_k(\ldots, \lambda_0) = \mathbf{y}(\ldots, \lambda_0)$ for all k. Since $\mathbf{y}$ remains in $S_{\beta/2}$, as was already mentioned, there exists a $\gamma > 0$ such that if $|\lambda - \lambda_0| < \gamma$, the first k_0 terms of the sequence $(\mathbf{u}_k)$ also lie in S_β. Since, however, $\mathbf{f}^* = \mathbf{f}$ is holomorphic in $\mathbf{y}$ on this set, all of the $\mathbf{u}_k$ are holomorphic in λ', and the same holds for $\mathbf{y}$, at least for $|\lambda - \lambda_0| < \gamma$. In this proof one can take any element of U_α for λ_0, and the holomorphy of the solution in U_α follows. ∎

Remark. (a) Note that the derivatives of $\mathbf{y}$ with respect to ξ, $\boldsymbol{\eta}$, and λ' all satisfy a linear integral equation of the form

$$\mathbf{z}(x; \boldsymbol{\mu}) = \mathbf{h}(x; \boldsymbol{\mu}) + \int_\xi^x \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}(t; \boldsymbol{\mu}), \lambda)\mathbf{z}(t; \boldsymbol{\mu})\, dt, \tag{19}$$

where $\boldsymbol{\mu} = (\xi, \boldsymbol{\eta}, \lambda)$, and

$$\begin{aligned} \mathbf{z} = \partial \mathbf{y}/\partial \xi : &\quad \mathbf{h} = -\mathbf{f}(\xi, \boldsymbol{\eta}, \lambda), \\ \mathbf{z} = \partial \mathbf{y}/\partial \eta_i : &\quad \mathbf{h} = \mathbf{e}_i \qquad (i\text{th unit vector}), \\ \mathbf{z} = \partial \mathbf{y}/\partial \lambda' : &\quad \mathbf{h} = \textstyle\int_\xi^x \mathbf{f}_{\lambda'}(t, \mathbf{y}(t; \xi, \boldsymbol{\eta}, \lambda), \lambda)\, dt. \end{aligned}$$

Thus in each case, one has a linear differential equation for $\mathbf{z}$,

$$\mathbf{z}'(x; \boldsymbol{\mu}) = \mathbf{f}_{\mathbf{y}}(x, \mathbf{y}(x; \boldsymbol{\mu}), \lambda)\mathbf{z}(x; \boldsymbol{\mu}) \quad [+\mathbf{f}_{\lambda'}(x, \mathbf{y}(x; \boldsymbol{\mu}), \lambda)], \tag{20}$$

where the term in square brackets appears in the case $\mathbf{z} = \partial \mathbf{y}/\partial \lambda'$.

(b) Based on these observations, the question of the existence of higher derivatives can be easily answered, since Theorems II, III, and VI apply to the (linear!) integral equation (19). If the partial derivatives of $\mathbf{f}$, which occur in a formal differentiation of (19), exist and are continuous, then the differentiation is "allowed," and the corresponding derivative of $\mathbf{y}$ exists and is continuous. Here one always obtains an integral equation of the form (19), but each time with a different $\mathbf{h}$. In particular, the following is true.

Corollary. *If* $\mathbf{f}$ *is* p*-times continuously differentiable with respect to all variables* x, $\mathbf{y}$, λ *in* $D \times K^\circ$, *then* $\mathbf{y}$ *is* p*-times continuously differentiable with respect to all the variables* x, ξ, $\boldsymbol{\eta}$, λ *in* $J \times S_\alpha \times U_\alpha$. *The same holds for* $\mathbf{y}'$, *because of the differential equation* (16).

XII. Exercise. Show that in the initial value problem for the differential equation with separated variables

$$y'(x) = f(x)g(y), \quad y(\xi) = \eta,$$

the derivatives of $y(x;\xi,\eta)$ in the case $g(\eta) \neq 0$ are given by

$$y_\xi(x;\xi,\eta) = -f(\xi)g(y(x;\xi,\eta)),$$
$$y_\eta(x;\xi,\eta) = g(y(x;\xi,\eta))/g(\eta).$$

Here only continuity of f and g are assumed.

Hint: Differentiate the identity (1.8).

XIII. Theorem. *Suppose that* $\mathbf{f}$ *satisfies a local Lipschitz condition with respect to* $\mathbf{y}$ *in the open set* $D \subset \mathbb{R}^{n+1}$. *Let* $\mathbf{y}(x;\xi,\boldsymbol{\eta})$ *denote the (maximally extended) solution of the initial value problem* (11) *and* $E \subset \mathbb{R}^{n+2}$ *the domain of this function (that is,* E *is the set of all* $(x;\xi,\boldsymbol{\eta})$ *such that the solution of* (11) *exists from* ξ *to* x*). Then* E *is open, and* $\mathbf{y}(x;\xi,\boldsymbol{\eta}) : E \to \mathbb{R}^n$ *is continuous.*

Proof as an exercise. *Hint:* If $(x,\xi,\boldsymbol{\eta}) \in E$ and, say, $\xi \leq x$, then the solution exists in $J = [\xi - \varepsilon, x + \varepsilon]$ ($\varepsilon > 0$ small). Construct a strip $S_\alpha \subset D$ as was done in Theorem X and apply Theorem II.

Chapter IV
Linear Differential Equations

§ 14. Linear Systems

I. Matrices. We denote $n \times n$ matrices by uppercase italic letters,

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = (a_{ij}),$$

where $a_{ij} \in \mathbb{R}$ or $\mathbb{C}$. With the usual definitions of addition and scalar multiplication of matrices,

$$A + B = (a_{ij} + b_{ij}), \qquad \lambda A = (\lambda a_{ij}),$$

the set of all $n \times n$ matrices forms a real or complex vector space. One can interpret this space as $\mathbb{R}^{n^2}$ (or, for complex a_{ij}, b_{ij}, λ, as $\mathbb{C}^{n^2}$). Matrix multiplication is defined by

$$AB = C \iff c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

It is not commutative. We recall the definition of the determinant of A:

$$\det A = \sum_p (-1)^{\nu(p)} a_{1p_1} a_{2p_2} \cdots a_{np_n}, \tag{1}$$

where $p = (p_1, \ldots, p_n)$ runs through all permutations of the numbers $1, \ldots, n$ and $\nu(p)$ is the number of inversions of p.

For $n \times 1$-matrices, i.e., column vectors, the previous notation is used

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \quad \text{occasionally also} \quad \mathbf{a} = (a_1, \ldots, a_n)^\top.$$

The expression $A^\top$ denotes the transpose of the matrix A.

The notation

$$A = (\mathbf{a}_1, \ldots, \mathbf{a}_n) \quad \text{with} \quad \mathbf{a}_i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix}$$

is self-explanatory. In particular,

$$I = (\mathbf{e}_1, \ldots, \mathbf{e}_n) = (\delta_{ij}) \quad \text{with} \quad \delta_{ij} = \begin{cases} 1 & \text{for} \quad i = j, \\ 0 & \text{for} \quad i \neq j \end{cases}$$

is the *identity matrix*, $\mathbf{e}_i$ the ith unit vector. Finally,

$$A\mathbf{x} = \mathbf{y} \quad \Longleftrightarrow \quad y_i = \sum_{j=1}^{n} a_{ij} x_j .$$

II. Compatible Norms. Let $|A|$ be a matrix norm, that is, a norm in $\mathbb{R}^{n^2}$ (or $\mathbb{C}^{n^2}$), and $|\mathbf{a}|$ a vector norm in $\mathbb{R}^n$ (or $\mathbb{C}^n$). Here we consider only *compatible norms*, i.e., those for which the inequalities

$$|AB| \leq |A| \cdot |B|, \tag{2}$$

$$|A\mathbf{x}| \leq |A| \cdot |\mathbf{x}| \tag{3}$$

are satisfied.

Example. Let

$$|A| = \sum_{i,j} |a_{ij}|, \qquad |\mathbf{x}| = \max |x_i|.$$

We show that these norms are compatible. First, for $C = AB$,

$$|C| \leq \sum_{i,j} \sum_{k} |a_{ik} b_{kj}| \leq \sum_{i,j,k,l} |a_{ik} b_{lj}| = |A| \cdot |B|,$$

which implies (2); and second, for $\mathbf{y} = A\mathbf{x}$,

$$|y_i| \leq \sum_{j} |a_{ij} x_j| \leq |\mathbf{x}| \sum_{j} |a_{ij}| \leq |A| \cdot |\mathbf{x}|,$$

and since this is true for every i, it follows that (3) holds.

Exercise. Show that the two Euclidean norms $|x|_e$ and

$$|A|_e = \sqrt{\sum_{i,j} |a_{ij}|^2}$$

are compatible.

III. Matrices with Variable Elements. For matrices $A(t) = (a_{ij}(t))$ with elements depending on t, the notation introduced in §10 for vectors is used, in particular,

$$A'(t) = (a'_{ij}(t)), \quad \int_a^b A(t)\,dt = \left(\int_a^b a_{ij}(t)\,dt \right).$$

If $A(t)$ is continuous, then by Lemma 10.VIII,

$$\left| \int_a^b A(t)\,dt \right| \leq \int_a^b |A(t)|\,dt \tag{4}$$

(one can consider A as a vector with n^2 components).

IV. Lemma. *If the functions $A(t)$, $B(t)$, $\mathbf{x}(t)$ are differentiable (at a point t_0), then each of the following functions is differentiable, and the corresponding derivative formulas hold:*

$$\begin{aligned} (AB)' &= A'B + AB', \\ (A\mathbf{x})' &= A'\mathbf{x} + A\mathbf{x}', \\ (\det A)' &= \sum_{i=1}^{n} \det(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}'_i, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n). \end{aligned} \tag{5}$$

Proof. The first two formulas follow immediately from the definition of the products. Formula (5) for the derivative of the determinant follows from the representation (1) and the product rule of differentiation. ∎

V. Systems of n Linear Differential Equations. We consider a system of differential equations of the following form:

$$\begin{array}{l} y'_1 = a_{11}(t)y_1 + \cdots + a_{1n}(t)y_n + b_1(t) \\ \vdots \quad\quad \vdots \quad\quad\quad\quad\quad \vdots \quad\quad\quad \vdots \\ y'_n = a_{n1}(t)y_1 + \cdots + a_{nn}(t)y_n + b_n(t) \end{array} \tag{6}$$

or, written in matrix notation,

$$\mathbf{y}' = A(t)\mathbf{y} + \mathbf{b}(t), \tag{6'}$$

where

$$A(t) = (a_{ij}(t)), \quad \mathbf{b}(t) = (b_1(t), \ldots, b_n(t))^\top. \tag{7}$$

It is common practice with linear systems to denote the independent variable by t instead of x, because in many physical applications the independent variable represents time. Note that while t is always assumed real, the functions that appear can be either real-valued or complex-valued.

VI. Existence, Uniqueness, and Estimation Theorem. *Let the real- or complex-valued functions $A(t)$, $\mathbf{b}(t)$ be continuous in an (arbitrary) interval J, and let $\tau \in J$. Then the initial value problem*

$$\mathbf{y}' = A(t)\mathbf{y} + \mathbf{b}(t), \quad \mathbf{y}(\tau) = \boldsymbol{\eta} \tag{8}$$

has exactly one solution $\mathbf{y}(t)$ for a given $\boldsymbol{\eta} \in \mathbb{R}^n$ or $\mathbb{C}^n$, respectively. The solution exists in all of J.

If J' is a subinterval of J, $\tau \in J'$ and

$$|A(t)| \le L \quad \text{and} \quad |\mathbf{b}(t)| \le \delta \quad \text{in} \quad J', \quad |\boldsymbol{\eta}| \le \gamma, \tag{9}$$

then $\mathbf{y}(t)$ satisfies

$$|\mathbf{y}(t)| \le \gamma e^{L|t-\tau|} + \frac{\delta}{L}(e^{L|t-\tau|} - 1) \quad \text{in} \quad J'. \tag{10}$$

The solution $\mathbf{y}(t)$ depends continuously on $A(t)$, $\mathbf{b}(t)$, and $\boldsymbol{\eta}$ on every compact subinterval $J' \subset J$, i.e., for every $\varepsilon > 0$ there exists $\beta > 0$ such that if the inequalities

$$|B(t) - A(t)| < \beta, \quad |\mathbf{b}(t) - \mathbf{c}(t)| < \beta \quad \text{in} \quad J', \quad |\boldsymbol{\eta} - \boldsymbol{\zeta}| < \beta \tag{11}$$

are satisfied and if $\mathbf{z}(t)$ is a solution of the initial value problem

$$\mathbf{z}' = B(t)\mathbf{z} + \mathbf{c}(t), \quad \mathbf{z}(\tau) = \boldsymbol{\zeta} \tag{12}$$

($B, \mathbf{c}$ are continuous), then the inequality

$$|\mathbf{z}(t) - \mathbf{y}(t)| < \varepsilon \quad \text{in} \quad J' \tag{13}$$

holds.

Proof. Suppose J' is a compact subinterval. Then, since a continuous function on J' assumes its maximum, it follows that there exist constants L, δ such that (9) holds. In $D = J' \times \mathbb{R}^n$, respectively $J' \times \mathbb{C}^n$, $\mathbf{f}(t, \mathbf{y}) = A\mathbf{y} + \mathbf{b}$ satisfies a Lipschitz condition (4)

$$|(A\mathbf{y} + \mathbf{b}) - (A\bar{\mathbf{y}} - \mathbf{b})| \le |A|\,|\mathbf{y} - \bar{\mathbf{y}}| \le L|\mathbf{y} - \bar{\mathbf{y}}|.$$

Therefore, by Theorem 10.VII, problem (8) has exactly one solution, defined in all of J'. We use Theorem 12.V with $\mathbf{f}(t, \mathbf{y}) = A\mathbf{y} + \mathbf{b}$ and $\mathbf{z}(t) \equiv 0$:

$$|\mathbf{z}' - \mathbf{f}(t, \mathbf{z})| = |\mathbf{b}(t)| \le \delta, \quad |\mathbf{z}(\tau) - \mathbf{y}(\tau)| = |\boldsymbol{\eta}| \le \gamma.$$

That is, (12.5) holds, as does the estimate (12.6), which is identical to (10).

Since in this reasoning J' can be any compact subinterval of J, the solutions of (8) exist in all of J.

Now let $J' \subset J$ be a compact interval, and let (9) and (13) hold, where it is assumed that $\beta \le 1$. Then there exists a constant c such that

$$|B(t)| \le c, \quad |\mathbf{c}(t)| \le c, \quad |\boldsymbol{\zeta}| \le c,$$

and thus every solution $\mathbf{z}(t)$ of (12) satisfies (10) with $\gamma = L = \delta = c$, and hence $|\mathbf{z}(t)| \le c_1$ in J'. We now apply Theorem 12.V to the two solutions $\mathbf{y}(t)$ and $\mathbf{z}(t)$ of (8) and (12) and to $\mathbf{f}(t, \mathbf{y}) = A\mathbf{y} + \mathbf{b}$. The result is

$$\begin{aligned} |\mathbf{z}' - \mathbf{f}(t, \mathbf{z})| &= |(B - A)\mathbf{z} + \mathbf{c} - \mathbf{b}| \\ &\le |B - A| \cdot |\mathbf{z}| + |\mathbf{c} - \mathbf{b}| \\ &\le \beta(1 + c_1). \end{aligned}$$

Therefore, (12.6) holds with $\gamma = \beta$, $\delta = \beta(1 + c_1)$. If $\beta > 0$ is chosen sufficiently small, the inequality (13) follows. ∎

VII. Complex Systems Versus Real Systems. Let

$$\mathbf{z}' = B(t)\mathbf{z} + \mathbf{b}(t) \tag{14}$$

be a complex system of n differential equations, where

$$\mathbf{z} = \mathbf{x} + \mathrm{i}\mathbf{y}, \; \mathbf{b} = \mathbf{c} + \mathrm{i}\mathbf{d}, \; B(t) = C(t) + \mathrm{i}D(t) \quad \text{with} \quad \mathrm{i} = \sqrt{-1}.$$

By separating (14) into real and imaginary parts, one obtains the pair of differential equations

$$\begin{aligned} \mathbf{x}' &= C(t)\mathbf{x} - D(t)\mathbf{y} + \mathbf{c}(t), \\ \mathbf{y}' &= C(t)\mathbf{y} + D(t)\mathbf{x} + \mathbf{d}(t). \end{aligned} \tag{15}$$

This real system of $2n$ equations can be written in the form

$$\mathbf{u}' = A(t)\mathbf{u} + \mathbf{a}(t), \tag{15$'$}$$

where $\mathbf{u} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$ and $\mathbf{a} = \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix}$ are $2n$-dimensional column vectors and

$$A = \begin{pmatrix} C & -D \\ D & C \end{pmatrix} \tag{16}$$

is a real $2n \times 2n$ matrix. Therefore, every complex system of order n is equivalent to a real system of order $2n$ of the special form (15′), (16).

VIII. Exercise. Let $c_k > 0$ for $k = 1, \ldots, n$. Show that

$$|\mathbf{x}| := \sum_{k=1}^{n} c_k |x_k|$$

and

$$|A| := \max_m \frac{1}{c_m} \sum_{k=1}^{n} c_k |a_{km}|$$

define compatible norms in the sense of II.

IX. Exercise. The Operator Norm. If $|\mathbf{x}|$ is an arbitrary vector norm, then

$$|A| := \max\{|A\mathbf{x}| : |\mathbf{x}| \le 1\}$$

defines a compatible matrix norm (proof as an exercise). This norm is called the *operator norm* of A. The operator norm is the smallest number γ such that $|A\mathbf{x}| \le \gamma|\mathbf{x}|$ for all $\mathbf{x}$.

Show that the norm given in VIII is the operator norm (for all A) and that the two matrix norms defined in II are, in general, not operator norms (calculate the operator norm of $A = I$).

§ 15. Homogeneous Linear Systems

A linear system of differential equations $\mathbf{y}' = A(t)\mathbf{y} + \mathbf{b}(t)$ is called *homogeneous* if $\mathbf{b}(t) \equiv 0$; otherwise, it is called *inhomogeneous*. Within this section, the term *solution* refers to a solution of the homogeneous system

$$\mathbf{y}' = A(t)\mathbf{y}, \tag{1}$$

where $A(t)$ is continuous in the (arbitrary) interval J. By Theorem 14.VI, there exists exactly one solution $\mathbf{y} = \mathbf{y}(t;\tau,\boldsymbol{\eta})$ for every $\tau \in J$, $\boldsymbol{\eta} \in \mathbb{R}^n$ or $\mathbb{C}^n$, and this solution exists in J.

I. Theorem. *If $A(t)$ is real-valued (complex-valued) and continuous in J, then the set of real (complex) solutions $\mathbf{y}(t)$ of the homogeneous equation* (1) *forms an n-dimensional real (complex) linear space.*

For fixed $\tau \in J$ the mapping

$$\boldsymbol{\eta} \mapsto \mathbf{y}(t;\tau,\boldsymbol{\eta})$$

defines an isomorphism (a linear, bijective mapping) between $\mathbb{R}^n$ ($\mathbb{C}^n$) and the space of solutions.

This theorem is a simple consequence of the *superposition principle*, which states that every linear combination

$$\mathbf{y} = c_1\mathbf{y}_1 + \cdots + c_k\mathbf{y}_k$$

of solutions is again a solution. It follows that

$$\mathbf{y}(t;\tau,\lambda\boldsymbol{\eta} + \lambda'\boldsymbol{\eta}') = \lambda\mathbf{y}(t;\tau,\boldsymbol{\eta}) + \lambda'\mathbf{y}(t;\tau,\boldsymbol{\eta}'),$$

since the functions on the left and the right are both solutions with the same initial value $\lambda\boldsymbol{\eta} + \lambda'\boldsymbol{\eta}'$ at τ. Therefore, the mapping $\boldsymbol{\eta} \mapsto \mathbf{y}(t;\tau,\boldsymbol{\eta})$ is linear. In particular, the zero solution $\mathbf{y} \equiv \mathbf{0}$ is the image of the zero vector $\boldsymbol{\eta} = \mathbf{0}$. The remaining statements are evident. ∎

The isomorphism $\boldsymbol{\eta} \mapsto \mathbf{y}(t;\tau,\boldsymbol{\eta})$, which is well-defined for every fixed $\tau \in J$, leads immediately to some important

II. Propositions. (a) If $\mathbf{y}$ is a solution and $\mathbf{y}(t_0) = \mathbf{0}$ for some $t_0 \in J$, then $\mathbf{y} \equiv \mathbf{0}$ in J.

(b) A set of k solutions $\mathbf{y}_1, \ldots, \mathbf{y}_k$ is called *linearly dependent* if there exist constants $c_1, \ldots, c_k$ with $|c_1| + \cdots |c_k| > 0$ such that

$$c_1\mathbf{y}_1 + \cdots + c_k\mathbf{y}_k = \mathbf{0}.$$

Because of (a) and the superposition principle, this equation holds identically if it holds at one point in J. The k solutions are said to be *linearly independent* if they are not linearly dependent.

(c) We recall that k vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{R}^n$ are linearly independent in $\mathbb{R}^n$ if and only if they are linearly independent when considered as vectors in $\mathbb{C}^n$. A similar statement holds for real solutions of (1) when $A(t)$ is real.

(d) For $k > n$, any k solutions $\mathbf{y}_1, \ldots, \mathbf{y}_k$ are linearly dependent.

(e) There exist n linearly independent solutions $\mathbf{y}_1, \ldots, \mathbf{y}_n$. Every such set of n linearly independent solutions is called a *fundamental system* of solutions. If $\mathbf{y}_1, \ldots, \mathbf{y}_n$ is a fundamental system, then *every* solution $\mathbf{y}$ can be written in a unique way as a linear combination

$$\mathbf{y} = c_1\mathbf{y}_1 + \cdots + c_n\mathbf{y}_n. \tag{2}$$

(f) A system of n solutions $\mathbf{y}_1, \ldots, \mathbf{y}_n$ can be assembled into an $n \times n$ *solution matrix*

$$Y(t) = (\mathbf{y}_1, \ldots, \mathbf{y}_n).$$

In this notation, the n differential equations $\mathbf{y}_i' = A(t)\mathbf{y}_i$ $(i = 1, \ldots, n)$ can be written as a matrix differential equation

$$Y' = A(t)Y; \tag{3}$$

it is easy to see that this matrix equation is equivalent to the n differential equations. The solution $Y(t)$ of (3) is uniquely determined by the specification of an initial condition $Y(\tau) = C$. Here $Y(t)$ is a fundamental system (also called *fundamental matrix*) if and only if the matrix C is nonsingular. It follows then from (b) that $Y(t)$ is nonsingular for every $t \in J$. If $Y(t)$ is a fundamental matrix, then every solution of (1) can be written in the form

$$\mathbf{y} = Y\mathbf{c}, \quad \mathbf{c} \in \mathbb{R}^n \text{ or } \mathbb{C}^n, \text{ respectively.} \tag{2'}$$

This equation is identical to (2).

(g) A special fundamental matrix $X(t)$ is obtained from the initial value problem

$$X' = A(t)X, \quad X(\tau) = I. \tag{4}$$

Problem (4) is identical to the n initial value problems

$$\mathbf{x}_i' = A(t)\mathbf{x}_i, \quad \mathbf{x}_i(\tau) = \mathbf{e}_i \qquad (i = 1, \ldots, n). \tag{4'}$$

Using the solution matrix $X(t)$, the solution to any initial value problem can be immediately given:

$$\mathbf{y}' = A(t)\mathbf{y}, \quad \mathbf{y}(\tau) = \boldsymbol{\eta} \iff \mathbf{y}(t) = X(t)\boldsymbol{\eta}.$$

(h) The following extension of (f) is true: If $Y(t)$ is a solution to (3) and C is a constant matrix, then $Z(t) = Y(t)C$ is also a solution to (3). We have namely

$$Z' = Y'C = AYC = AZ.$$

If $Y(t)$ is a fundamental matrix and C nonsingular, then $Z(t)$ is again a fundamental matrix; moreover, every fundamental matrix can be represented in the form $Y(t)C$ with C nonsingular.

In particular, every solution matrix $Y(t)$ satisfies

$$Y(t) = X(t)Y(\tau), \tag{5}$$

where $X(t)$ is the solution to (4). This follows because the right side is a solution, and it has the correct initial values $X(\tau)Y(\tau) = IY(\tau) = Y(\tau)$.

III. The Wronskian. If $Y(t) = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ is a solution of (3), then its determinant $\phi(t) := \det Y(t)$ is called the *Wronskian determinant* or *Wronskian*[1], of $Y(t)$.

Theorem. *If $A(t)$ is continuous in J, then the Wronskian $\phi(t) := \det Y(t)$ of a solution of (3) satisfies the differential equation*

$$\phi' = (\operatorname{tr} A(t))\phi \quad \textit{in} \quad J, \tag{6}$$

where

$$\operatorname{tr} A(t) = a_{11}(t) + a_{22}(t) + \cdots + a_{nn}(t)$$

is the trace of A. Hence

$$\phi(t) = \phi(\tau) \exp\left(\int_\tau^t \operatorname{tr}(A(s))\, ds\right). \tag{7}$$

In particular, the Wronskian of the solution $X(t)$ of (4) is given by

$$\det X(t) = \exp\left(\int_\tau^t \operatorname{tr} A(s)\, ds\right). \tag{8}$$

The theorem shows that the Wronskian can be calculated from the initial value $Y(\tau)$ alone without knowledge of the solution.

[1] After J.M. Hoene-Wronski (1778–1853), Polish mathematician.

Proof. By (14.5) we have

$$(\det X(t))' = \sum_{i=1}^{n} \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_i', \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n).$$

Hence, because $\mathbf{x}_i(\tau) = \mathbf{e}_i$, $\mathbf{x}_i'(\tau) = A(\tau)\mathbf{e}_i$, we have

$$\begin{aligned}(\det X(\tau))' &= \sum_{i=1}^{n} \det(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{i-1}, A(\tau)\mathbf{e}_i, \mathbf{e}_{i+1}, \ldots, \mathbf{e}_n) \\ &= \sum_{i=1}^{n} a_{ij}(\tau) = \operatorname{tr} A(\tau).\end{aligned}$$

By (5), the function $\phi(t) = \det Y(t)$ satisfies $\phi(t) = \phi(\tau) \det X(t)$. Hence

$$\phi'(t) = \phi(\tau)\,(\det X(t))',$$

and in particular,

$$\phi'(\tau) = \phi(\tau) \operatorname{tr} A(\tau).$$

Since this argument can be applied at every point τ, it follows that $\phi(t)$ satisfies the linear differential equation (6) in all of J. ∎

Corollary. *The Wronskian is either* $\equiv 0$ *or* $\neq 0$ *in* J. *The nonvanishing of the Wronskian is a necessary and sufficient condition for* $Y(t)$ *to be a fundamental matrix (this follows already from* II.(b)*).*

IV. D'Alembert's Method of Reduction of Order. In general, it is not possible to give the solutions of a homogeneous system in closed form. However, if one solution is known, it is possible to reduce the system to a system of $n-1$ differential equations. If $\mathbf{x}(t)$ is a (known) solution of the differential equation (1), then for the remaining solutions one makes the ansatz

$$\mathbf{y}(t) = \phi(t)\mathbf{x}(t) + \mathbf{z}(t) \qquad \text{with} \qquad \mathbf{z}(t) = \begin{pmatrix} 0 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \tag{9}$$

(ϕ scalar). This function is a solution of (1) precisely if

$$\mathbf{y}' = \phi'\mathbf{x} + \phi\mathbf{x}' + \mathbf{z}' = \phi A\mathbf{x} + A\mathbf{z},$$

i.e., if

$$\mathbf{z}' = A\mathbf{z} - \phi'\mathbf{x}.$$

The equation for the first component is

$$\sum_{j=2}^{n} a_{1j} z_j = \phi' x_1, \tag{10}$$

and for the ith component $(2 \le i \le n)$ it is

$$z_i' = \sum_{j=2}^{n} a_{ij} z_j - \phi' x_i.$$

Thus, for the components z_i, one obtains the differential equations

$$z_i' = \sum_{j=2}^{n} \left(a_{ij} - \frac{x_i}{x_1} a_{1j} \right) z_j \qquad (i = 2, \ldots, n), \tag{11}$$

that is, a homogeneous linear system of $n-1$ equations. Here it is assumed, without loss of generality, that $x_1(t) \neq 0$ (instead of the first component, any other component can be chosen). If $(z_2, \ldots, z_n)$ is a solution to this system, then from (10),

$$\phi(t) = \int \frac{1}{x_1} \sum_{j=2}^{n} a_{1j} z_j \, dt, \tag{12}$$

and a solution $\mathbf{y}(t)$ of (1) is obtained by (9).

If (11) has been completely solved, i.e., if a fundamental system has been found, then this procedure leads to $n-1$ solutions $\mathbf{y}_1, \ldots, \mathbf{y}_{n-1}$ of the original differential equation (1). Combining these solutions with $\mathbf{x}$ gives a fundamental system for (1).

To *prove* the linear independence of these n solutions, let

$$\mathbf{y}_i = \phi_i \mathbf{x} + \mathbf{z}_i \qquad (i = 1, \ldots, n-1)$$

and consider the equation

$$\lambda \mathbf{x} + \lambda_1 \mathbf{y}_1 + \cdots + \lambda_{n-1} \mathbf{y}_{n-1} = \mathbf{0}.$$

Since the first component of each $\mathbf{z}_i$ vanishes, the first component of this equation, divided by x_1, reads

$$\lambda + \lambda_1 \phi_1 + \cdots + \lambda_{n-1} \phi_{n-1} = 0.$$

Multiplying this equation by $\mathbf{x}$ and subtracting from the previous one gives

$$\lambda_1 \mathbf{z}_1 + \cdots + \lambda_{n-1} \mathbf{z}_{n-1} = \mathbf{0}.$$

Therefore, since the $\mathbf{z}_i$ are linearly independent, $\lambda_1 = \cdots = \lambda_{n-1} = 0$, and consequently $\lambda = 0$. ∎

V. Example. The system

$$y_1' = \frac{1}{t}y_1 - y_2, \qquad\qquad A(t) = \begin{pmatrix} \dfrac{1}{t} & -1 \\ \dfrac{1}{t^2} & \dfrac{2}{t} \end{pmatrix}$$
$$y_2' = \frac{1}{t^2}y_1 + \frac{2}{t}y_2,$$

has the solution

$$\mathbf{x}(t) = \begin{pmatrix} t^2 \\ -t \end{pmatrix}.$$

Here (11) reduces to a single equation for $z_2(t) = z(t)$,

$$z' = \left(\frac{2}{t} - \frac{t}{t^2}\right) z = \frac{1}{t}z.$$

One solution is $z(t) = t$. We choose the basic interval to be $J = (0, \infty)$. By (12), then,

$$\phi(t) = \int \frac{1}{t^2}(-t)\,dt = -\ln t.$$

Therefore, in view of (9),

$$\mathbf{y}(t) = -\mathbf{x}(t)\ln t + \begin{pmatrix} 0 \\ t \end{pmatrix} = \begin{pmatrix} -t^2 \ln t \\ t + t\ln t \end{pmatrix}$$

is another solution of the original system. The system of solutions

$$Y(t) = (\mathbf{x}, \mathbf{y}) = \begin{pmatrix} t^2 & -t^2 \ln t \\ -t & t + t\ln t \end{pmatrix} \quad \text{with} \quad Y(1) = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \tag{13}$$

is a fundamental system. If the second column in $Y(t)$ is added to the first and the second is left unchanged, then one obtains the identity matrix for $t = 1$, i.e.,

$$X(t) = (\mathbf{x} + \mathbf{y}, \mathbf{y}) = \begin{pmatrix} t^2(1 - \ln t) & -t^2 \ln t \\ t \ln t & t(1 + \ln t) \end{pmatrix}$$

is a solution with $X(1) = I$.

VI. Exercise. The Adjoint Equation. Let $C^* = \bar{C}^\top$ be the complex conjugate transpose of the matrix C, that is, $c_{ij}^* = \bar{c}_{ji}$. If $CC^* = I$, then C is called *unitary*. The star operation obeys the rules

$$(BC)^* = C^*B^*, \quad (C^*)^{-1} = (C^{-1})^*, \quad (C^*)^* = C,$$

and $\langle C\mathbf{y}, \mathbf{z}\rangle = \langle \mathbf{y}, C^*\mathbf{z}\rangle$ for $\mathbf{y}, \mathbf{z} \in \mathbb{C}^n$, where $\langle \mathbf{a}, \mathbf{b}\rangle = a_1\bar{b}_1 + \cdots + a_n\bar{b}_n$ is the inner product in $\mathbb{C}^n$. Further, if $C = C(t)$ is differentiable, then

$$(C'(t))^* = (C^*(t))'.$$

The equation *adjoint* to equation (1), $\mathbf{y}' = A(t)\mathbf{y}$, is given by

$$\mathbf{z}' = -A^*(t)\mathbf{z}. \tag{14}$$

Similarly, the differential operator M corresponding to (14), that is, $M\mathbf{z} = \mathbf{z}' + A^*(t)\mathbf{z}$, is called the *adjoint operator* to the operator L, which corresponds to (1) and is given by $L\mathbf{y} = \mathbf{y}' - A(t)\mathbf{y}$. The adjoint to M is again L. The operator L is called *self-adjoint* if $L = M$, i.e., $A = -A^*$.

(a) Let $Y(t)$ be a fundamental matrix for equation (1). Then $Z(t)$ is a fundamental matrix for equation (14) if and only if $Y^*(t)Z(t) = C$ is a nonsingular constant matrix.

(b) *The Lagrange identity* for $\mathbf{y}(t), \mathbf{z}(t) \in C^1(J)$ reads

$$\langle L\mathbf{y}, \mathbf{z}\rangle + \langle \mathbf{y}, M\mathbf{z}\rangle = \frac{d}{dt}\langle \mathbf{y}, \mathbf{z}\rangle.$$

(c) If L is self-adjoint, $\tau \in J$, and $Y(t)$ is a fundamental matrix for (1) with the property that $Y(\tau)$ is unitary, then $Y(t)$ is unitary for all $t \in J$.

Hint for (a): $(Y^*Z)' = Y^*A^*Z + Y^*Z' = 0 \iff Z' = -A^*Z$.

Remark on the Real Case. If C is real, then $C^* = C^\top$. In this case unitary matrices are called *orthogonal*. If (1) is a real system, then L is self-adjoint if and only if $A(t)$ is skew-symmetric.

§ 16. Inhomogeneous Systems

As earlier, $A(t)$, $\mathbf{b}(t)$ are defined and continuous in an interval J (real- or complex-valued).

The following theorem gives the relationship between solutions of the inhomogeneous differential equation

$$\mathbf{y}' = A(t)\mathbf{y} + \mathbf{b}(t) \tag{1}$$

and solutions of the corresponding homogeneous differential equation.

I. Theorem. *Let $\bar{\mathbf{y}}(t)$ be a fixed solution of the inhomogeneous equation (1). If $\mathbf{x}(t)$ is an arbitrary solution of the homogeneous equation, then*

$$\mathbf{y}(t) = \bar{\mathbf{y}}(t) + \mathbf{x}(t)$$

is a solution of the inhomogeneous equation, and all solutions of the inhomogeneous equation are obtained in this way.

As in the case $n = 1$, the *proof* rests on the simple fact that the difference between two solutions of the inhomogeneous differential equation is a solution of the homogeneous equation. ∎

Thus our task is to find just *one* solution of the inhomogeneous equation. We make use of a procedure that originated with Lagrange, the

II. Method of Variation of Constants. If $Y(t)$ is a fundamental matrix to the homogeneous differential equation, then by 15.II.(e), every solution of the homogeneous equation can be represented in the form $Y(t)\mathbf{v}$, where $\mathbf{v}$ runs through all (constant) vectors. In the method of variation of constants the constants $(v_1, \ldots, v_n)$ are "varied," i.e., replaced by functions of t,

$$\mathbf{z}(t) = Y(t)\mathbf{v}(t).$$

The function $\mathbf{v}(t)$ is to be determined such that $\mathbf{z}(t)$ is a solution of the inhomogeneous differential equation (1). Substituting $\mathbf{z}(t)$ into (1) gives

$$\mathbf{z}' = Y'\mathbf{v} + Y\mathbf{v}' = AY\mathbf{v} + Y\mathbf{v}' = AY\mathbf{v} + \mathbf{b},$$

which leads to the condition

$$Y(t)\mathbf{v}' = \mathbf{b}(t). \tag{2}$$

Since Y is a fundamental matrix, the Wronskian $\det Y$ is $\neq 0$. Therefore, the inverse matrix Y^{-1} exists and is continuous in J. Multiplying equation (2) on the left by this matrix and integrating gives

$$\mathbf{v}(t) = \mathbf{v}(\tau) + \int_\tau^t Y^{-1}(s)\mathbf{b}(s)\,ds.$$

For instance, the solution $\mathbf{z}(t)$ with initial value $\mathbf{z}(\tau) = \mathbf{0}$ is given by

$$\mathbf{z}(t) = Y(t)\int_\tau^t Y^{-1}(s)\mathbf{b}(s)\,ds. \tag{3}$$

III. Theorem. *The initial value problem ($A(t)$, $\mathbf{b}(t) \in C(J)$, $\tau \in J$)*

$$\mathbf{y}' = A(t)\mathbf{y} + \mathbf{b}(t), \quad \mathbf{y}(\tau) = \boldsymbol{\eta}$$

has the (uniquely determined) solution

$$\mathbf{y}(t) = X(t)\boldsymbol{\eta} + \int_\tau^t X(t)X^{-1}(s)\mathbf{b}(s)\,ds, \tag{4}$$

where $X(t)$ is the fundamental matrix of the homogeneous differential equation with $X(\tau) = I$.

Proof. The first summand on the right side is a solution of the homogeneous equation with the initial value $\boldsymbol{\eta}$, and the second summand is a solution of the inhomogeneous equation with initial value $\mathbf{0}$ (see (3)). ∎

Remark. If $Y(t)$ is a fundamental matrix, then by (15.5) $Y(t) = X(t)Y(\tau)$, whence $Y^{-1}(t) = Y^{-1}(\tau)X^{-1}(t)$, and it follows that

$$Y(t)Y^{-1}(s) = X(t)X^{-1}(s).$$

Thus a representation of the solution $\mathbf{y}$ to the initial value problem in terms of $Y(t)$ is given by

$$\mathbf{y}(t) = Y(t)Y^{-1}(\tau)\boldsymbol{\eta} + \int_\tau^t Y(t)Y^{-1}(s)\mathbf{b}(s)\,ds. \tag{4'}$$

Example.

$$y_1' = \frac{1}{t}y_1 - y_2 + t, \qquad \mathbf{b}(t) = \begin{pmatrix} t \\ -t^2 \end{pmatrix} \qquad (t > 0).$$
$$y_2' = \frac{1}{t^2}y_1 + \frac{2}{t}y_2 - t^2,$$

The general solution to the corresponding homogeneous equation was found in 15.V. Using the well-known formula for the inverse of a 2×2-matrix,

$$B = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Longrightarrow B^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

and (15.13), one can easily calculate $Y^{-1}(t)$. The resulting matrix is

$$Y^{-1}(t) = \frac{1}{t^3}\begin{pmatrix} t(1 + \ln t) & t^2 \ln t \\ t & t^2 \end{pmatrix}.$$

Hence

$$Y^{-1}(t)\mathbf{b}(t) = \frac{1}{4}\begin{pmatrix} \ln t + 1 - t^2 \ln t \\ 1 - t^2 \end{pmatrix},$$

$$\int_1^t Y^{-1}(s)\mathbf{b}(s)\,ds = \frac{1}{4}\begin{pmatrix} t^2 - 1 + (4 - 2t^2 + 2\ln t)\ln t \\ 4\ln t - 2t^2 + 2 \end{pmatrix},$$

and therefore from (3),

$$\mathbf{z}(t) = Y(t)\int_1^t Y^{-1}(s)\mathbf{b}(s)\,ds = \frac{1}{4}\begin{pmatrix} t^2(t^2 - 1 + 2\ln t - 2\ln^2 t) \\ t(3 - 3t^2 + 2\ln t + 2\ln^2 t) \end{pmatrix}.$$

Thus we have found a particular solution of the inhomogeneous differential equation with initial value $\mathbf{z}(1) = \mathbf{0}$.

IV. Exercise. Show that the real linear system

$$x' = a(t)x - b(t)y,$$
$$y' = b(t)x + a(t)y$$

can be reduced to a single complex linear differential equation

$$z' = c(t)z$$

for $z(t) = x(t) + \mathrm{i}y(t)$. Derive a linear differential equation for $v(t) = z(t)\bar{z}(t) = x^2(t) + y^2(t)$.

Use this method to solve the system

$$x' = x\cos t - y\sin t,$$
$$y' = x\sin t + y\cos t.$$

In particular, determine a fundamental system $X(t)$ with $X(0) = I$ and compute its Wronskian $\det X(t)$. Show that every solution is periodic. What is the period?

Sketch the orbit $z(t) = (x(t), y(t))$ of the solution with initial values $(x(0), y(0)) = (1, 0)$ in the xy-plane. Determine $v(t) = |z(t)|^2$ and find two bounds $0 < \alpha \le v(t) \le \beta$ for this solution.

V. Exercise. Determine the general solution of the system

$$x' = (3t-1)x - (1-t)y + te^{t^2},$$
$$y' = -(t+2)x + (t-2)y - e^{t^2}.$$

Hint. The homogeneous system has a solution of the form $(x(t), y(t)) = (\phi(t), -\phi(t))$.

Supplement: L^1-Estimation of C-Solutions

We consider solutions in the sense of Carathéodory of the problem

$$\mathbf{y}' = A(t)\mathbf{y} + \mathbf{b}(t) \quad \text{in} \quad J = [\tau, \tau + a], \quad \mathbf{y}(\tau) = \boldsymbol{\eta} \tag{5}$$

under the assumption that (all components of) $A(t)$ and $\mathbf{b}(t)$ belong to $L(J)$. According to Theorem 10.XII, there exists a unique C-solution in J, and it is not difficult to show that the earlier results, in particular Theorems 15.I and 15.III for the homogeneous system and the representation formula (4) for the solution of problem (5), are also valid under these assumptions.

Our aim now is to establish *pointwise estimates* on $\mathbf{y}(t)$ and on the difference $\mathbf{y}(t) - \mathbf{z}(t)$, where $\mathbf{z}(t)$ satisfies

$$\mathbf{z}' = B(t)\mathbf{z} + \mathbf{c}(t) \quad \text{in} \quad J, \quad \mathbf{z}(\xi) = \boldsymbol{\zeta}, \tag{6}$$

in terms of *integral estimates* of the given functions. Note that in the corresponding Theorem 14.VI pointwise bounds (and not L^1 bounds) on these functions are required.

VI. Estimation Theorem. *Assume that A, B, $\mathbf{b}$, $\mathbf{c}$ belong to $L(J)$ and that $|A(t)|$, $|B(t)| \le h(t) \in L(J)$. Then the solutions $\mathbf{y}(t)$ of (5) and $\mathbf{z}(t)$ of (6) satisfy*

$$|\mathbf{y}(t)|e^{-H(t)} \le |\boldsymbol{\eta}| + \int_\tau^t e^{-H(s)} |\mathbf{b}(s)|\, ds, \tag{7}$$

where $H(t) = \int_\tau^t h(s)\, ds$ and

$$|\mathbf{y}(t) - \mathbf{z}(t)|e^{-H(t)} \le |\boldsymbol{\eta} - \boldsymbol{\zeta}| + \int_\tau^t e^{-H(s)} \{|\mathbf{b}(s) - \mathbf{c}(s)| + |A(s) - B(s)|\, |\mathbf{y}(s)|\}\, ds. \tag{8}$$

For the maximum norm $\|\mathbf{f}\|_\infty = \max_J |\mathbf{f}(t)|$ *and the* L^1*-norm* $\|\mathbf{f}\|_{L^1} = \int_\tau^{\tau+a} |\mathbf{f}(t)|\,dt$, *the estimates*

$$\|\mathbf{y}\|_\infty \le C(|\boldsymbol{\eta}| + \|\mathbf{b}\|_{L^1}), \quad C = \exp(\|h\|_{L^1}), \tag{9}$$

$$\|\mathbf{y}-\mathbf{z}\|_\infty \le C_1(|\boldsymbol{\eta}-\boldsymbol{\zeta}| + \|\mathbf{b}-\mathbf{c}\|_{L^1} + \|A-B\|_{L^1}) \tag{10}$$

hold, where C_1 *depends only on* $|\boldsymbol{\eta}|$, $\|h\|_{L^1}$ *and* $\|\mathbf{b}\|_{L^1}$.

Proof. By 10.XVI,

$$|\mathbf{y}(t)|' \le |\mathbf{y}'(t)| = |A(t)\mathbf{y} + \mathbf{b}(t)| \le h(t)|\mathbf{y}(t)| + |\mathbf{b}(t)|.$$

Hence $\phi(t) = |\mathbf{y}(t)|\mathrm{e}^{-H(t)}$ satisfies $\phi'(t) \le \mathrm{e}^{-H(t)}|\mathbf{b}(t)|$, $\phi(\tau) = |\boldsymbol{\eta}|$, which leads to (7) after integration.

The difference $\mathbf{u} = \mathbf{z} - \mathbf{y}$ satisfies

$$\mathbf{u}' = B\mathbf{z} + \mathbf{c} - A\mathbf{y} - \mathbf{b} = B\mathbf{u} + (\mathbf{c} - \mathbf{b}) + (B - A)\mathbf{y}.$$

The estimate (7), applied to $\mathbf{u}$ (with $\boldsymbol{\eta} - \boldsymbol{\zeta}$ instead of $\boldsymbol{\eta}$, B instead of A, and $(\mathbf{c}-\mathbf{b}) + (B-A)\mathbf{y}$ instead of $\mathbf{b}$) gives (8). ∎

Our next theorem deals with the linear case of the comparison theorem 10.XII in the context of C-solutions.

VII. Positivity Theorem. *Assume that the real matrix* $A(t) \in L(J)$, $J = [\tau, \tau + a]$, *is essentially positive; i.e.,* $a_{ij}(t) \ge 0$ *for* $i \ne j$. *Then, for* $\mathbf{u} \in AC(J)$,

$$\mathbf{u}' \ge A(t)\mathbf{u} \ \text{ a.e. in } J, \quad \mathbf{u}(\tau) \ge 0 \ \text{ implies } \mathbf{u}(t) \ge 0 \ \text{ in } J.$$

Moreover, if $u_i(t_1) > 0$, *then* $u_i(t) > 0$ *for* $t > t_1$.

Proof. Let $|A(t)| \le h(t)$, where $|\cdot|$ is the maximum norm, and $H(t) = \int_\tau^t h(s)\,ds$. Then $B(t) = A(t) + h(t)I \ge 0$, i.e., $b_{ij} \ge 0$ for all i, j, and $|B(t)| \le 2h(t)$. The function $\mathbf{w}(t) = \mathrm{e}^{H(t)}\mathbf{u}(t)$ satisfies $\mathbf{w}'(t) \ge B(t)\mathbf{w}$, and the function $\boldsymbol{\sigma} = (\rho, \rho, \ldots, \rho)$ with $\rho(t) = \mathrm{e}^{2nH(t)}$ also satisfies $\boldsymbol{\sigma}' \ge B(t)\boldsymbol{\sigma}$; both inequalities are easily established. Hence $\mathbf{w}_\varepsilon = \mathbf{w} + \varepsilon\boldsymbol{\sigma}$ satisfies $\mathbf{w}_\varepsilon' \ge B\mathbf{w}_\varepsilon$ and $\mathbf{w}_\varepsilon(0) > 0$. As long as $\mathbf{w}_\varepsilon(t) \ge 0$, we have $\mathbf{w}_\varepsilon'(t) \ge 0$, and this shows that $\mathbf{w}_\varepsilon(t)$ is increasing and positive in J. Since ε is arbitrary, $\mathbf{w}(t)$ is increasing, and both propositions about $\mathbf{u}(t)$ are obtained as a result. ∎

This theorem can be used to give an alternative proof for the comparison theorem 10.XII that is valid for C-solutions. As before, $P\mathbf{u} = \mathbf{u}' - \mathbf{f}(t, \mathbf{u})$ is the defect of $\mathbf{u}$.

VIII. Comparison Theorem. *Suppose* $\mathbf{f}(t, \mathbf{y}) : J \times \mathbb{R}^n \to \mathbb{R}^n$ *is quasi-monotone increasing in* $\mathbf{y}$ *and satisfies a Lipschitz condition in the maximum norm* $|\cdot|$ *with* $h(t) \in L(J)$,

$$|\mathbf{f}(t, \mathbf{y}) - \mathbf{f}(t, \mathbf{z})| \le h(t)|\mathbf{y} - \mathbf{z}| \quad \textit{for} \quad \mathbf{y}, \mathbf{z} \in \mathbb{R}^n.$$

Then, for $\mathbf{v}, \mathbf{w} \in AC(J)$,

$$\mathbf{v}(\tau) \le \mathbf{w}(\tau) \ \textit{ and } \ P\mathbf{v} \le P\mathbf{w} \ \textit{ a.e. in } J \ \textit{ implies } \ \mathbf{v} \le \mathbf{w} \ \textit{ in } J.$$

If $v_i(t_1) < w_i(t_1)$ *for an index* i *and* $t_1 \in J$, *then* $v_i < w_i$ *for* $t \ge t_1$.

Proof as an Exercise. Hint: Show that for $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, the difference $\mathbf{f}(t, \mathbf{y}) - \mathbf{f}(t, \mathbf{z})$ can be written in the form $A(\mathbf{y} - \mathbf{x})$ where A is essentially positive and bounded in norm by $h(t)$, and apply VII (use a decomposition of the $\mathbf{f}$-difference as given for $n = 2$ by $g(y_1, y_2) - g(z_1, z_2) = [g(y_1, y_2) - g(z_1, y_2)] + [g(z_1, y_2) - g(z_1, z_2)]$).

§ 17. Systems with Constant Coefficients

I. The Exponential Ansatz. Eigenvalues and Eigenvectors. In this section suppose $A = (a_{ij})$ in the homogeneous linear system

$$\mathbf{y}' = A\mathbf{y} \tag{1}$$

is a constant complex matrix. Solutions can be obtained using the ansatz

$$\mathbf{y}(t) = \mathbf{c} \cdot e^{\lambda t} = \begin{pmatrix} c_1 e^{\lambda t} \\ \vdots \\ c_n e^{\lambda t} \end{pmatrix}, \tag{2}$$

where λ, c_i are complex constants. Substituting $\mathbf{y} = \mathbf{c} \cdot e^{\lambda t}$ into equation (1) leads to

$$\mathbf{y}' = \lambda \mathbf{c} e^{\lambda t} = A\mathbf{c} e^{\lambda t};$$

i.e., $\mathbf{y}(t)$ is a solution of (1) if and only if

$$A\mathbf{c} = \lambda \mathbf{c}. \tag{3}$$

A vector $\mathbf{c} \ne \mathbf{0}$ that satisfies equation (3) is called an *eigenvector* of the matrix A; the number λ is called the *eigenvalue* of A corresponding to $\mathbf{c}$.

We recall a couple of facts from linear algebra. Equation (3), or what amounts to the same thing,

$$(A - \lambda I)\mathbf{c} = \mathbf{0}, \tag{3'}$$

is a linear homogeneous system of equations for $\mathbf{c}$. This system has a nontrivial solution if and only if

$$\det(A-\lambda I) = \begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix} = 0; \tag{4}$$

in other words, the eigenvalues of A are the zeros of the polynomial

$$P_n(\lambda) = \det(A-\lambda I), \tag{5}$$

called the *characteristic polynomial.* This polynomial is of degree n, as one can see, for instance, from the definition (14.1) of a determinant. Thus it has n (real or complex) zeros, where each zero is counted according to its multiplicity. An eigenvector $\mathbf{c} \neq \mathbf{0}$ corresponding to a zero λ (an eigenvector is $\neq \mathbf{0}$ by definition) is obtained by solving the system (3′). It is determined only up to a multiplicative constant. The set $\sigma(A)$ of eigenvalues is called the *spectrum* of A.

II. Theorem (Complex Case). *The function (λ, $\mathbf{c}$, A complex, $\mathbf{c} \neq \mathbf{0}$)*

$$\mathbf{y}(t) = \mathbf{c} \cdot e^{\lambda t}$$

is a solution of equation (1) *if and only if λ is an eigenvalue of the matrix A and $\mathbf{c}$ is a corresponding eigenvector.*

The solutions

$$\mathbf{y}_i(t) = e^{\lambda_i t}\mathbf{c}_i \qquad (i = 1, \ldots, p)$$

are linearly independent if and only if the vectors $\mathbf{c}_i$ are linearly independent. In particular, they are linearly independent if all eigenvalues $\lambda_1, \ldots, \lambda_p$ are distinct.

Thus if A has n linearly independent eigenvectors (this is the case, for example, if A has n distinct eigenvalues), then the system obtained in this manner is a fundamental system of solutions.

Proof. By the isomorphism statement proved in Theorem 15.I, the solutions $\mathbf{y}_i$ are linearly independent if and only if their initial values $\mathbf{y}_i(0) = \mathbf{c}_i$ are linearly independent. The statement that p eigenvectors corresponding to distinct eigenvalues are linearly independent is certainly true for $p = 1$. It is proved in general by establishing the inductive step from p to $p+1$. If the eigenvectors $\mathbf{c}_1, \ldots, \mathbf{c}_p$ are linearly independent and if $\mathbf{c}$ is an additional eigenvector corresponding to the eigenvalue λ and $\lambda \neq \lambda_i$ (here and in the following equations i runs through the numbers 1 to p), then, as we will now show, a representation of the form

$$\mathbf{c} = \sum \alpha_i \mathbf{c}_i$$

is not possible. By applying A to both sides, one would obtain

$$\lambda \mathbf{c} = \sum \alpha_i \lambda_i \mathbf{c}_i,$$

and because such representations are unique,

$$\lambda \alpha_i = \lambda_i \alpha_i, \quad \text{i.e.,} \quad \alpha_i = 0.$$

∎

III. Real Systems. Obviously, the theorem also holds for real systems. In this case, however, one is interested in real solutions. Here one runs into the difficulty that a real matrix may have complex eigenvalues, which lead in turn to complex solutions $\mathbf{y}(t)$. Now, it is immediately obvious that for real A both the real part and the imaginary part of a complex solution are real solutions to (1). Thus from a complex eigenvalue one obtains two real solutions. Note, however, that if the complex quantities λ and $\mathbf{c}$ satisfy equation (3), then their complex conjugates $\bar{\lambda}$ and $\bar{\mathbf{c}}$ do also. Therefore, $\bar{\lambda}$ and $\bar{\mathbf{c}}$ are also an eigenvalue and eigenvector of A and lead to a solution $\bar{\mathbf{y}} = \bar{\mathbf{c}} \cdot e^{\bar{\lambda} t}$, which is the complex conjugate to $\mathbf{y} = \mathbf{c} \cdot e^{\lambda t}$. The decomposition of the complex conjugate solution into real and imaginary parts leads to exactly the same two real solutions.

IV. Theorem (Real Case). *If $\lambda = \mu + i\nu$ $(\nu \neq 0)$ is a complex eigenvalue of the real matrix A and $\mathbf{c} = \mathbf{a} + i\mathbf{b}$ is a corresponding eigenvector, then the complex solution $\mathbf{y} = \mathbf{c}e^{\lambda t}$ produces two real solutions:*

$$\mathbf{u}(t) = \operatorname{Re} \mathbf{y} = e^{\mu t}\{\mathbf{a}\cos \nu t - \mathbf{b}\sin \nu t\},$$
$$\mathbf{v}(t) = \operatorname{Im} \mathbf{y} = e^{\mu t}\{\mathbf{a}\sin \nu t + \mathbf{b}\cos \nu t\}.$$

Suppose there are $2p$ distinct, nonreal eigenvalues

$$\lambda_1, \ldots, \lambda_p;\ \lambda_{p+1} = \bar{\lambda}_1, \ldots, \lambda_{2p} = \bar{\lambda}_p$$

and q distinct real eigenvalues λ_i $(i = 2p+1, \ldots, 2p+q)$. If for the $2p$ distinct, nonreal eigenvalues one constructs $2p$ real solutions

$$\mathbf{u}_i = \operatorname{Re} \mathbf{c}_i e^{\lambda_i t}, \quad \mathbf{v}_i = \operatorname{Im} \mathbf{c}_i e^{\lambda_i t} \qquad (i = 1, \ldots, p)$$

in the manner described above and q real solutions $\mathbf{y}_i$ corresponding to the q distinct real eigenvalues using (2), then the resulting $2p+q$ solutions are linearly independent.

A corresponding result also holds if some of the λ_i are equal, i.e., if there are multiple eigenvalues. If the $2p + q$ corresponding eigenvectors are complex linearly independent, then the same is true for the corresponding $2p + q$ solutions of the form (2), and it remains true for the $2p + q$ real solutions obtained after splitting into real and imaginary parts. In particular, if A has n linearly independent eigenvectors, then one obtains a real fundamental system.

The independence of these real solutions follows from the fact that the original solutions $\mathbf{y}_i = \mathbf{c}_i e^{\lambda_i t}$ ($i = 1, \dots, 2p+q$) are linearly independent by Theorem II and can be represented as linear combinations of the above real solutions; cf. 15.II.(c).

V. Example.

$$\begin{aligned} y_1' &= y_1 - 2y_2 \\ y_2' &= 2y_1 \qquad\quad - y_3 \\ y_3' &= 4y_1 - 2y_2 - y_3 \end{aligned} \qquad A = \begin{pmatrix} 1 & -2 & 0 \\ 2 & 0 & -1 \\ 4 & -2 & -1 \end{pmatrix}.$$

We have

$$P_3(\lambda) = \begin{vmatrix} 1-\lambda & -2 & 0 \\ 2 & -\lambda & -1 \\ 4 & -2 & -1-\lambda \end{vmatrix} = (1-\lambda)(\lambda^2+\lambda+2).$$

The eigenvalues are

$$\lambda_1 = -\frac{1}{2} + i\frac{\sqrt{7}}{2}, \quad \lambda_2 = -\frac{1}{2} - i\frac{\sqrt{7}}{2}, \quad \lambda_3 = 1.$$

The corresponding eigenvectors are solutions to the system (3′). For example, the equation for $\mathbf{c}_1 = (x, y, z)^\top$ is

$$\begin{pmatrix} \frac{3}{2} - i\frac{\sqrt{7}}{2} & -2 & 0 \\ 2 & \frac{1}{2} - i\frac{\sqrt{7}}{2} & -2 \\ 4 & -2 & -\frac{1}{2} - i\frac{\sqrt{7}}{2} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which has a solution

$$\mathbf{c}_1 = \left(\tfrac{3}{2} + i\tfrac{\sqrt{7}}{2}, 2, 4\right)^\top.$$

It follows then that $\mathbf{c}_2 = \bar{\mathbf{c}}_1 = (\frac{3}{2} - i\frac{\sqrt{7}}{2}, 2, 4)^\top$ is also an eigenvector. Another simple calculation gives $\mathbf{c}_3 = (1, 0, 2)^\top$. Taking the solution

$$\mathbf{y}_1(t) = \mathbf{c}_1 \cdot e^{(-\frac{1}{2} + i\frac{\sqrt{7}}{2})t}$$

and separating into real and imaginary parts leads to the two real solutions

$$\mathbf{u}_1(t) = e^{-\frac{1}{2}t} \left[\begin{pmatrix} \frac{3}{2} \\ 2 \\ 4 \end{pmatrix} \cos \tfrac{\sqrt{7}}{2} t - \begin{pmatrix} \frac{\sqrt{7}}{2} \\ 0 \\ 0 \end{pmatrix} \sin \tfrac{\sqrt{7}}{2} t \right],$$

$$\mathbf{v}_1(t) = e^{-\frac{1}{2}t} \left[\begin{pmatrix} \frac{\sqrt{7}}{2} \\ 0 \\ 0 \end{pmatrix} \cos \tfrac{\sqrt{7}}{2} t + \begin{pmatrix} \frac{3}{2} \\ 2 \\ 4 \end{pmatrix} \sin \tfrac{\sqrt{7}}{2} t \right],$$

which, combined with the solution

$$\mathbf{y}_3(t) = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} e^t,$$

constitute a real fundamental system.

VI. Linear Transformations. We consider the results obtained above from a somewhat different point of view. If C is a nonsingular constant matrix, then the mapping

$$\mathbf{y} = C\mathbf{z}, \ \mathbf{z} = C^{-1}\mathbf{y} \qquad (\det C \neq 0) \tag{6}$$

transforms a solution $\mathbf{y}(t)$ of (1) into a solution $\mathbf{z}(t)$ of the system

$$\mathbf{z}' = B\mathbf{z}, \qquad \text{with} \qquad B = C^{-1}AC, \tag{7}$$

and conversely.

Suppose now that A has n linearly independent eigenvectors $\mathbf{c}_1, \ldots, \mathbf{c}_n$. If one sets

$$C = (\mathbf{c}_1, \ldots, \mathbf{c}_n),$$

then

$$AC = (A\mathbf{c}_1, \ldots, A\mathbf{c}_n) = (\lambda_1\mathbf{c}_1, \ldots, \lambda_n\mathbf{c}_n) = CD,$$

where

$$D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \qquad (\text{i.e.,} \quad d_{ii} = \lambda_i, \ d_{ij} = 0 \quad \text{otherwise}),$$

is a diagonal matrix. Thus for this choice of C,

$$B = C^{-1}AC = D,$$

and then (7) reads simply

$$\begin{aligned} z_1' &= \lambda_1 z_1 \\ \vdots \quad & \quad \vdots \\ z_n' &= \lambda_n z_n. \end{aligned}$$

It is easy to find a fundamental system of solutions for this system, namely

$$Z(t) = (\mathbf{z}_1, \ldots, \mathbf{z}_n) = \begin{pmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n t} \end{pmatrix}. \tag{8}$$

Going back to $\mathbf{y} = C\mathbf{z}$, we obtain the fundamental system of Theorem II,

$$Y = CZ = (C\mathbf{z}_1, \ldots, C\mathbf{z}_n) = (\mathbf{c}_1 e^{\lambda_1 t}, \ldots, \mathbf{c}_n e^{\lambda_n t}). \tag{9}$$

Summary. In the case where there are n distinct eigenvalues and, more generally, in the case of n linearly independent eigenvectors there is a fundamental system of solutions of the form (2) (the simplest example, $A = I$ with eigenvectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$, shows that it is also possible to have n linearly independent eigenvectors in the case of multiple zeros of the characteristic polynomial).

VII. Jordan Normal Form of a Matrix. In order to handle the general case, we make use of a result from matrix theory without proof. It says that for every real or complex matrix A there exists a nonsingular matrix C (in general, C will be complex), such that $B = C^{-1}AC$ has the so-called *Jordan normal form*

$$B = \begin{pmatrix} \boxed{J_1} & & & \\ & \boxed{J_2} & & \\ & & \ddots & \\ & & & \boxed{J_k} \end{pmatrix}, \tag{10}$$

where the *Jordan block* J_i is a square matrix of the form

$$J_i = \begin{pmatrix} \lambda_i & 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_i & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \lambda_i & 1 & \ddots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \ddots & \lambda_i & 1 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & \lambda_i & 1 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & \lambda_i \end{pmatrix} \tag{11}$$

with r_i rows and columns; outside of the Jordan blocks, B consists entirely of zeros. Here $r_1 + \cdots + r_k = n$, and

$$P_n(\lambda) = (-1)^n (\lambda - \lambda_1)^{r_1} \cdots (\lambda - \lambda_k)^{r_k}.$$

Note that the main diagonal of B consists of eigenvalues of A and that each block is made up of one and the same eigenvalue. However, the same eigenvalue can appear in more than one block; for example, the matrix I is in Jordan normal form ($k = n$, $r_i = 1$, $\lambda_i = 1$).

The system corresponding to a Jordan block J with r rows and diagonal element λ is given by

$$\mathbf{x}' = J\mathbf{x}, \qquad \text{or} \qquad \left\{ \begin{aligned} x_1' &= \lambda x_1 + x_2 \\ x_2' &= \lambda x_2 + x_3 \\ &\vdots \\ x_{r-1}' &= \lambda x_{r-1} + x_r \\ x_r' &= \lambda x_r \end{aligned} \right. \tag{12}$$

and can be easily solved (one begins with the last equation). The matrix

$$X(t) = \begin{pmatrix} e^{\lambda t} & te^{\lambda t} & \frac{1}{2}t^2 e^{\lambda t} & \cdots & \frac{1}{(r-1)!}t^{r-1}e^{\lambda t} \\ 0 & e^{\lambda t} & te^{\lambda t} & \cdots & \frac{1}{(r-2)!}t^{r-2}e^{\lambda t} \\ 0 & 0 & e^{\lambda t} & \cdots & \frac{1}{(r-3)!}t^{r-3}e^{\lambda t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & e^{\lambda t} \end{pmatrix} \tag{13}$$

is a fundamental matrix for equation (12) with $X(0) = I$.

Proceeding in this way, a fundamental matrix $Z(t)$ for equation (7) can be constructed if B is a Jordan matrix; one has simply to insert the corresponding solution (13) into each Jordan block. For example, if

$$B = \begin{pmatrix} \boxed{\begin{matrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{matrix}} & & \\ & \boxed{\mu} & \\ & & \boxed{\begin{matrix} \nu & 1 \\ 0 & \nu \end{matrix}} \end{pmatrix},$$

then the corresponding fundamental matrix $Z(t)$ with $Z(0) = I$ reads

$$Z(t) = \begin{pmatrix} \boxed{\begin{matrix} e^{\lambda t} & te^{\lambda t} & \frac{1}{2}t^2 e^{\lambda t} \\ 0 & e^{\lambda t} & te^{\lambda t} \\ 0 & 0 & e^{\lambda t} \end{matrix}} & & \\ & \boxed{e^{\mu t}} & \\ & & \boxed{\begin{matrix} e^{\nu t} & te^{\nu t} \\ 0 & e^{\nu t} \end{matrix}} \end{pmatrix}.$$

Thus, if $B = C^{-1}AC$ has Jordan normal form, then each column of $Z(t)$ is a solution of (7) of the form

$$\mathbf{z}(t) = \left(0, \ldots, 0, \frac{t^m}{m!}e^{\lambda t}, \ldots, te^{\lambda t}, e^{\lambda t}, 0, \ldots, 0\right)^\top,$$

where λ is an eigenvalue of A (note that $\sigma(A) = \sigma(B)$). Consequently, $\mathbf{y} = C\mathbf{z}$ is a solution of equation (1) of the form

$$\mathbf{y}(t) = \mathbf{p}_m(t)e^{\lambda t} \qquad \text{with} \qquad \mathbf{p}_m(t) = (p_1^m(t), \ldots, p_n^m(t))^\top,$$

where $p_i^m(t)$ is a polynomial of degree $\leq m$.

VIII. Summary. *For every k-fold zero λ of the characteristic polynomial there exist k linearly independent solutions*

$$\mathbf{y}_1 = \mathbf{p}_0(t)e^{\lambda t}, \ldots, \mathbf{y}_k = \mathbf{p}_{k-1}(t)e^{\lambda t}, \tag{14}$$

in which every component of

$$\mathbf{p}_m(t) = (p_1^m(t), \ldots, p_n^m(t))^\top \qquad (m = 0, 1, \ldots, k-1)$$

is a polynomial of degree $\leq m$. If carried out for every eigenvalue, this construction leads to n solutions, which form a fundamental system.

If A is real, then a real fundamental system is obtained by taking, in case λ is nonreal, two real solutions $\mathbf{u}_i = \operatorname{Re}\mathbf{y}_i$, $\mathbf{v}_i = \operatorname{Im}\mathbf{y}_i$ from each of the k solutions $\mathbf{y}_i$ of (14) *and ignoring the corresponding k solutions for the complex conjugate eigenvalue $\bar{\lambda}$.*

The degree of the polynomials that arise can be determined from the Jordan normal form. In the previous example, where B is a Jordan matrix with $n = 6$, there is a solution $\mathbf{y} = \mathbf{p}(t)e^{\lambda t}$ with degree $\mathbf{p} = 2$, but no solution with higher degree, and this is true even if $\lambda = \mu = \nu$. If $\mu = \nu \neq \lambda$, then there exists a solution $\mathbf{y} = \mathbf{p}(t)e^{\nu t}$ with degree $\mathbf{p} = 1$, but no solution with degree $\mathbf{p} = 2$, etc. The following terminology is useful here.

Algebraic and Geometric Multiplicity. If λ is a k-fold zero of the characteristic polynomial of A, then the number $m(\lambda) := k$ is called the *algebraic multiplicity* of the eigenvalue, and the dimension $m'(\lambda)$ of the corresponding eigenspace, that is, the maximal number of its linearly independent eigenvectors, is called the *geometric multiplicity*. Here $1 \leq m'(\lambda) \leq m(\lambda) \leq n$. If $m(\lambda) = m'(\lambda)$, the eigenvalue is called *semisimple*. In this case, the number λ appears $m(\lambda)$ times in the main diagonal of the Jordan normal form, but there is no 1 in the superdiagonal, and in the corresponding $m(\lambda)$ solutions (14) the $\mathbf{p}_m(t)$ are constant polynomials (namely the eigenvectors). If this is true of all eigenvalues, then the Jordan matrix corresponding to A is a diagonal matrix, and the matrix A is said to be *diagonalizable*.

The calculation of the solutions is easily accomplished once the Jordan normal form $B = C^{-1}AC$ and the transformation matrix C have been determined. However, the $k = m(\lambda)$ solutions belonging to the eigenvalue λ can also be obtained, a step at a time, by first determining the corresponding eigenvectors $\mathbf{c}$ that lead to the solutions $\mathbf{y} = \mathbf{c}e^{\lambda t}$. Then, one after another, the ansätze ($\mathbf{a}$, $\mathbf{b}$, $\ldots \in \mathbb{C}$)

$$\mathbf{y} = (\mathbf{a} + \mathbf{c}t)e^{\lambda t}, \ \mathbf{y} = (\mathbf{a} + \mathbf{b}t + \mathbf{c}t^2)e^{\lambda t}, \ \ldots$$

are applied until $m(\lambda)$ solutions have been found. By equating coefficients of like terms, one sees that the coefficient $\mathbf{c}$ of the highest power of t is always an eigenvector.

IX. Example. $n = 2$, $\mathbf{y}(t) = (x(t), y(t))^\top$,

$$\begin{aligned} x' &= x - y \\ y' &= 4x - 3y \end{aligned} \qquad A = \begin{pmatrix} 1 & -1 \\ 4 & -3 \end{pmatrix}.$$

From

$$\det(A - \lambda I) = \lambda^2 + 2\lambda + 1$$

it follows that $\lambda = -1$ with algebraic multiplicity $m(\lambda) = 2$ and

$$A - \lambda I = A + I = \begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix}.$$

The corresponding homogeneous system (3′) has only one linearly independent solution,

$$\mathbf{c} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Thus we have $m'(\lambda) = 1$. The corresponding solution is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \mathrm{e}^{-t}.$$

A second, linearly independent, solution can be obtained using the ansatz

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a + bt \\ c + dt \end{pmatrix} \mathrm{e}^{-t}.$$

We have that

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} b - a - bt \\ d - c - dt \end{pmatrix} \mathrm{e}^{-t} = A \begin{pmatrix} a + bt \\ c + dt \end{pmatrix} \mathrm{e}^{-t}$$

holds if and only if

$$A \begin{pmatrix} b \\ d \end{pmatrix} = - \begin{pmatrix} b \\ d \end{pmatrix} \quad \text{and} \quad A \begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} b - a \\ d - c \end{pmatrix}.$$

The first equation has the eigenvector $\mathbf{c}$ as a solution, i.e., $b = 1$, $d = 2$. The second equation is satisfied, for example, if $a = 0$, $c = -1$. The corresponding solution

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} t \\ -1 + 2t \end{pmatrix} \mathrm{e}^{-t}$$

is linearly independent from the first solution.

X. Real Systems for $n = 2$. We consider the real system for $\mathbf{y} = (x, y)^\top$

$$\begin{pmatrix} x \\ y \end{pmatrix}' = A \begin{pmatrix} x \\ y \end{pmatrix}, \qquad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \tag{15}$$

under the assumption $D = \det A \neq 0$. This implies that $\lambda = 0$ is not an eigenvalue. The corresponding characteristic polynomial

$$P(\lambda) = \det(A - \lambda I) = \lambda^2 - S\lambda + D \quad \text{with} \quad S = \operatorname{tr} A = a_{11} + a_{22}$$

has zeros

$$\lambda = \frac{1}{2}\left(S - \sqrt{S^2 - 4D}\right), \qquad \mu = \frac{1}{2}\left(S + \sqrt{S^2 - 4D}\right).$$

Real Normal Forms. Our first goal is to show that every real system (15) can be reduced by means of a *real* affine transformation (6), (7) to one of the following normal forms:

$$R(\lambda,\mu) = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}, \quad R_a(\lambda) = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \quad K(\alpha,\omega) = \begin{pmatrix} \alpha & \omega \\ -\omega & \alpha \end{pmatrix}.$$

Here λ, μ, α, ω are real numbers with $\mu \neq 0$ and $\omega > 0$. If $S^2 > 4D$, we have the real case (R). If $S^2 < 4D$, the complex case (K) occurs, while if $S^2 = 4D$, the case (R) or (R_a) occurs depending on whether $\lambda = \mu$ has two linearly independent eigenvectors. We construct now the affine transformation C.

Case (R). There are two (real) eigenvectors $\mathbf{c}$, $\mathbf{d}$ with $A\mathbf{c} = \lambda\mathbf{c}$, $A\mathbf{d} = \mu\mathbf{d}$. If $C = (\mathbf{c}, \mathbf{d})$, then $C^{-1}AC = R(\lambda,\mu)$; cf. VI.

Case (R_a). We have $\lambda = \mu$ and only one eigenvector $\mathbf{c}$. However, as is shown in linear algebra, there is a vector $\mathbf{d}$ linearly independent of $\mathbf{c}$ such that $(A - \lambda I)\mathbf{d} = \mathbf{c}$. The matrix $C = (\mathbf{c}, \mathbf{d})$ again satisfies $C^{-1}AC = R_a(\lambda)$.

Case (K). $\mu = \bar{\lambda}$; hence $A\mathbf{c} = \lambda\mathbf{c}$ and $A\bar{\mathbf{c}} = \bar{\lambda}\bar{\mathbf{c}}$. The matrix $(\mathbf{c}, \bar{\mathbf{c}})$ transforms the system to the normal form $B = \operatorname{diag}(\lambda, \bar{\lambda})$. However, we want to find a *real* normal form. This can be obtained as follows.

Let $\mathbf{c} = \mathbf{a} + \mathrm{i}\mathbf{b}$, $\lambda = \alpha + \mathrm{i}\omega$ with $\omega > 0$. Separating the equation $A\mathbf{c} = \lambda\mathbf{c}$ into real and imaginary parts leads to

$$\left.\begin{aligned} A\mathbf{a} &= \alpha\mathbf{a} - \omega\mathbf{b} \\ A\mathbf{b} &= \alpha\mathbf{b} + \omega\mathbf{a} \end{aligned}\right\} \Leftrightarrow A(\mathbf{a},\mathbf{b}) = (A\mathbf{a}, A\mathbf{b}) = (\mathbf{a},\mathbf{b}) \begin{pmatrix} \alpha & \omega \\ -\omega & \alpha \end{pmatrix}.$$

Since $\mathbf{c}$, $\bar{\mathbf{c}}$ are linearly independent and can be representedin terms of $\mathbf{a}$, $\mathbf{b}$, it follows that $\mathbf{a}$ and $\mathbf{b}$ are also linearly independent; i.e., the matrix $C = (\mathbf{a}, \mathbf{b})$ is regular and transforms the system to the real form $K(\alpha,\omega)$. ∎

We investigate now each of these cases and construct a phase portrait of the differential equation, from which the global behavior of the solutions can be seen. If equations (1) and (7) are coupled by the transformation (6), then their phase portraits are also coupled by the same affine mapping $\mathbf{y} = C\mathbf{z}$ of $\mathbb{R}^n$, which transforms straight lines into straight lines, circles into ellipses, ..., but preserves the characteristic features such as the behavior as $t \to \infty$. In this way we obtain an insight into the global properties of all systems with $\det A \neq 0$.

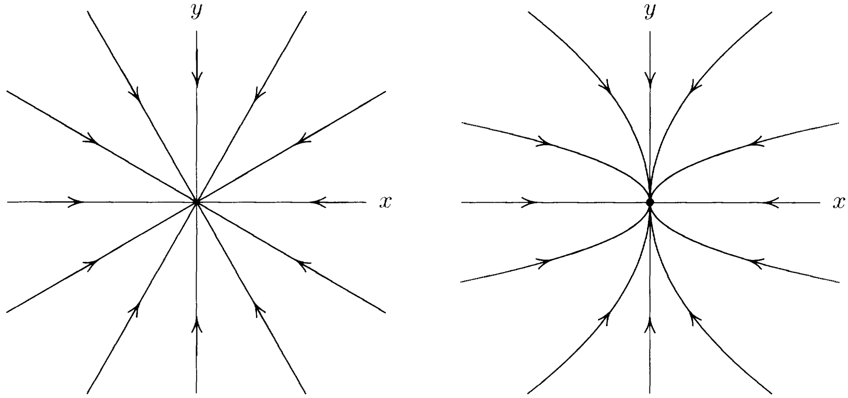
(a) $A = R(\lambda,\mu)$ with $\lambda \leq \mu < 0$. The solutions of the system $x' = \lambda x$, $y' = \mu y$ are given by $(x(t), y(t)) = (a\mathrm{e}^{\lambda t}, b\mathrm{e}^{\mu t})$ (a, b real), their trajectories by

$$\left(\frac{x}{a}\right)^{\mu} = \left(\frac{y}{b}\right)^{\lambda} \qquad (a, b \neq 0 \quad \text{with} \quad x/a,\ y/b > 0).$$

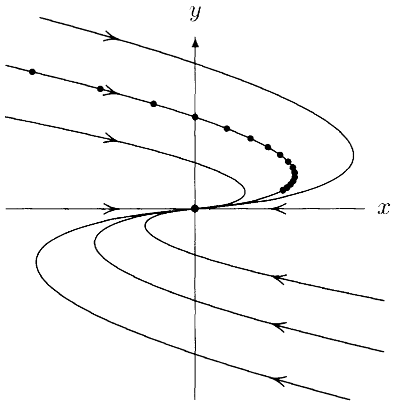
The special cases $a = 0$ or $b = 0$ are simple. All solutions tend to 0 as $t \to \infty$. In the case $\lambda = \mu$, the trajectories are half-lines; in the general case, corresponding power curves. The origin is called a (*stable*) *node*.

(b) $A = R_a(\lambda)$ with $\lambda < 0$. From $x' = \lambda x + y$, $y' = y$, one obtains

$$x(t) = a\mathrm{e}^{\lambda t} + bt\mathrm{e}^{\lambda t}, \qquad y(t) = b\mathrm{e}^{\lambda t}.$$



Stable nodes. $A = R(\lambda, \lambda)$ with $\lambda < 0$ (left) and $A = R(\lambda, \mu)$ with $\lambda < \mu < 0$, $\lambda/\mu = 2$ (right)



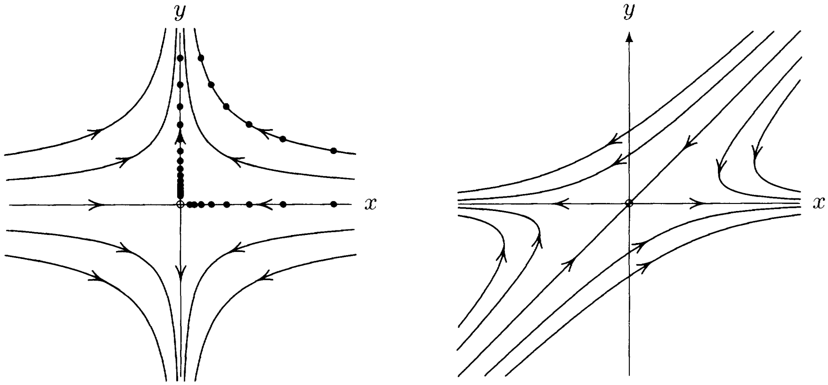
Stable node for $A = R_a(\lambda)$ with $\lambda < 0$

For $a = 0$ (this means that $(x(0), y(0)) = (0, b)$), we have $x = ty$ and $\lambda t = \log(y/b)$. Thus the trajectories are given by

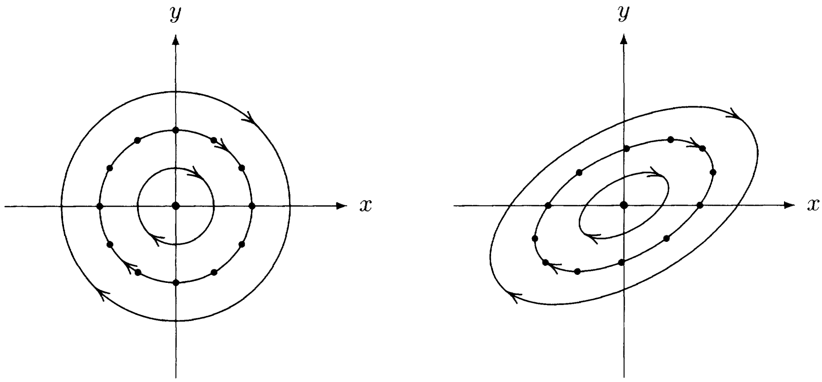
$$\lambda x = y \log\left(\frac{y}{b}\right) \quad \text{for} \quad b \neq 0 \qquad \left(\text{with} \quad \frac{y}{b} > 0\right).$$

The positive and negative x-axis are also trajectories. Here, too, all solutions tend to the origin as $t \to \infty$, which is again called a (*stable*) *node.*

(c) $A = R(\lambda, \mu)$ with $\lambda < 0 < \mu$. The solutions and their trajectories are determined formally as in (a). However, the phase portrait has a completely different appearance. There are only two trajectories that point toward the origin ($b = 0$). All of the other solutions (with $(a, b) \neq \mathbf{0}$) tend to infinity; $x^2(t) + y^2(t) \to \infty$ as $t \to \infty$. The origin is called a *saddle point.*



Saddle point. $A = R(\lambda, \mu)$ with $\lambda < 0 < \mu$, $\lambda/\mu = -2$ (left) and $A = \begin{pmatrix} 1 & -3 \\ 0 & -2 \end{pmatrix}$ (affine distortion, right)



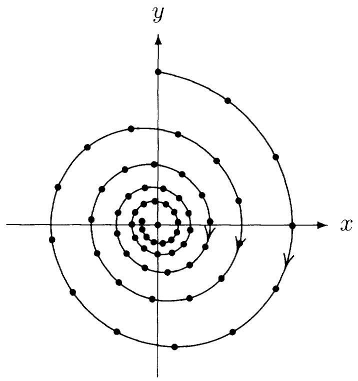
Center for $A = K(0, \omega)$ (left) and $A = \begin{pmatrix} -5 & 13 \\ -7 & 5 \end{pmatrix}$ (affine distortion, right)

(d) $A = K(\alpha, \omega)$ with $\alpha \le 0$. It is easy to check that

$$(x_1, y_1) = e^{\alpha t}(\cos \omega t,\ -\sin \omega t),$$
$$(x_2, y_2) = e^{\alpha t}(\sin \omega t,\ \cos \omega t)$$

are two solutions of (15), from which one can construct a fundamental system $X(t)$ with $X(0) = I$. Using complex notation, in which complex numbers are identified with pairs of real numbers, the above solutions are represented by $z_1(t) = e^{\alpha t}e^{-i\omega t}$ and $z_2(t) = iz_1(t)$. In this notation, the form of the trajectories can be read off.

If $\alpha = 0$, the trajectory is a circle around the origin, which is traced out in the negative sense with circular frequency ω. If $\alpha < 0$, then an additional factor $e^{\alpha t}$, is included and the trajectories are spirals that approach the origin. In the



Stable vortex for $A = K(\alpha, \omega)$ with $\alpha < 0$

case $\alpha = 0$, the origin is called a *center*, in the case $\alpha < 0$ a (*stable*) *vortex*.

(e) *Switching from t to $-t$.* If (x, y) is a solution of (15), then the pair of functions $(\bar{x}(t), \bar{y}(t)) := (x(-t), y(-t))$ is a solution of a related equation, in which A is replaced by $-A$ (and consequently λ, μ are replaced by $-\lambda$, $-\mu$). The functions (x, y) and $(\bar{x}, \bar{y})$ have the same trajectories, only the direction of the arrows is reversed. This takes care of all possible cases.

(f) *Summary.* The following table summarizes the properties of the origin in each of the various cases. See Exercise XII for the case $D = \det A = 0$.

$S^2 \geq 4D$,	$D > 0$,	$S < 0$	stable node	[asymptotically stable]
"	"	$S > 0$	unstable node	[unstable]
"	$D < 0$,	"	saddle point	[unstable]
$S^2 < 4D$,	"	$S < 0$	stable vortex	[asymptotically stable]
"	"	$S = 0$	center	[stable]
"	"	$S > 0$	unstable vortex	[unstable]

The entries in brackets are explained in the next section.

XI. Stability. We consider the homogeneous linear system (15.1)

$$\mathbf{y}' = A(t)\mathbf{y} \quad \text{in} \quad J = [a, \infty)$$

and assume that $A(t)$ is continuous in J. The zero solution $\mathbf{y}(t) \equiv 0$ is called *stable* if all solutions are bounded in J, *asymptotically stable* if every solution tends to $\mathbf{0}$ as $t \to 0$, and *unstable* if there exists a solution that is unbounded in J. If the zero solution is stable and if $X(t)$ denotes the fundamental system with $X(a) = I$, then there exists $K > 0$ such that $|X(t)| \leq K$ for $t \geq a$. If $\mathbf{y}$ is the solution with initial value $\mathbf{y}(a) = \mathbf{c}$, then $|\mathbf{y}(t)| = |X(t)\mathbf{c}| \leq K|\mathbf{c}|$. Thus stability means, roughly speaking, that solutions with small initial values remain small for all future time. The differential equation $\mathbf{y}' = A(t)\mathbf{y}$ is also sometimes called stable, asymptotically stable, or unstable, when the zero solution is stable, asymptotically stable, or unstable, respectively.

If the matrix A is constant, then the Summary VIII gives us a complete description of the stability behavior of the zero solution. The zero solution is

asymptotically stable	if $\operatorname{Re}\lambda < 0$ for all $\lambda \in \sigma(A)$
stable	if $\operatorname{Re} \leq 0$ for $\lambda \in \sigma(A)$ and if $m'(\lambda) = m(\lambda)$ for all eigenvalues λ with $\operatorname{Re}\lambda = 0$
unstable	in all other cases, i.e., if there exists a $\lambda \in \sigma(A)$ with $\operatorname{Re}\lambda > 0$ or with $\operatorname{Re}\lambda = 0$ and $m'(\lambda) < m(\lambda)$

In the first two cases, there exists a fundamental matrix $Y(t)$ by VIII, which tends to 0 as $t \to 0$ or remains bounded, respectively. The same holds then for an arbitrary solution, since every solution can be represented in the form $\mathbf{y}(t) = Y(t)Y^{-1}(a)\mathbf{y}(a)$. In the third case, there exists an unbounded solution $\mathbf{y} = \mathbf{c}e^{\lambda t}$ with $\operatorname{Re}\lambda > 0$ or $\mathbf{y} = \mathbf{p}(t)e^{i\omega t}$ with real ω and degree $\mathbf{p} \geq 1$.

Exercise. Let $A(t)$ be a complex matrix satisfying

$$\operatorname{Re}\langle A(t)\mathbf{y}, \mathbf{y}\rangle \leq \gamma(t)|\mathbf{y}|_e^2 \text{ for } t \geq a \text{ and } \mathbf{y} \in \mathbb{C}^n,$$

where $\langle\cdot,\cdot\rangle$ is the usual scalar product in $\mathbb{C}^n$; cf. 28.II.(a). Let $h(t) = \int_a^t \gamma(s)\,ds$. Show that if $h(t)$ is bounded, the zero solution is stable, and asymptotically stable if $h(t) \to -\infty$ as $t \to \infty$ (holds also for real matrices).

Hint: Derive the inequality $\phi' \leq 2\gamma(t)\phi$ for $\phi(t) = |\mathbf{y}|_e^2$.

XII. Exercise. Investigate the two-dimensional linear system (15) in the case $D = \det A = 0$.

(a) Determine the normal forms that arise.

(b) Solve the corresponding systems, determine the critical points, and sketch their phase portraits.

(c) Solve the system $x' = 2x - 4y$, $y' = -x + 2y$ (include a phase portrait).

XIII. Exercises. Determine a real fundamental system of solutions for the following systems:

$$(a)\quad \begin{aligned} x' &= 3x + 6y, \\ y' &= -2x - 3y. \end{aligned} \qquad\qquad (b)\quad \begin{aligned} x' &= 8x + y, \\ y' &= -4x + 4y. \end{aligned}$$

$$(c)\quad \begin{aligned} x' &= x - y + 2z, \\ y' &= -x + y + 2z, \\ z' &= x + y. \end{aligned} \qquad\qquad (d)\quad \begin{aligned} x' &= -x + y - z, \\ y' &= 2x - y + 2z, \\ z' &= 2x + 2y - z. \end{aligned}$$

§ 18. Matrix Functions. Inhomogeneous Systems

I. Power Series of Matrices. In this section, the constants and matrices can be complex. If B is an $n \times n$ matrix and $p(s)$ is the polynomial

$$p(s) = c_0 + c_1 s + \cdots + c_k s^k, \tag{1}$$

then $p(B)$ is defined to be the matrix

$$p(B) = c_0 I + c_1 B + \cdots + c_k B^k. \tag{2}$$

In particular, for $B = At$ (i.e., $b_{ij} = a_{ij}t$),

$$p(At) = c_0 I + c_1 At + \cdots + c_k A^k t^k.$$

The derivative of this matrix with respect to t is given by

$$\frac{d}{dt} p(At) = Ap'(At), \tag{3}$$

where $p'(s)$ is the derivative of $p(s)$ (note the analogy to the chain rule).

We now consider infinite series of $n \times n$ matrices C_k,

$$C = \sum_{k=0}^{\infty} C_k.$$

Convergence is defined as usual: $S_p = C_0 + \cdots + C_p \to C$ as $p \to \infty$, i.e., $|S_p - C| \to 0$, where $|\cdot|$ is a compatible matrix norm; cf. 14.II. This is true if and only if with the notation $C_k = (c_{ij}^{(k)})$, each of the n^2 series $\sum_k c_{ij}^{(k)}$ converges to c_{ij}. We say that the matrix series is absolutely convergent if the real series $\sum |C_k|$ converges, or, equivalently, if each of the n^2 scalar series is absolutely convergent. This equivalence follows from the fact (10.III) that for each norm there are constants $\alpha, \beta > 0$ such that

$$\alpha |b_{ij}| \le |B| \le \beta \sum |b_{ij}|. \tag{4}$$

In particular, every power series

$$f(s) = \sum_{k=0}^{\infty} c_k s^k \qquad (|s| < r) \tag{1'}$$

with radius of convergence $r > 0$ generates a matrix function

$$f(B) = \sum_{k=0}^{\infty} c_k B^k \quad \text{(absolutely convergent for } |B| < r). \tag{2'}$$

To be precise, if $|B| =: s < r$, then, since $|BC| \le |B||C|$ by (14.2),

$$|B^2| \le |B|^2 = s^2, \ldots, |B^k| \le s^k,$$

and hence $\sum |c_k B^k| < \infty$. Thus

$$f(At) = c_0 I + c_1 At + c_2 A^2 t^2 + \cdots$$

is absolutely convergent for

$$|t| < \frac{r}{|A|} = t_0$$

and uniformly convergent in every compact subinterval of $(-t_0, t_0)$. Since the formally differentiated series is again uniformly convergent, one can differentiate $f(At)$ term by term and obtain, similar to (3), the formula

$$\frac{d}{dt} f(At) = Af'(At). \tag{3'}$$

II. Example. The Exponential Function. The series

$$\mathrm{e}^B = I + B + \frac{B^2}{2!} + \frac{B^3}{3!} + \cdots$$

converges absolutely for all B. Here we have

$$(\mathrm{e}^{At})' = A\mathrm{e}^{At} \tag{3''}$$

by (3′). This is a really surprising result: We have found a fundamental matrix for the linear system

$$\mathbf{y}' = A\mathbf{y} \tag{5}$$

in a second way, independent of §17, namely

$$X(t) = \mathrm{e}^{At} \quad \text{with} \quad X(0) = I. \tag{6}$$

Formally this result agrees completely with the one-dimensional case: The solution of $x' = ax$, $x(0) = 1$ is $x(t) = \mathrm{e}^{at}$. Additionally, we note that one obtains the series for $X(t)$ by writing the initial value problem for $X(t)$ as a matrix integral equation

$$X(t) = I + \int_0^t AX(s)\, ds$$

and applying the method of successive approximations

$$X_0 = I,\ X_{k+1} = I + \int_0^t AX_k(s)\, ds \qquad (k = 0, 1, 2, \ldots).$$

A simple calculation shows that

$$X_k(t) = I + At + \cdots + \frac{1}{k!} A^k t^k,$$

which is the kth partial sum of the series for e^{At}.

We are going to derive some properties of the exponential function.

III. Lemma. *We have*

(a) $e^{B+C} = e^B \cdot e^C$ *if* $BC = CB$;

(b) $e^{C^{-1}BC} = C^{-1}e^B C$ *if* $\det C \neq 0$;

(c) $e^{\operatorname{diag}(\lambda_1,\dots,\lambda_n)} = \operatorname{diag}(e^{\lambda_1},\dots,e^{\lambda_n})$,

where $D = \operatorname{diag}(\mu_1,\dots,\mu_n)$ *means that* $d_{ii} = \mu_i$, *and* $d_{ij} = 0$ *for* $i \neq j$. *Proposition* (a) *is the addition theorem; it does not hold in general.*

Proof. Because of the absolute convergence of the series for e^B and e^C, these series can be multiplied termwise (this actually involves the termwise multiplication of n^2 scalar series), and one obtains

$$\begin{aligned} e^{B+C} &= \sum_{n=0}^{\infty} \frac{(B+C)^n}{n!} = \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{B^k C^{n-k}}{k!(n-k)!} \\ &= \sum_{p=0}^{\infty} \frac{B^p}{p!} \cdot \sum_{q=0}^{\infty} \frac{C^q}{q!} = e^B \cdot e^C. \end{aligned}$$

This proves (a). To prove (b), it can be shown by induction that

$$(C^{-1}BC)^k = C^{-1}B^k C \qquad (k = 0, 1, 2, \dots)$$

and hence

$$\sum_{k=0}^{n} \frac{1}{k!}(C^{-1}BC)^k = C^{-1}\left(\sum_{k=0}^{n}\frac{1}{k!}B^k\right)C \qquad (n = 0, 1, 2, \dots).$$

The assertion follows from this relation by taking limits as $n \to \infty$.

It can also be shown by induction that

$$(\operatorname{diag}(\lambda_1,\dots,\lambda_n))^k = \operatorname{diag}(\lambda_1^k,\dots,\lambda_n^k).$$

The final assertion in the lemma now follows after multiplication by $1/k!$ and then forming the infinite series. ∎

Remark. The addition theorem can also be derived from property (3″). According to this property, $U(t) := e^{(B+C)t}$ is a solution of

$$U' = (B+C)U \quad \text{with} \quad U(0) = I. \tag{7}$$

It follows from the product rule 14.IV that $V(t) := e^{Bt} \cdot e^{Ct}$ also satisfies (7),

$$V' = Be^{Bt}e^{Ct} + e^{Bt}Ce^{Ct} = (B+C)V,$$

since $Ce^{Bt} = e^{Bt}C$. Hence $U = V$ by uniqueness.

We note some simple consequences of the addition theorem.

IV. Corollaries. *For an arbitrary square matrix A,*

(a) $(e^A)^{-1} = e^{-A}$;

(b) $e^{A(s+t)} = e^{As} \cdot e^{At}$;

(c) $e^{A+\lambda I} = e^{\lambda} \cdot e^{A}$.

V. Remark. If the matrix A has the form of a Jordan block (17.11), then the fundamental matrix $X(t)$ with $X(0) = I$ can be explicitly given; cf. (17.13). It is also given by e^{Jt}, where J is the $r \times r$ matrix from (17.12). The uniqueness theorem guarantees that e^{Jt} actually has the form (17.13); however, this result can also be explicitly verified without difficulty. We briefly indicate the necessary steps. If $F = (f_{ij})$ is the $r \times r$ matrix with elements $f_{i,i+1} = 1$, $f_{ij} = 0$ otherwise, then it is easy to check that

$$\begin{array}{lll} G = F^2 & \text{is given by} & g_{i,i+2} = 1,\ g_{ij} = 0 \text{ otherwise}, \\ H = F^3 & \text{is given by} & h_{i,i+3} = 1,\ h_{ij} = 0 \text{ otherwise}, \\ \vdots & \vdots & \vdots \end{array}$$

In particular, $F^k = 0$ for $k \geq r$. It follows that

$$e^{Ft} = \begin{pmatrix} 1 & t & \frac{1}{2!}t^2 & \cdots & \frac{1}{(r-1)!}t^{r-1} \\ 0 & 1 & t & \cdots & \frac{1}{(r-2)!}t^{r-2} \\ 0 & 0 & 1 & \cdots & \frac{1}{(r-3)!}t^{r-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}; \tag{8}$$

hence, if $J = \lambda I + F$, then by IV.(c),

$$e^{Jt} = e^{\lambda t} e^{Ft}. \tag{9}$$

This is precisely the matrix given in (17.13).

VI. Inhomogeneous Systems. By (16.4), the initial value problem

$$\mathbf{y}' = A\mathbf{y} + \mathbf{b}(t), \quad \mathbf{y}(\tau) = \boldsymbol{\eta} \qquad (A \text{ constant}) \tag{10}$$

has the solution

$$\mathbf{y}(t) = e^{A(t-\tau)}\boldsymbol{\eta} + \int_\tau^t e^{A(t-s)}\mathbf{b}(s)\, ds. \tag{11}$$

This is true because $X(t) = e^{A(t-\tau)}$ is the fundamental matrix with $X(\tau) = I$ and, by IV.(a),

$$(X(t))^{-1} = e^{-A(t-\tau)}. \tag{12}$$

VII. Exercise. Power Series. (a) If A is a constant $n \times n$ matrix, then $\sin A$ and $\cos A$ are defined by the power series

$$\sin A := \sum_{k=0}^{\infty} (-1)^k \frac{A^{2k+1}}{(2k+1)!}, \qquad \cos A := \sum_{k=0}^{\infty} (-1)^k \frac{A^{2k}}{(2k)!}.$$

Prove the Euler formulas

$$\mathrm{e}^{\mathrm{i}A} = \cos A + \mathrm{i} \cdot \sin A$$

$$\cos A = \frac{1}{2}(\mathrm{e}^{\mathrm{i}A} + \mathrm{e}^{-\mathrm{i}A}), \qquad \sin A = \frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}A} - \mathrm{e}^{-\mathrm{i}A})$$

and the addition theorems

$$\cos(A+B) = \cos A \cos B - \sin A \sin B,$$

$$\sin(A+B) = \sin A \cos B + \cos A \sin B,$$

under the assumptions that A and B commute ($AB = BA$).

Show that the functions $\cos At$ and $\sin At$ can be defined by initial value problems:

$$Y'' + A^2 Y = 0,\ Y(0) = I,\ Y'(0) = 0 \iff Y(t) = \cos At,$$

$$Y'' + A^2 Y = 0,\ Y(0) = 0,\ Y'(0) = A \iff Y(t) = \sin At.$$

For $n = 1$ this reduces to well-known properties of $\cos at$ and $\sin at$.

(b) Let $f(t) = \sum f_i t^i$, $g(t) = \sum g_i t^i$, and $h(t) = \sum h_i t^i$ (i runs from 0 to ∞) be power series with positive radii of convergence r_f, r_g, and $r_h \geq r_f + r_g$. Assume that $f(s)g(t) = h(s+t)$ for $|s| < r_f$, $|t| < r_g$. Show that if $|A| < r_f$, $|B| < r_g$, and $AB = BA$, then $f(A)g(B) = h(A+B)$. Here $|A|$ is an operator norm generated from an arbitrary vector norm in $\mathbb{R}^n$ (or $\mathbb{C}^n$).

VIII. Exercise. Special Inhomogeneous Systems. In the following $\mathbf{p}(t)$ and $\mathbf{q}(t)$ are vector polynomials. Show:

(a) If $\alpha \notin \sigma(A)$, then the differential equation

$$\mathbf{y}' = A\mathbf{y} + \mathbf{c}\mathrm{e}^{\alpha t} \qquad (\mathbf{c} \in \mathbb{C}^n) \tag{13}$$

has exactly one solution of the form $\mathbf{y} = \mathbf{d}\mathrm{e}^{\alpha t}$, $\mathbf{d} \in \mathbb{C}^n$.

More generally, the differential equation

$$\mathbf{y}' = A\mathbf{y} + \mathbf{p}(t)\mathrm{e}^{\alpha t} \tag{14}$$

has exactly one solution of the form $\mathbf{y} = \mathbf{q}(t)\mathrm{e}^{\alpha t}$, and degree $\mathbf{p}$ = degree $\mathbf{q}$.

In particular, the differential equation

$$\mathbf{y}' = A\mathbf{y} + \mathbf{c}$$

has exactly one constant solution if $\det A \neq 0$.

(b) If A, α, and $\mathbf{p}(t)$ are real-valued and if $\mathrm{i}\alpha \notin \sigma(A)$, then the equation

$$\mathbf{y}' = A\mathbf{y} + \mathbf{p}(t)\cos\alpha t \tag{15}$$

has exactly one solution of the form $\mathbf{y}(t) = \mathbf{q}_1(t)\cos\alpha t + \mathbf{q}_2(t)\sin\alpha t$ with real polynomials $\mathbf{q}_1$, $\mathbf{q}_2$, and $\operatorname{degree}\mathbf{p} = \max\{\operatorname{degree}\mathbf{q}_1, \operatorname{degree}\mathbf{q}_2\}$.

(c) Also in the case $\alpha \in \sigma(A)$ the differential equation (14) has a solution of the form $\mathbf{y} = \mathbf{q}(t)\mathrm{e}^{\alpha t}$, but we can say only that $\operatorname{degree}\mathbf{q} \le \operatorname{degree}\mathbf{p} + m(\alpha)$, where $m(\alpha)$ is the algebraic multiplicity of the eigenvalue α.

Hint: For (b) consider the equation $\mathbf{y}' = A\mathbf{y} + \mathbf{p}(t)\mathrm{e}^{\mathrm{i}\alpha t}$; for (c) transform first into Jordan normal form and prove the statement for $A = J$.

Supplement: Floquet Theory

We deal with linear systems with periodic coefficients. The following theory goes back to the French mathematician Gaston Floquet (1847–1920). A main result states that systems with periodic coefficients can be reduced to systems with constant coefficients (at least in principle).

IX. Homogeneous Systems with Periodic Coefficients. Let $\omega > 0$. A function f is called ω-periodic if f is defined in $\mathbb{R}$ and satisfies the equation $f(t+\omega) = f(t)$. We consider systems with a continuous, ω-periodic (real or complex) coefficient matrix,

$$\mathbf{y}' = A(t)\mathbf{y} \quad \text{with} \quad A(t+\omega) = A(t). \tag{16}$$

In the following, the term *solution* will always refer to solutions of (16). Every solution exists in $\mathbb{R}$.

(a) If $\mathbf{y}(t)$ is a solution, then so is $\mathbf{z}(t) = \mathbf{y}(t+\omega)$.

(b) If $\mathbf{y}$ is a solution and $\mathbf{y}(\omega) = \lambda\mathbf{y}(0)$ ($\lambda \in \mathbb{R}$ or $\mathbb{C}$), then it follows that $\mathbf{y}(t+\omega) = \lambda\mathbf{y}(t)$ and more generally $\mathbf{y}(t+k\omega) = \lambda^k\mathbf{y}(t)$ for all t (k an integer).

The *proof* of (a) is elementary; (b) follows for $k = 1$ from the observation that $\lambda\mathbf{y}(t)$ and $\mathbf{y}(t+\omega)$ both satisfy the same initial condition at $t = 0$, the result for $k > 1$ is obtained by induction and for $k < 0$ using the change of independent variable $t' = t + k\omega$. ∎

Let $X(t)$ be the fundamental matrix for (16) with $X(0) = I$. Then by (a) and Proposition 15.III, $Z(t) = X(t+\omega)$ is also a fundamental matrix, and from 15.II.(h) it follows that

$$X(t+\omega) = X(t)C \quad \text{with nonsingular} \quad C = X(\omega). \tag{17}$$

The *transition matrix* C will play a decisive role in the following. Its eigenvalues λ_i are called *characteristic* (or *Floquet*) *multipliers*. Since C is nonsingular, they are nonzero, and there exist numbers $\mu_i \in \mathbb{C}$ with $\lambda_i = \mathrm{e}^{\omega\mu_i}$. The μ_i are called *characteristic exponents*. They are determined only up to a multiple of $2\pi\mathrm{i}/\omega$ (because $\mathrm{e}^{2\pi\mathrm{i}} = 1$). However, $\operatorname{Re}\mu_i$ is uniquely determined.

Since an arbitrary solution can be represented in the form $\mathbf{y}(t) = X(t)\mathbf{y}(0)$, the relation $\mathbf{y}(\omega) = C\mathbf{y}(0)$ holds. Thus the equation $\mathbf{y}(\omega) = \lambda\mathbf{y}(0)$ is equivalent to $C\mathbf{y}(0) = \lambda\mathbf{y}(0)$, and from (b) one obtains the following theorem.

Theorem. *There exist nontrival solutions to* (16) *satisfying* $\mathbf{y}(t+\omega) = \lambda\mathbf{y}(t)$ *if and only if* λ *is an eigenvalue of* C. *Every such solution has the form* $\mathbf{y} = X(t)\mathbf{c}$, *where* $\mathbf{c}$ *is an eigenvector of* C *corresponding to* λ. *If the matrix* C *is diagonalizable, then one obtains a fundamental system of solutions in this manner.*

Therefore, there exist nontrivial ω*-periodic solutions if and only if* $\lambda = 1$ *is an eigenvalue of* C, *and periodic solutions with minimal period* $k\omega > 0$ $(k \in \mathbb{N})$ *if and only if* C *has an eigenvalue* λ *satisfying* $\lambda^k = 1$ *and* $\lambda^j \neq 1$ *for* $1 \le j < k$.

We need a result that will not be proved until 22.VI: For a nonsingular matrix C, there exists a matrix B with $C = \mathrm{e}^{\omega B}$ (in general, B is complex, even for real C). Because $\mathrm{e}^{2k\pi\mathrm{i}} = 1$, the matrix B is not uniquely determined; for example, one can add $(2k\pi\mathrm{i}/\omega)I$.

X. Theorem of Floquet. *The fundamental matrix* $X(t)$ *of* (16) *with* $X(0) = I$ *has a Floquet representation*

$$X(t) = Q(t)\mathrm{e}^{Bt} \qquad \textit{Floquet representation,} \tag{18}$$

where $Q \in C^1(\mathbb{R})$ *is* ω*-periodic and* B *satisfies the equation* $C = X(\omega) = \mathrm{e}^{\omega B}$. *Clearly,* $Q(0) = I$, *and* $Q(t)$ *is a nonsingular matrix for all* t.

Proof. We define Q by (18), i.e., $Q(t) = X(t)\mathrm{e}^{-Bt}$. Then

$$X(t+\omega) = Q(t+\omega)\mathrm{e}^{B(t+\omega)}.$$

On the other hand, $X(t+\omega) = X(t)C = Q(t)\mathrm{e}^{Bt}C = Q(t)\mathrm{e}^{B(t+\omega)}$. The assertion $Q(t) = Q(t+\omega)$ follows (after multiplication by $\mathrm{e}^{-B(t+\omega)}$) by comparing these two results. ∎

Our analysis of the Floquet representation uses the following lemma.

Lemma. *From the Jordan normal form* V *of the matrix* U *one obtains the Jordan normal form of* e^U *by replacing the diagonal elements* λ_i *of* V *by* e^{λ_i} *(the corresponding Jordan blocks are thus of the same size). An eigenvalue* λ *of* U *has the same algebraic and geometric multiplicity and the same eigenvectors as the corresponding eigenvalue* e^λ *of* e^U.

Proof. If $V = D^{-1}UD$ (D nonsingular), then by III.(b), $\mathrm{e}^V = D^{-1}\mathrm{e}^U D$. In our investigation of the matrix e^V we can confine ourselves to a single Jordan block $J = \lambda I + F$ with r columns (cf. V for the notation). We show that for $\mathbf{x} \in \mathbb{C}^n$,

$$\begin{aligned} J\mathbf{x} &= \lambda\mathbf{x} \iff F\mathbf{x} = 0 \iff \mathbf{x} = \alpha\mathbf{e}_1, \\ \mathrm{e}^J\mathbf{x} &= \mathrm{e}^\lambda\mathbf{x} \iff (\mathrm{e}^F - I)\mathbf{x} = 0 \iff \mathbf{x} = \alpha\mathbf{e}_1, \end{aligned}$$

where $\mathbf{e}_1 = (1, 0, \ldots, 0)$. The first line is easily established, and by (8) the matrix $B = e^F - I$ satisfies $b_{ij} = 0$ for $j \leq i$, $b_{i,i+1} = 1$, which leads to the same result $\mathbf{x} = \alpha \mathbf{e}_1$ (the reader should consider, e.g., the case $r = 4$). The matrix J has only one eigenvalue λ, and the matrix e^J has only one eigenvalue e^λ, and both matrices have $\mathbf{e}_1$ as the only eigenvector. Hence $e^\lambda I + F$ is the Jordan normal form of the matrix e^J. If one observes, in addition, that from a common eigenvector $\mathbf{x}$ of V and e^V, a common eigenvector $\mathbf{c} = D\mathbf{x}$ of U and e^U is obtained, then all assertions of the lemma are proved. ∎

We apply the lemma to the matrix $U = \omega B$. If μ_i are the eigenvalues of B, then $\omega \mu_i$ are the eigenvalues of U and $\lambda_i = e^{\omega \mu_i}$ are the eigenvalues of $C = e^{\omega B}$ $(i = 1, \ldots, n)$; i.e., the μ_i are *characteristic exponents.*

Suppose both sides of (18) are multiplied on the right by a nonsingular matrix D. Then the matrix $Y(t) = X(t)D$ on the left is a fundamental matrix of (16), and the matrix $Z(t) = e^{Bt}D$ on the right is a fundamental matrix of the equation $\mathbf{z}' = B\mathbf{z}$ (both with the initial value D at $t = 0$). The resulting equation $Y(t) = Q(t)Z(t)$ shows how the fundamental solutions of (16) are obtained from those of $\mathbf{z}' = B\mathbf{z}$. The summary given in 17.VIII (with B instead of A) leads then to the following

XI. Summary. An eigenvalue $\lambda = e^{\omega \mu}$ of C corresponds to an eigenvalue μ of B, and both have the same algebraic multiplicity k. Moreover, there exist k linearly independent solutions

$$\mathbf{y} = Q(t)\mathbf{p}_m(t)e^{\mu t} \qquad (m = 0, 1, \ldots, k-1),$$

where $\mathbf{p}_m(t)$ is a vector polynomial of degree $\leq m$. The function

$$\mathbf{q}_m(t) = Q(t)\mathbf{p}_m(t) = \mathbf{c}_0(t) + \mathbf{c}_1(t)t + \cdots + \mathbf{c}_m(t)t^m$$

is a "polynomial with ω-periodic coefficients" $\mathbf{c}_j$. This construction, carried out for all characteristic exponents μ_i, leads to a fundamental system of solutions to equation (16).

Stability. Since there exist positive constants α, β with $\alpha \leq |Q(t)| \leq \beta$, the stability analysis from 17.XI carries over to equation (16) with λ replaced by μ. Thus the zero solution of equation (16) is

asymptotically stable	if $\|\lambda\| < 1$ for all $\lambda \in \sigma(C)$
stable	if $\|\lambda\| \leq 1$ for all $\lambda \in \sigma(C)$ and the eigenvalues λ with $\|\lambda\| = 1$ are semisimple
unstable	in all other cases.

Note that the condition $|\lambda| < 1$ or ≤ 1 or > 1 for the characteristic exponents is equivalent to $\operatorname{Re} \mu < 0$ or ≤ 0 or > 0.

The fact that the characteristic exponents are not uniquely determined plays no role here. From $\lambda = e^{\omega\mu'}$ it follows that $\mu = \mu' + 2k\pi i/\omega$ (k an integer). If μ is replaced by μ' in $e^{\mu t}$, then an ω-periodic factor $e^{(2k\pi i/\omega)t}$ appears, which can be incorporated into the term $\mathbf{q}_m(t)$.

XII. The Inhomogeneous System. The system

$$\mathbf{y}' = A(t)\mathbf{y} + \mathbf{b}(t) \tag{19}$$

is now considered under the assumption that $A(t)$ and $\mathbf{b}(t)$ are continuous and ω-periodic. The following theorem clarifies the relationship of (19) with the system

$$\mathbf{z}' = B\mathbf{z} + \mathbf{c}(t) \quad \text{with} \quad \mathbf{c}(t) = Q^{-1}(t)\mathbf{b}(t). \tag{20}$$

Theorem. *The solutions* $\mathbf{y}$ *of equation* (19) *and* $\mathbf{z}$ *of equation* (20) *with the same initial value* $\mathbf{y}(0) = \mathbf{z}(0) = \boldsymbol{\eta}$ *are coupled by the relation* $\mathbf{y}(t) = Q(t)\mathbf{z}(t)$ *(equivalently,* $\mathbf{z}(t) = Q(t)^{-1}\mathbf{y}(t)$*).*

Proof. From the Floquet representation $X = Qe^{Bt}$ it follows that $X' = (Q' + QB)e^{Bt} = AX = AQe^{Bt}$, hence

$$Q' + QB = AQ. \tag{$*$}$$

Let $\mathbf{y}$ be a solution of (19) and $\mathbf{z}$ be defined by $\mathbf{y} = Q\mathbf{z}$. Then $\mathbf{y}' = Q'\mathbf{z} + Q\mathbf{z}'$ and $\mathbf{y}' = A\mathbf{y} + \mathbf{b}$, from which

$$Q'\mathbf{z} + Q\mathbf{z}' = AQ\mathbf{z} + \mathbf{b} = Q'\mathbf{z} + QB\mathbf{z} + \mathbf{b}$$

follows because of $(*)$. By multiplication on the left by Q^{-1} one obtains (20). The reverse direction is proved similarly. ∎

§ 19. Linear Differential Equations of Order n

A linear differential equation of order n

$$Lu := u^{(n)} + a_{n-1}(t)u^{(n-1)} + \cdots + a_0(t)u = b(t) \tag{1}$$

is equivalent to the system

$$\begin{aligned} y_1' &= y_2 \\ \vdots\ \ & \quad \vdots \\ y_{n-1}' &= y_n \\ y_n' &= -(a_0 y_1 + \cdots + a_{n-1}y_n) + b(t); \end{aligned} \tag{2}$$

cf. 11.I. This can be written in the form

$$\mathbf{y}' = A(t)\mathbf{y} + \mathbf{b}(t), \tag{2'}$$

where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} u \\ u' \\ \vdots \\ u^{(n-2)} \\ u^{(n-1)} \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b(t) \end{pmatrix},$$

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-2} & -a_{n-1} \end{pmatrix}.$$

On the basis of Theorem 14.VI, we have the following theorem.

I. Existence and Uniqueness Theorem. *If the real- or complex-valued coefficients $a_i(t)$, $b(t)$ ($i = 1, \ldots, n$, $a_n \equiv 1$) are continuous in an interval J and if $\tau \in J$, then the initial value problem*

$$Lu \equiv \sum_{i=0}^{n} a_i(t) u^{(i)}(t) = b(t), \; u^{(\nu)}(\tau) = \eta_\nu, \quad (\nu = 0, 1, \ldots, n-1) \tag{3}$$

has exactly one solution. The solution exists in all of J and depends continuously on η_ν, and on $a_i(t)$, $b(t)$ in each compact subinterval of J.

II. The Homogeneous Differential Equation $Lu = 0$. *If the coefficients $a_i(t)$ are real (or complex), then the real (complex) solutions of the homogeneous differential equation form an n-dimensional vector space over the field of real (complex) numbers.*

Each vector $(\eta_0, \eta_1, \ldots, \eta_{n-1}) \in \mathbb{R}^n$ or $\mathbb{C}^n$ is associated with a solution satisfying the initial conditions in (3), and this mapping is again a linear isomorphism; cf. 15.I. Thus there exist n linearly independent solutions

$$u_1(t), \ldots, u_n(t), \tag{4}$$

and they form a *fundamental system.* Every solution is a linear combination of the u_i.

In going from (1) to (2), a solution $u(t)$ of (1) is associated with the vector $\mathbf{y}(t) = (u(t), u'(t), \ldots, u^{n-1}(t))^\top$, which is a solution of the corresponding system (2). Thus the Wronskian of the n solutions (4) is the determinant

$$W(t) = \begin{vmatrix} u_1 & \cdots & u_n \\ u_1' & \cdots & u_n' \\ \vdots & & \vdots \\ u_1^{(n-1)} & \cdots & u_n^{(n-1)} \end{vmatrix}.$$

By (15.6), (15.7), the Wronskian satisfies the equation

$$W' = -a_{n-1}W;$$

hence

$$W(t) = W(\tau)\mathrm{e}^{-\int_\tau^t a_{n-1}(s)\,ds}. \tag{5}$$

In §15 we constructed a special fundamental system $X(t)$ with $X(t) = I$. Here that system corresponds to a fundamental system $u_1, \ldots, u_n$ of (3), where

$$Lu_i = 0, \quad u_i^{(j)}(\tau) = \begin{cases} 1 & \text{for} \quad j = i-1, \\ 0 & \text{otherwise.} \end{cases}$$

III. The D'Alembert Reduction Method. The reduction method of §15 is valid for every homogeneous linear system. However, if applied directly to the system (2), it has the disadvantage that the new system of order $n-1$ no longer has the special form of (2), i.e., it cannot be written as a system of linear differential equations of order $n-1$. Therefore, it is expedient to modify the ansatz as follows:

Suppose $v(t) \not\equiv 0$ is a particular solution of $Lv = 0$ and

$$u(t) = v(t)w(t).$$

The function $w(t)$ is to be determined in such a way that u is also a solution. Applying the differential operator L, we have

$$Lu = \sum_{i=0}^n a_i \sum_{j=0}^i \binom{i}{j} w^{(j)}v^{(i-j)} = \sum_{j=0}^n w^{(j)} \sum_{i=j}^n \binom{i}{j} a_i(t)v^{(i-j)}.$$

The term corresponding to $j = 0$ in the sum on the right equals $w \cdot Lv$ and hence equals zero; thus we have

$$Lu = \sum_{j=1}^n w^{(j)}b_j(t) \quad \text{with} \quad b_j(t) := \sum_{i=j}^n \binom{i}{j} a_i(t)v^{(i-j)}$$

(note that the sum starts with $j = 1$). Therefore, $Lu = 0$ holds if and only if w satisfies the differential equation

$$L^* w = \sum_{j=1}^{n} b_j(t) w^{(j)} = 0.$$

This equation, however, is a differential equation of order $n-1$ for w'. Suppose that $n-1$ linearly independent solutions $w_1', \ldots, w_{n-1}'$ have been determined and $w_1, \ldots, w_{n-1}$ are corresponding antiderivatives. Then the n functions

$$v,\ vw_1, \ldots,\ vw_{n-1}$$

are a fundamental system for the original differential equation $Lu = 0$.

For proof, we consider a linear combination

$$c_0 v + c_1 v w_1 + \cdots + c_{n-1} v w_{n-1} = 0.$$

After division by v and differentiation one obtains

$$c_1 w_1' + \cdots + c_{n-1} w_{n-1}' = 0.$$

Therefore, $c_1 = \cdots = c_{n-1} = 0$ because of the linear independence of the w_i'.

IV. The Case $\boldsymbol{n = 2}$. If $v(t)$ is a solution of the equation $u'' + a_1(t)u' + a_0(t)u = 0$, then a second solution $u = vw$ is obtained by solving

$$w' \left(a_1 + 2\frac{v'}{v} \right) + w'' = 0.$$

Example.

$$u'' - u' \cos t + u \sin t = 0.$$

A solution is given by

$$v = \mathrm{e}^{\sin t}.$$

For $w(t)$, we have

$$w'' + w' \cos t = 0, \quad \text{i.e.} \quad w'(t) = \mathrm{e}^{-\sin t}.$$

Thus a second solution is

$$u(t) = \mathrm{e}^{\sin t} \int_0^t \mathrm{e}^{-\sin s}\, ds.$$

V. The Inhomogeneous Differential Equation. The result that was proved in 16.I for systems remains true:

Every solution $w(t)$ of the inhomogeneous differential equation

$$Lw = b(t) \tag{6}$$

can be written in the form

$$w = w^* + u,$$

where w^* is a particular solution of (6) and u is the general solution to the homogeneous differential equation.

A particular solution w of the differential equation (6) can be obtained by means of the

VI. Method of Variation of Constants. Let

$$w(t) = u_1(t)c_1(t) + \cdots + u_n(t)c_n(t),$$

where $u_1, \ldots, u_n$ is a fundamental system and $c_1, \ldots, c_n$ are functions that are yet to be determined. Instead of recomputing this ansatz, we refer to the result in §16, in particular (16.3). There it was shown that

$$\mathbf{z}(t) = Y(t) \int_\tau^t Y^{-1}(s)\mathbf{b}(s)\,ds$$

is a solution of the inhomogeneous differential equation $\mathbf{z}' = A(t)\mathbf{z} + \mathbf{b}(t)$. In order to carry this over to the system (2), we define

$$\mathbf{z} = \begin{pmatrix} w \\ w' \\ \vdots \\ w^{(n-1)} \end{pmatrix}, \quad Y(t) = \begin{pmatrix} u_1 & \cdots & u_n \\ \vdots & & \vdots \\ u_1^{(n-1)} & \cdots & u_n^{(n-1)} \end{pmatrix}, \quad \mathbf{b}(t) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ b(t) \end{pmatrix}.$$

Due to the special form of $\mathbf{b}$, the calculation of the expression $Y^{-1}(t)\mathbf{b}(t) = \mathbf{a}(t)$ is particularly simple. Since $\mathbf{a}(t)$ is the solution of the linear system of equations

$$Y \cdot \mathbf{a} = \mathbf{b},$$

then using Cramer's rule, one obtains the components of $\mathbf{a}$ in the form

$$a_i = \frac{V_i}{W},$$

with $W = \det Y$ and

$$V_i = \det \begin{pmatrix} u_1 & \cdots & u_{i-1} & 0 & u_{i+1} & \cdots & u_n \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ u_1^{(n-1)} & \cdots & u_{i-1}^{(n-1)} & b(t) & u_{i+1}^{(n-1)} & \cdots & u_n^{(n-1)} \end{pmatrix}.$$

Expanding V_i in terms of cofactors of the ith column yields

$$V_i(t) = (-1)^{n+i} b(t) W_i(t),$$

where W_i is the Wronskian determinant (of order $n-1$) of the functions u_1, $\ldots$, u_{i-1}, u_{i+1}, $\ldots$, u_n.

Thus a solution w of the inhomogeneous differential equation reads (w is the first component of $\mathbf{z}$)

$$w(t) = \sum_{i=1}^{n} u_i(t)(-1)^{n+i} \int_\tau^t \frac{b(t)}{W(s)} W_i(s)\, ds. \tag{7}$$

VII. The Case $n = 2$. If $u_1(t)$, $u_2(t)$ is a fundamental system for the homogeneous differential equation, then

$$w(t) = -u_1(t) \int_\tau^t \frac{b(s)u_2(s)}{W(s)}\, ds + u_2(t) \int_\tau^t \frac{b(s)u_1(s)}{W(s)}\, ds \tag{8}$$

is a solution of the inhomogeneous differential equation.

Example.

$$w'' - w' \cos t + w \sin t = \sin t.$$

The corresponding homogeneous equation was dealt with in VI. Using the fundamental system found there, $v = \mathrm{e}^{\sin t}$, $u = \phi v$ with $\phi(t) = \displaystyle\int_0^t \mathrm{e}^{-\sin s}\, ds$, we have

$$W(t) = \begin{vmatrix} \mathrm{e}^{\sin t} & \mathrm{e}^{\sin t}\phi(t) \\ \mathrm{e}^{\sin t}\cos t & 1 + \mathrm{e}^{\sin t}\phi(t)\cos t \end{vmatrix} = \mathrm{e}^{\sin t},$$

which also follows from (5). By (8),

$$w(t) = -\mathrm{e}^{\sin t} \int_0^t \sin r \left(\int_0^r \mathrm{e}^{-\sin s}\, ds \right) dr + \mathrm{e}^{\sin t}\phi(t) \int_0^t \sin r\, dr$$

is a solution to the given inhomogeneous differential equation. From the relations

$$\begin{aligned} \int_0^t \sin r \left(\int_0^r \mathrm{e}^{-\sin s}\, ds \right) dr &= \int_0^t \mathrm{e}^{-\sin s} \left(\int_s^t \sin r\, dr \right) ds \\ &= -\int_0^t \mathrm{e}^{-\sin s}(\cos t - \cos s)\, ds \\ &= -\phi(t)\cos t - \mathrm{e}^{-\sin t} + 1 \end{aligned}$$

it follows that

$$w(t) = \int_0^t \mathrm{e}^{\sin t - \sin s}\, ds + 1 - \mathrm{e}^{\sin t} = u(t) + 1 - v(t).$$

Hence $w_1(t) \equiv 1$ is also a solution.

§ 20. Linear Equations of Order n with Constant Coefficients

Now let

$$Lu = \sum_{i=0}^{n} a_i u^{(i)}(t) = 0, \quad a_i \text{ constant}, \ a_n = 1. \tag{1}$$

The characteristic polynomial

$$P(\lambda) = \begin{vmatrix} -\lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -\lambda & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-2} & -a_{n-1}-\lambda \end{vmatrix}$$

can be given explicitly. Expanding the determinant in cofactors of the last row, one obtains

$$P(\lambda) = (-1)^n[\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0]. \tag{2}$$

I. Theorem. *If λ is a zero of the characteristic polynomial of multiplicity k, then there are k solutions of the differential equation* (1)

$$\mathrm{e}^{\lambda t}, t\mathrm{e}^{\lambda t}, \ldots, t^{k-1}\mathrm{e}^{\lambda t}, \tag{3}$$

that correspond to λ. In this manner, one obtains n linearly independent solutions from the n zeros of the characteristic polynomial $P(\lambda)$ (each counted according to its multiplicity), that is, a fundamental system.

If the a_i are real and there exist complex zeros, then this fundamental system contains complex solutions. A real fundamental system can be obtained by splitting the k solutions in (3) *corresponding to a complex zero $\lambda = \mu + i\nu$ $(\nu \neq 0)$ into real and imaginary parts,*

$$t^q\mathrm{e}^{\mu t}\cos\nu t, \quad t^q\mathrm{e}^{\mu t}\sin\nu t \quad (q = 0, 1, \ldots, k-1)$$

(and discarding the solutions corresponding to $\bar{\lambda}$).

An elementary *proof*, which is independent of §17, will be given for this important theorem. Because of (2), the ansatz $u = \mathrm{e}^{\lambda t}$ leads to

$$L(\mathrm{e}^{\lambda t}) = \sum a_i(\mathrm{e}^{\lambda t})^{(i)} = \sum a_i\lambda^i\mathrm{e}^{\lambda t} = (-1)^n\mathrm{e}^{\lambda t}P(\lambda); \tag{4}$$

i.e., $u = e^{\lambda t}$ is a solution of (1) if and only if λ is a zero of the characteristic polynomial. In order to show that for a zero λ of multiplicity k the functions $t^q e^{\lambda t}$ $(0 \le q < k)$ are solutions, one makes use of the following trick: We have

$$t^q e^{\lambda t} = \frac{d^q}{d\lambda^q} e^{\lambda t},$$

and hence, because of (4),

$$L(t^q e^{\lambda t}) = L\left(\frac{d^q}{d\lambda^q} e^{\lambda t}\right) = \frac{d^q}{d\lambda^q} L(e^{\lambda t}) = (-1)^n \frac{d^q}{d\lambda^q}(e^{\lambda t} P(\lambda)).$$

The interchange of derivatives with respect to t and λ is clearly permissible. It was assumed that $P(\lambda)$ has a zero of multiplicity k at the point λ, i.e., that the derivatives of $P(\lambda)$ up to order $k-1$ vanish at the point λ. The same is then also true for the the derivatives of the function $e^{\lambda t} P(\lambda)$ (t fixed; product rule!). Hence we have $L(t^q e^{\lambda t}) = 0$ for $q = 0, 1, \ldots, k-1$.

In order to check that these n solutions are linearly independent, we consider an arbitrary linear combination of these solutions (with real or complex coefficients). It is clearly of the form

$$\phi(t) = \sum_{i=1}^{m} p_i(t) e^{\lambda_i t},$$

where $p_i(t)$ is a polynomial (with complex coefficients, in general) and $\lambda_1, \ldots, \lambda_m$ $(m \le n)$ are *distinct* numbers, namely the zeros of the characteristic polynomial (multiple zeros are only counted once).

We must show that $\phi(t)$ vanishes identically only if all $p_i(t)$ vanish. For $m = 1$ this is immediately clear. In the induction proof, it will be assumed that the result is true for m summands (with arbitrary polynomials) and that

$$\sum_{i=1}^{m} p_i(t) e^{\lambda_i t} + p(t) e^{\lambda t} \equiv 0 \quad (\lambda \ne \lambda_i).$$

Multiplication by $e^{-\lambda t}$ gives

$$\sum_{i=1}^{m} p_i(t) e^{\varrho_i t} + p(t) \equiv 0, \quad \varrho_i = \lambda_i - \lambda \ne 0.$$

If this equation is repeatedly differentiated until $p(t)$ vanishes, then what remains is an expression of the form

$$\sum_{i=1}^{m} q_i(t) e^{\varrho_i t} = 0,$$

and by the induction hypothesis $q_i(t) \equiv 0$, since the q_i are again polynomials. However, this is possible only if $p_i(t) \equiv 0$, since differentiation of an expression $r(t)e^{\varrho t}$ (r a polynomial $\ne 0$, $\varrho \ne 0$) gives rise to an expression $(r' + \varrho r)e^{\varrho t} = q(t)e^{\varrho t}$, where $q(t)$ is a polynomial of the same degree, hence $\ne 0$. ∎

Procedure for Finding Solutions. By Theorem I, the ansatz $u = e^{\lambda t}$ immediately produces the characteristic polynomial and with it all solutions.

II. Example. $u^{(5)} + 4u^{(4)} + 2y''' - 4u'' + 8u' + 16u = 0.$
The characteristic polynomial is

$$\begin{aligned}-P(\lambda) &= \lambda^5 + 4\lambda^4 + 2\lambda^3 - 4\lambda^2 + 8\lambda + 16 \\ &= (\lambda+2)^3(\lambda^2 - 2\lambda + 2) = (\lambda+2)^3(\lambda - 1 + i)(\lambda - 1 - i).\end{aligned}$$

A real fundamental system of solutions is given by

$$\mathrm{e}^{-2t},\ t\mathrm{e}^{-2t},\ t^2\mathrm{e}^{-2t},\ \mathrm{e}^t \sin t, \mathrm{e}^t \cos t.$$

III. Second Order Linear Differential Equations. The differential equation

$$Lu = u'' + 2au' + bu = 0 \tag{5}$$

arises in physics, for example as the differential equation of damped oscillations

$$m\ddot{s} + \beta\dot{s} + ks = 0 \quad \text{for} \quad s = s(t). \tag{6}$$

In the mechanical interpretation m is the mass, $s(t)$ the displacement from the equilibrium $s = 0$, $\beta > 0$ the coefficient of friction, $k > 0$ the coefficient of elasticity, i.e., of the linear restoring force ("spring constant").

The characteristic equation

$$P(\lambda) = \lambda^2 + 2a\lambda + b = 0$$

has the two roots,

$$\lambda = -a - \sqrt{a^2 - b}, \quad \mu = -a + \sqrt{a^2 - b}.$$

If only real coefficients are considered, then the following three cases need to be distinguished:

(a) $\boxed{a^2 > b: \quad u_1 = \mathrm{e}^{(-a+\sqrt{a^2-b})t},\ u_2 = \mathrm{e}^{(-a-\sqrt{a^2-b})t}}$

are real solutions. In the case $a > 0$, $b > 0$ both solutions tend to zero exponentially as $t \to \infty$.

Oscillator equation: $\beta^2 > 4km$, *overdamped motion* (nonoscillatory, aperiodic case).

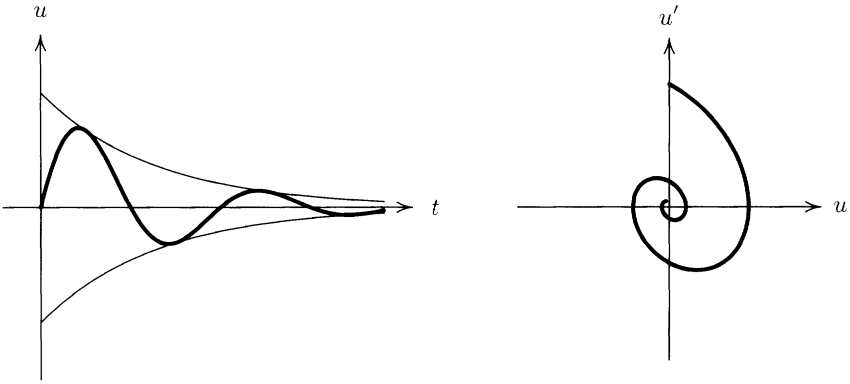
(b) $\boxed{a^2 = b: \quad u_1 = \mathrm{e}^{-at},\ u_2 = t\mathrm{e}^{-at}.}$

Oscillator equation: $\beta^2 = 4km$, *critically damped motion* (nonoscillatory, aperiodic case).

(c) $\boxed{a^2 < b: \quad u_1 = \mathrm{e}^{-at} \cos \sqrt{b - a^2}t,\ u_2 = \mathrm{e}^{-at} \sin \sqrt{b - a^2}t.}$

Oscillator equation: *Damped oscillations* with frequency

$$\nu = \frac{1}{2\pi}\sqrt{b - a^2} = \frac{1}{4\pi m}\sqrt{4km - \beta^2}.$$



Damped oscillation curve in the tu-plane (left) and trajectory (right)

If equation (5) is regarded as a plane system for $(x, y) = (u, u')$, then the classification introduced in 17.X leads to the following results (here it is assumed that $a > 0$, $b > 0$):

Case (a) (overdamping) corresponds to 17.X.(a) with the normal form $R(\lambda, \mu)$, where $\lambda < \mu < 0$. The origin is a stable node.

Case (b) (critical damping) falls under 17.X.(b); the normal form is $R_a(-a)$. Here again the origin is a stable node.

The case (c) of damped oscillations has the normal form $K(-a, \sqrt{b - a^2})$; cf. 17.X.(d). The origin is a stable vortex. The reader should study the two figures to get a clear understanding of the connection between the behavior of the function $u(t)$ and the trajectory in the phase plane.

A detailed discussion of damped oscillations is found in elementary texts.

IV. The Inhomogeneous Equation of the form

$$Lu = u'' + 2au' + bu = c \cdot \cos \alpha t \quad (a, b, c \text{ and } \alpha \neq 0 \text{ real}) \tag{7}$$

can be solved using the technique given in §19. However, one arrives at this goal more quickly by taking advantage of Exercise 18.VIII.(b). One considers the complex differential equation

$$u'' + 2au' + bu = c \cdot \mathrm{e}^{\mathrm{i}\alpha t} \tag{7'}$$

and uses the ansatz $u(t) = A\mathrm{e}^{\mathrm{i}\alpha t}$ (A complex). It leads to the equation

$$A(-\alpha^2 + 2\mathrm{i}a\alpha + b) = c,$$

from which A can be calculated (the term in parentheses vanishes only if $a = 0$ and $\alpha^2 = b$). The real part of u is a solution to the original equation (7).

In the case $b > 0$, the equation (7) describes an oscillatory system (with damping if $a > 0$) driven by an *external force* $c \cdot \cos \alpha t$ that acts on the system.

The solution represents a *forced oscillation* with the frequency of the external force.

The special case $a = 0$, $b = \alpha^2$ is called the *resonance case*; equation (7′) has then an unbounded solution

$$u = Ate^{\mathrm{i}\alpha t} \quad \text{with} \quad A = \frac{c}{2\alpha\mathrm{i}}.$$

The phenomena occurring for $a \approx 0$ are discussed in Exercise VIII.

V. Euler's Differential Equation. This name is given to a differential equation of the form (a_i constant)

$$Ly = a_n x^n y^{(n)} + a_{n-1} x^{n-1} y^{(n-1)} + \cdots + a_1 x y' + a_0 y = 0 \tag{8}$$

for $y = y(x)$. If $y(x)$ is a solution, then so is $y(-x)$; therefore, it is sufficient to study the case $x > 0$. Using the change of variables

$$x = \mathrm{e}^t,\ y(\mathrm{e}^t) = u(t),\ y(x) = u(\ln x)$$

we obtain the derivative formulas

$$\begin{aligned}
\frac{du}{dt} &= y'x & &\iff & xy' &= \frac{du}{dt},\\
\frac{d^2u}{dt^2} &= y'x + y''x^2 & &\iff & x^2y'' &= \frac{d^2u}{dt^2} - \frac{du}{dt},\\
\frac{d^3u}{dt^3} &= y'x + 3y''x^2 + y'''x^3 & &\iff & x^3y''' &= \frac{d^3u}{dx^3} - 3\frac{d^2u}{dx^2} + 2\frac{du}{dt},
\end{aligned}$$

etc., which lead to a linear differential equation with constant coefficients for $u(t)$,

$$Mu = b_n \frac{d^n u}{dt^n} + b_{n-1} \frac{d^{n-1} u}{dt^{n-1}} + \cdots + b_0 u = 0.$$

This equation can be solved in closed form using the techniques in I. By the way, $a_0 = b_0$ and $a_n = b_n$.

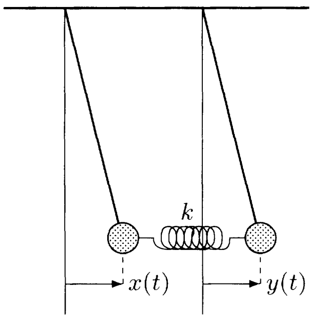
Procedure for Calculating the Solutions. The two operators L and M are connected through the equation

$$(Ly)(\mathrm{e}^t) = (Mu)(t) \quad \text{with} \quad u(t) = y(\mathrm{e}^t).$$

In particular, by equation (4),

$$L(x^\lambda) = M(\mathrm{e}^{\lambda t}) = (-1)^n P(\lambda) x^\lambda \quad (x = \mathrm{e}^t),$$

where P is the characteristic polynomial of M. Therefore, in order to obtain P, it is not necessary to calculate the operator M; it is sufficient to calculate $L(x^\lambda)$. Then all of the solutions can be given using the result of Theorem I.



Coupled pendulums

Example.

$$x^2 y'' - 3xy' + 7y = 0.$$

Here

$$L(x^\lambda) = [\lambda(\lambda - 1) - 3\lambda + 7]x^\lambda = P(\lambda)x^\lambda.$$

The characteristic equation $\lambda^2 - 4\lambda + 7 = 0$ has the roots $\lambda = 2 \pm \mathrm{i}\sqrt{3}$. Thus the differential equation $Mu = 0$ reads

$$\frac{d^2u}{dt^2} - 4\frac{du}{dt} + 7u = 0.$$

From the two real linearly independent solutions

$$u_1(t) = \mathrm{e}^{2t} \sin \sqrt{3}t, \quad u_2(t) = \mathrm{e}^{2t} \cos \sqrt{3}t$$

one obtains the solutions

$$y_1(x) = x^2 \sin(\sqrt{3} \ln x), \quad y_2(x) = x^2 \cos(\sqrt{3} \ln x)$$

of the original differential equation.

VI. Exercises. Determine all real solutions of the differential equations
(a) $y'' + 4y' + 4y = \mathrm{e}^x$,
(b) $y'' - 2y' + 5y = \mathrm{e}^x$.
In particular, find the solutions satisfying the initial conditions $y(0) = 1$, $y'(0) = 0$.

VII. Exercise. *Coupled Pendulums.* For two coupled pendulums of equal mass m and equal length l the equations of motion read

$$\begin{aligned} m\ddot{x} &= -\alpha x - k(x - y), \\ m\ddot{y} &= -\alpha y - k(y - x), \end{aligned} \qquad \text{with} \quad \alpha = mg/l$$

(g is the gravitational constant, k is the spring constant). Here the coordinate systems are chosen in such a way that $x = y = 0$ corresponds to the equilibrium point, and it is assumed that the pendulums hang vertically at rest. These are linearized equations, which are valid for small oscillations. Give a fundamental system of solutions (either by transforming to a system of first order and calculating the characteristic polynomial or by making two physically suggestive ansätze). Discuss the course of the motion if one pendulum is given a push at time $t = 0$, that is, $x(0) = 0$, $\dot{x}(0) = 1$, $y(0) = 0$, $\dot{y}(0) = 0$.

VIII. Exercise. Determine all (real) solutions of the differential equation

$$u'' + 2au' + \omega^2 u = c \cdot \cos \omega t \qquad (c > 0,\ 0 \le a < \omega).$$

Show that $L = \limsup_{t\to\infty} |u(t)|$ depends only on a, c, ω, and calculate $L(a, c, \omega)$ ($a = 0$ is a special case).

Remark. The above differential equation represents the simplest mathematical model for a *resonance phenomenon* in a periodically excited mechanical system (usually to be avoided in mechanical systems, but sometimes a desirable property in electrical circuits). In the differential equation for a harmonic oscillator $u'' + \omega^2 u = 0$, the solutions $u = \alpha \cos \omega t + \beta \sin \omega t$ describe a harmonic oscillation with frequency $\omega/(2\pi)$. If the system is excited with the same frequency (right side $= c \cdot \cos \omega t$), then resonance occurs, and the solutions grow without bound (in the case $a = 0$) as $t \to \infty$. In the case of damping ($a > 0$) (which is always present in practice), the solutions remain indeed bounded. However, the maximum amplitude tends to infinity as $a \to 0+$.

Supplement: Linear Differential Equations with Periodic Coefficients

IX. Second Order Equations with Periodic Coefficents. We consider the differential equation

$$u'' + 2a(t)u' + b(t)u = 0 \tag{9}$$

with real-valued, continuous, and ω-periodic coefficients a, b. We apply the Floquet theory, developed at the end of §18, to the equivalent system for $\mathbf{y} = (u, u')^T$,

$$\mathbf{y}' = A(t)\mathbf{y} \quad \text{with} \quad A(t) = \begin{pmatrix} 0 & 1 \\ -b & -2a \end{pmatrix}. \tag{9'}$$

The transition matrix $C = X(\omega)$ must be determined first. Thus let (u, v) be a fundamental system for (9) with initial values $u(0) = 1$, $u'(0) = 0$ and $v(0) = 0$, $v'(0) = 1$ and

$$X(t) = \begin{pmatrix} u & v \\ u' & v' \end{pmatrix} \Longrightarrow X(\omega) = C = \begin{pmatrix} u(\omega) & v(\omega) \\ u'(\omega) & v'(\omega) \end{pmatrix}.$$

Thus we have $\det(C - \lambda I) = \lambda^2 - \lambda \cdot \operatorname{tr} C + \det C$. The last term can be computed using formula (15.8): $\det C = \exp(\int_0^\omega \operatorname{tr} A(s)\, ds)$. Thus it remains only to calculate $\operatorname{tr} C$; then the stability behavior is essentially determined. Using the notation $\det C = \gamma > 0$, $\operatorname{tr} C = u(\omega) + v'(\omega) = 2\alpha$, the eigenvalues of C can be obtained from the equation

$$\lambda^2 - 2\alpha\lambda + \gamma = 0 \quad \text{as} \quad \lambda_{1,2} = \alpha \pm \sqrt{\alpha^2 - \gamma}.$$

From the root theorem of Vieta, we obtain the relations

$$\lambda_1\lambda_2 = \gamma > 0 \quad \text{and (suitably normalized)} \quad \omega(\mu_1 + \mu_2) = \log \gamma$$

for the characteristic multipliers λ_i and exponents μ_i (the latter are determined by $\lambda_i = \mathrm{e}^{\omega\mu_i}$). Theorem 18.XI yields the following separation into cases. Note here that the first component of a solution $\mathbf{y}(t) = \mathbf{q}(t)\mathrm{e}^{\mu t}$ of (9′) represents a solution of (9) and that α is real.

(a) $\alpha^2 \neq \gamma$. There exist two real or complex conjugate eigenvalues λ_1, λ_2 and correspondingly a fundamental system of solutions of the form

$$u_1(t) = p_1(t)\mathrm{e}^{\mu_1 t}, \quad u_2(t) = p_2(t)\mathrm{e}^{\mu_2 t},$$

where the p_i are ω-periodic functions. Recall that these two solutions satisfy $u_i(t + \omega) = \lambda_i u_i(t)$ $(i = 1, 2)$.

(b) $\alpha^2 = \gamma$ and the only eigenvalue α is semisimple. In this case C has two linearly independent eigenvectors and hence, similar to case (a), there is a fundamental system

$$u_1(t) = p_1(t)\mathrm{e}^{\mu t}, \quad u_2(t) = p_2(t)\mathrm{e}^{\mu t},$$

where the p_i are again ω-periodic and μ is determined by $\mathrm{e}^{\omega\mu} = \alpha$. Since every vector $\mathbf{y} \in \mathbb{R}^2$ satisfies the equation $C\mathbf{y} = \alpha\mathbf{y}$, we have $C = \alpha I$, whence $X(t + \omega) = \alpha X(t)$. Thus for every solution u we have $u(t + \omega) = \alpha u(t)$.

(c) $\alpha^2 = \gamma$, the eigenvalue α is not semisimple. There exists a fundamental system of solutions (p_i ω-periodic, $\alpha = \mathrm{e}^{\omega\mu}$)

$$u_1(t) = p_1(t)\mathrm{e}^{\mu t}, \quad u_2(t) = (p_2(t) + p_3(t)t)\mathrm{e}^{\mu t}.$$

In addition, one can assume that $p_1 = p_3$ here; the basis for this result is the last sentence in 17.VIII. These solutions may be complex. Real solutions are obtained by splitting into real and imaginary parts, similarly to 17.IV.

X. Hill's Differential Equation. If $a(t) \equiv 0$, one obtains from (9)

$$u'' + b(t)u = 0 \quad (b(t)\ \omega\text{-periodic}) \qquad \textit{Hill's equation.} \tag{10}$$

Here $\operatorname{tr} A(t) = 0$; hence $\gamma = \det C = 1$. The characteristic multipliers are given by

$$\lambda_{1,2} = \alpha \pm \sqrt{\alpha^2 - 1}, \quad \lambda_1\lambda_2 = 1, \quad \mu_1 + \mu_2 = 0.$$

Note that α is real. According to 18.XI, there are three cases:

$\lvert\alpha\rvert > 1$:	$\lambda_1 > 1$, the zero solution is unstable.
$\lvert\alpha\rvert < 1$:	$\lambda_{1,2} = \alpha \pm i\beta$ ($\beta > 0$) and thus $\lvert\lambda_1\rvert = \lvert\lambda_2\rvert = 1$. The zero solution is stable, but not asymptotically stable.
$\lvert\alpha\rvert = 1$:	$\lambda_1 = \lambda_2 = 1$ or -1. If the eigenvalue is semisimple, then the zero solution is stable, otherwise unstable.

If $|\alpha| = 1$ and the eigenvalue is semisimple, then by IX.(b), $X(\omega) = X(0)$ in the case $\alpha = 1$ and $X(2\omega) = X(0)$ in the case $\alpha = -1$. In the first (or second) case every solution of the differential equation is periodic with period ω (or 2ω).

A Special Case. Let the coefficient $b(t)$ be an even function. Then in the fundamental system considered above, $u(t) = u(-t)$ and $v(t) = -v(-t)$. From this observation and from $C^{-1} = X(-\omega)$ one deduces that $u(\omega) = v'(\omega)$ (exercise!). Thus the stability behavior (except for the case $\alpha = \pm 1$) is completely determined by a single function value $\alpha = u(\omega)$.

A well-known example with numerous physical applications is

$$u'' + (\delta + \gamma \cos 2t)u = 0 \quad (\omega = \pi) \qquad \textit{Mathieu's equation},$$

named after the French mathematician Emile-Léonard Mathieu (1835–1900). The *domain of stability*, the set of all points (γ, δ) with stable zero solution, can be represented in a figure in the $\gamma\delta$-plane (this is also true for other differential equations in which $b(t)$ depends on two parameters). Such a representation is called a stability map. It can be found, among others, in the book by L. Collatz (1988).

Numerous stability criteria have been established for Hill's equation (10). Two examples:

(a) If $b(t) \le 0$, then the differential equation is unstable.

(b) If $b(t) > 0$ and $\int_0^\omega b(t)\,dt \le 4/\omega$, then the differential equation is stable (Lyapunov 1839).

(c) *Exercise.* Carry out the above analysis for the differential equations $u'' + u = 0$ and $u'' = u$ in terms of ω (the calculation of C, α, λ_i, μ_i, stability). The coefficient $b = \pm 1$ is ω-periodic for every $\omega > 0$.

The *proof* of (a) is simple. By (10), $u'' = -b(t)u \ge 0$ as long as u is positive. Thus the solution u with $u(0) = 1$, $u'(0) = 1$ is convex and $\ge 1 + t$. ∎

The book by L. Cesari (1971) contains additional examples and proofs.

Chapter V
Complex Linear Systems

§ 21. Homogeneous Linear Systems in the Regular Case

I. Notation. The Space $H(G)$. The subject of this chapter is the homogeneous linear system

$$\mathbf{w}'(z) = A(z)\mathbf{w}(z), \tag{1}$$

where $\mathbf{w}(z) = (w_1(z), \ldots, w_n(z))^\top$ is a complex-valued vector function and $A(z) = (a_{ij}(z))$ is a complex-valued $n \times n$ matrix. We also investigate homogeneous linear differential equations of higher order. Let $G \subset \mathbb{C}$ be open and denote by $H(G)$ the complex linear space of functions that are single-valued and holomorphic on G. We write $\mathbf{w}(z) \in H(G)$ or $A(z) \in H(G)$ if every component $w_i(z)$ or $a_{ij}(z)$ belongs to $H(G)$. Compatible norms for complex column vectors and $n \times n$ matrices will be denoted by single vertical bars, and the properties (14.2–3),

$$|AB| \leq |A|\,|B| \quad \text{and} \quad |A\mathbf{w}| \leq |A|\,|\mathbf{w}|,$$

are taken for granted. Throughout this chapter, matrices are understood to be complex $n \times n$ matrices.

II. Theorem. *If G is simply connected and $A(z) \in H(G)$, then the initial value problem*

$$\mathbf{w}' = A(z)\mathbf{w}, \quad \mathbf{w}(z_0) = \mathbf{w}_0 \qquad (z_0 \in G,\ \mathbf{w}_0 \in \mathbb{C}^n) \tag{2}$$

has exactly one solution $\mathbf{w}(z) = \mathbf{w}(z; z_0, \mathbf{w}_0) \in H(G)$.

The solutions of (1) *form an n-dimensional (complex) linear subspace of $H(G)$. For a fixed z_0, the mapping $\mathbf{w}_0 \to \mathbf{w}(z; z_0, \mathbf{w}_0)$ is a linear isomorphism between $\mathbb{C}^n$ and this "solution space."*

This theorem is in almost complete agreement with the real Theorem 15.I. It is important to note that if G is simply connected, then every solution can be extended to all of G.

Proof. Local existence and uniqueness in a disk $|z - z_0| < \alpha$ follow immediately from Theorem 10.X. If z is an arbitrary point in G, then z can be connected to z_0 by a path lying in G, and the solution $\mathbf{w}$ can be extended along this path from z_0 to z by applying the local existence theorem 10.X.

It is easy to give a positive lower bound for the radii of disks appearing in the individual steps of the extension. Thus by the monodromy theorem the solution exists in G. The statements about the isomorphism are trivial; cf. 15.I.

Here is another proof that avoids analytic continuation and the monodromy theorem and at the same time yields an error estimate that will be important later on.

Let $p(z)$ be a real-valued continuous function defined on G such that

$$|A(z)| \le p(z) \quad \text{in} \quad G. \tag{3}$$

Let $C : \zeta = \zeta(s)$ $(0 \le s \le l)$ be a smooth curve, parametrized by arc length, connecting the points z_0 and z in G. Define

$$Q(z; C) = \int_C p\, ds = \int_0^l p(\zeta(s))\, ds$$

and

$$P(z) = \inf_C Q(z; C), \tag{4}$$

where the infimum is taken over all curves lying in G that connect the points z_0 and z. It is easy to see that $P(z)$ is bounded on every compact subset of G. Indeed, P is continuous; however, we do not need this fact. The set $B(G)$ of all vector functions $\mathbf{u} \in H(G)$ with

$$\|\mathbf{u}\| := \sup_G |\mathbf{u}(z)| \mathrm{e}^{-2P(z)} < \infty \tag{5}$$

is a Banach space. The completeness of $B(G)$ follows from the fact that convergence in the norm implies uniform convergence on compact subsets of G.

We consider the linear operator T defined by

$$(T\mathbf{u})(z) = \int_{z_0}^z A(\zeta)\mathbf{u}(\zeta)\, d\zeta, \quad \mathbf{u} \in B(G);$$

the integral is independent of path. Let $C : \zeta = \zeta(s)$ $(0 \le s \le l)$ be a smooth curve connecting the points z_0 and z and let

$$q(s) = \int_0^s p(\zeta(s'))\, ds', \quad \text{whence} \quad q(l) = Q(z; C).$$

Clearly, $q(s) \ge P(\zeta(s))$, since the function $\zeta(s')$, $0 \le s' \le s$, is a path connecting the points z_0 and $\zeta(s)$. Therefore, $|\mathbf{u}(\zeta(s))| \le \|\mathbf{u}\| \mathrm{e}^{2q(s)}$ and

$$\begin{aligned} |T\mathbf{u}(z)| &\le \int_C p(\zeta)|\mathbf{u}(\zeta)|\, ds \\ &\le \|\mathbf{u}\| \int_0^l p(\zeta(s)) \mathrm{e}^{2q(s)} ds \le \frac{\|\mathbf{u}\|}{2} \mathrm{e}^{2q(l)}, \end{aligned}$$

where the relation $(e^{2q(s)})' = 2p(\zeta(s))e^{2q(s)}$ is used to establish the final inequality. Since C is arbitrary, the term $q(l) = Q(z; C)$ on the right can be replaced by $P(z)$. Multiplying both sides of the resulting inequality by $e^{-2P(z)}$ and taking the supremum over G leads to

$$\|T\mathbf{u}\| \le \frac{1}{2}\|\mathbf{u}\|. \tag{6}$$

The initial value problem is equivalent to the operator equation $\mathbf{w} = \mathbf{w}_0 + T\mathbf{w} =: S\mathbf{w}$, where the operator S satisfies in $B(G)$ a Lipschitz condition $\|S\mathbf{u} - S\mathbf{v}\| = \|T(\mathbf{u} - \mathbf{v})\| \le \frac{1}{2}\|\mathbf{u} - \mathbf{v}\|$ with Lipschitz constant $1/2$. In addition, $\mathbf{w}_0 \in B(G)$. Therefore, by the fixed point theorem 5.IX, there exists exactly one solution $\mathbf{w}$ in $B(G)$. Since this proof works also in compact subsets $\overline{G}_1 \subset G$ and every solution $\mathbf{v} \in H(G)$ belongs to the Banach space $B(G_1)$, we get $\mathbf{v} = \mathbf{w}$ in G_1 and hence in G, i.e., all solutions belong to $B(G)$. ∎

III. Corollary. *The solution of the initial value problem* (2) *satisfies the estimate*

$$|\mathbf{w}(z)| \le 2|\mathbf{w}_0|e^{2P(z)} \quad in \quad G.$$

Proof. Because of (6), we have

$$\mathbf{w} = \mathbf{w}_0 + T\mathbf{w} \Longrightarrow \|\mathbf{w}\| \le \|\mathbf{w}_0\| + \frac{1}{2}\|\mathbf{w}\|,$$

whence $\|\mathbf{w}\| \le 2\|\mathbf{w}_0\| = 2|\mathbf{w}_0|$. The assertion follows. ∎

IV. Fundamental Matrices. By Theorem II, the solutions of (1) form an n-dimensional complex linear space, and the propositions discussed in 15.II, III also hold for (1). We recall them briefly. If a "solution matrix" W is formed with n solutions $\mathbf{w}_1, \ldots, \mathbf{w}_n$ as columns, then W satisfies the differential equation

$$W'(z) = A(z)W(z). \tag{7}$$

In particular, there exist n linearly independent solutions (a fundamental system) $\mathbf{w}_1, \ldots, \mathbf{w}_n$ from which every solution can be obtained as a linear combination

$$\mathbf{w} = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n \quad (c_i \in \mathbb{C}).$$

In this case, the solution matrix $W = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$ is again called a fundamental matrix. The following four statements are equivalent:

(a) $W(z)$ is a fundamental matrix.

(b) $W(z_0)$ is a nonsingular matrix for some $z_0 \in G$.

(c) $W(z_0)$ is nonsingular for every $z_0 \in G$.

(d) the Wronskian $\phi(z) = \det W(z)$ is $\neq 0$ in G.

The Wronskian $\phi(z)$ belongs to $H(G)$ and satisfies

$$\phi' = \operatorname{tr}(A(z)) \cdot \phi. \tag{8}$$

Finally, we recall Proposition 15.II.(h): Given a fundamental matrix $W(z)$, one can obtain every fundamental matrix in the form

$$U(z) = W(z)C, \quad C \text{ nonsingular.}$$

Exercise. Show that every solution of the equation $w'' = e^z w$ satisfies an estimate $|w(z)| + |w'(z)| \leq C \cdot \exp\{2\phi(x)|z|\}$, $z = x + iy$, $\phi(x) = 1 + (e^x - 1)/x$.

§ 22. Isolated Singularities

I. Statement of the Problem and Examples. We investigate the behavior of the solutions to the differential equation

$$\mathbf{w}' = A(z)\mathbf{w} \tag{1}$$

in a neighborhood of an isolated singular point z_0 of the matrix $A(z)$. It can be assumed that $z_0 = 0$ (one introduces the change of variables $z' = z - z_0$). Thus $A(z)$ is assumed to be single-valued and holomorphic for $0 < |z| < r$ $(r > 0)$.

An understanding of some elementary properties of the complex logarithm

$$\log z = \ln|z| + i \arg z + 2k\pi i \quad (k \text{ an integer}),$$

and the generalized power,

$$z^c = e^{c \log z} \quad (c \in \mathbb{C}),$$

is needed in this section. The argument of z is normalized to

$$-\pi < \arg z \leq \pi.$$

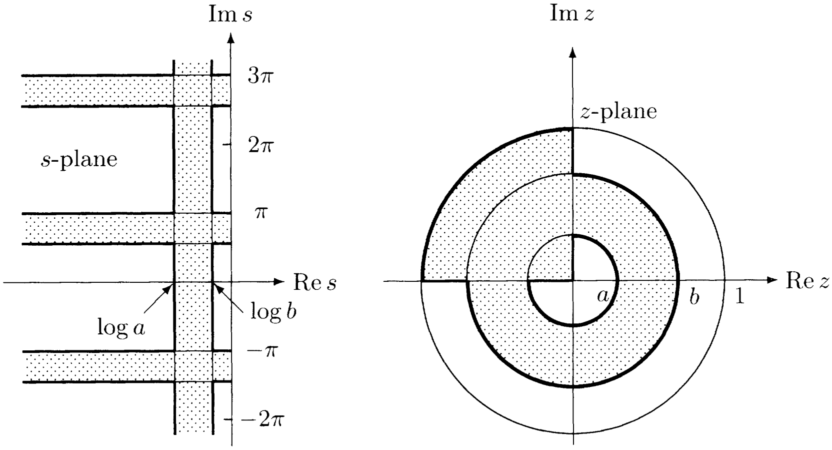
The logarithm is an analytic, infinitely many-valued function in the domain $G = \mathbb{C} \setminus \{0\}$. With $k = 0$ one obtains the principal value of the logarithm.

We begin with two examples.

(a) Let $n = 1$, $c \in \mathbb{C}$ and consider the differential equation

$$w' = \frac{c}{z} w.$$

A solution is given by $w = z^c$ (since it is $\neq 0$ and $n = 1$, this solution is a fundamental system). The function $A(z) = c/z$ is holomorphic in $G = \mathbb{C} \setminus \{0\}$. However, the domain G is not simply connected. Therefore, Theorem 21.II does not apply in G, but it does apply on simply connected subsets of G. For real integers c, the solution is single-valued and holomorphic in G. On the other hand, for $c = 1/2$ the solution $w = \sqrt{z}$ is double-valued, etc.



The mapping $s = \log z$

(b) Consider the system $(n = 2)$

$$w_1' = \frac{w_2}{z}, \quad w_2' = 0.$$

The corresponding matrix $A(z)$ is likewise holomorphic in $G = \mathbb{C}\backslash\{0\}$. From the solution $w_2 = c$ of the second equation one obtains $w_1 = c \log z$. A fundamental system of solutions is given by

$$W(z) = \begin{pmatrix} \log z & 1 \\ 1 & 0 \end{pmatrix}.$$

The first of these solutions is infinitely many-valued in G, the second is single-valued.

These examples show that the solutions of (1) can be infinitely many-valued functions in the neighborhood of an isolated singular point of the matrix $A(z)$. We introduce a transformation that clarifies this situation and shows a way to avoid some of the associated problems.

II. The Transformation $s = \log z$**.** The transformation

$$s = \log z, \qquad \text{or} \qquad z = \mathrm{e}^s,$$

maps the punctured disk $K_r^0 : 0 < |z| < r$ in the z-plane into the half-plane $\operatorname{Re} s < \log r$ of the s-plane.

Let R_α denote the half-plane $\operatorname{Re} s < \alpha$. The equation

$$v(s) := w(\mathrm{e}^s)$$

associates each many-valued analytic function w in K_r^0 with a holomorphic, single-valued function v in $R_{\log r}$. Conversely, every $v \in H(R_{\log r})$ gives rise to a possibly many-valued function w in K_r^0 given by

$$w(z) := v(\log z).$$

Here w is single-valued or m-fold many-valued if and only if $v(s)$ is periodic with period $2\pi\mathrm{i}$ or periodic with smallest period $2m\pi\mathrm{i}$, respectively. If neither of these cases is present, then $w(z)$ is infinitely many-valued in K_r^0. Some examples are

$$w(z) = (\log z)^2 \iff v(s) = s^2,$$

$$w(z) = z^c \iff v(s) = \mathrm{e}^{cs}.$$

As a first application of this transformation, we investigate

III. Euler Systems. These systems are of the form

$$\mathbf{w}' = \frac{A}{z}\mathbf{w}, \qquad A = (a_{ij}) \quad \text{constant.} \tag{2}$$

The function $\mathbf{w}(z)$ is a solution of (2) if and only if $\mathbf{v}(s) := \mathbf{w}(\mathrm{e}^s)$ satisfies the differential equation

$$\frac{d\mathbf{v}}{ds} = \mathbf{w}' \cdot \mathrm{e}^s = A\mathbf{w}(\mathrm{e}^s) = A\mathbf{v}(s).$$

We know from 18.II that such a system of differential equations with constant coefficients has a fundamental matrix of the form

$$V(s) = \mathrm{e}^{As} = \sum_{k=0}^{\infty} \frac{A^k s^k}{k!} \in H(\mathbb{C});$$

the proof that $\dfrac{d}{ds}(\mathrm{e}^{As}) = A\mathrm{e}^{As}$ given in §18 is also valid for $s \in \mathbb{C}$. Hence

$$W(z) = z^A \text{ with } z^A := \mathrm{e}^{A\log z} = \sum_{k=0}^{\infty} \frac{A^k(\log z)^k}{k!} \tag{3}$$

is an analytic, in general infinitely many-valued fundamental matrix for equation (2) in $K^0 = \mathbb{C} \setminus \{0\}$.

IV. The Structure of z^A. Formula (3) defines the power function z^A. Its structure is easily determined from the analysis of e^{As} carried out in §18. If A is a Jordan block, $A = \lambda I + F$ (F is defined as in 18.V), then by (18.9),

$$\mathrm{e}^{(\lambda I+F)s} = \mathrm{e}^{\lambda s}\mathrm{e}^{Fs},$$

where e^{Fs} is given by (18.8). It follows that

$$z^{\lambda I+F} = z^{\lambda} \cdot \begin{pmatrix} 1 & \log z & \frac{1}{2!}(\log z)^2 & \frac{1}{3!}(\log z)^3 & \cdots \\ 0 & 1 & \log z & \frac{1}{2!}(\log z)^2 & \cdots \\ 0 & 0 & 1 & \log z & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{4}$$

If A is an arbitrary matrix and C a nonsingular matrix with $A = C^{-1}BC$, then by 18.III, $\mathrm{e}^A = C^{-1}\mathrm{e}^B C$ holds, and hence

$$z^A = \mathrm{e}^{A \log z} = C^{-1}\mathrm{e}^{B \log z} C = C^{-1} z^B C.$$

If C is chosen such that B has Jordan normal form (17.10), then z^B is obtained from B by replacing the individual Jordan blocks of B with square blocks of the form (4).

The main result of this section is the following theorem. It shows that the examples previously considered in I and III are representative of arbitrary systems with isolated singularities at $z = 0$. More precisely, it says that the solutions are products of at most three factors: a holomorphic function, z^λ, and $\log z$. Other kinds of many-valued functions do not occur.

V. Theorem. *If $A(z)$ is single-valued and holomorphic in $K_r^0 : 0 < |z| < r$, then* (1) *has a fundamental matrix of the form*

$$W(z) = U(z) z^B, \tag{5}$$

where $U(z)$ is a single-valued holomorphic function in K_r^0 and B is a constant matrix.

Proof. If $W(z)$ is a fundamental matrix of (1) and $V(s) = W(\mathrm{e}^s)$, then

$$\frac{d}{ds} V(s) = \mathrm{e}^s A(\mathrm{e}^s) V(s). \tag{6}$$

By 21.II, this differential equation has a fundamental matrix $V(s)$ that is holomorphic in $R_{\log r}$. Since $\mathrm{e}^s A(\mathrm{e}^s)$ is periodic with period $2\pi\mathrm{i}$, the matrix $V(s + 2\pi\mathrm{i})$ is also a solution of (6) and indeed is again a fundamental matrix by 21.IV.(b),(c). Hence, by 21.IV,

$$V(s + 2\pi\mathrm{i}) = V(s)C, \qquad C \text{ nonsingular.}$$

By a theorem from matrix theory, there exists a matrix B such that $C = \mathrm{e}^{2\pi\mathrm{i}B}$; cf. Lemma VI below. The function

$$T(s) := V(s)\mathrm{e}^{-Bs}$$

satisfies the equation

$$T(s + 2\pi\mathrm{i}) = V(s + 2\pi\mathrm{i})\mathrm{e}^{-B(s+2\pi\mathrm{i})} = V(s)\mathrm{e}^{2\pi\mathrm{i}B}\mathrm{e}^{-B(s+2\pi\mathrm{i})} = T(s),$$

i.e., $T(s)$ is $2\pi\mathrm{i}$-periodic. Therefore, $U(z) = T(\log z)$ is single-valued in K_r^0, and $W(z) := V(\log z) = T(\log z) z^B$ has the form given by (5). ∎

Remark. Every fundamental matrix has the form (5), since for a nonsingular matrix C

$$W(z)C = U(z)CC^{-1}z^B C = U(z)Cz^D \quad \text{with} \quad D = C^{-1}BC.$$

VI. Lemma. *For every nonsingular $n \times n$ matrix C, there exists a $n \times n$ matrix X such that*

$$e^X = C.$$

The *proof* uses matrix versions of some basic facts on infinite series. As in 18.I, matrices are of size $n \times n$, and $|\cdot|$ is a compatible matrix norm.

(a) For a complex double series, we have the theorem

$$\sum_i \Big(\sum_j c_{ij}\Big) = \sum_j \Big(\sum_i c_{ij}\Big) = \sum_k \Big(\sum_{i+j=k} c_{ij}\Big) \text{ if } \sum_i \Big(\sum_j |c_{ij}|\Big) < \infty.$$

This carries immediately over to matrices C_{ij}; cf. 18.I.

We now consider power series $f(z) = \sum f_i z^i$, $g(z) = \sum g_i z^i$, $h(z) = \sum h_i z^i$ (i runs from 0 to ∞) with positive radii of convergence r_f, r_g, r_h.

(b) Consider the Cauchy product $f(z)g(z) = h(z)$, $h_i = f_0 g_i + \cdots + f_i g_0$. If $|B| < \min(r_f, r_g)$, then $f(B)g(B) = h(B)$. This follows from (a).

(c) We consider $h(z) = f(g(z))$. Let $G(z) = \sum \gamma_i z^i$, $\gamma_i = |g_i|$, and denote the power series expansions of the power g^k or G^k by $\sum g_i^k z^i$ or $\sum \gamma_i^k z^i$, resp. If $G(\rho) < r_f$, where $0 < \rho < r_g$, then h has a power series expansion

$$h(z) = f(g(z)) = \sum h_i z^i, \quad \text{where } h_i = \sum_k f_k g_i^k.$$

The series is absolutely convergent for $|z| \leq \rho$. This is a classical result that can be proved by observing that $|g_i^k| \leq \gamma_i^k$, $|f(G(\rho))| \leq \sum_{i,k} |f_k| \gamma_i^k \rho^i < \infty$. It follows easily that

$$f(g(B)) = h(B) \text{ if } |B| \leq \rho.$$

Taking $f(z) = e^z$, $g(z) = \log(1+z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \cdots$, $r_g = 1$, $h(z) = 1 + z$, one obtains

$$e^{g(B)} = I + B \text{ if } |B| < 1. \tag{7}$$

We come now to the proof. The equation $e^X = C$ holds if and only if

$$T^{-1} e^X T = e^{T^{-1}XT} = T^{-1}CT \qquad (T \text{ nonsingular}).$$

Therefore, one may assume that C has Jordan normal form. In fact, it can be assumed that C has the form of a Jordan block, $C = \lambda I + F$; cf. 17.VII and 18.V. If for each Jordan block J_k of C a matrix X_k with $e^{X_k} = J_k$ has been found, one simply builds a matrix X by putting X_k in the place of J_k. This matrix X has the desired property $e^X = C$.

Therefore, let $C = \lambda I + F$, where $\lambda \neq 0$, since C is nonsingular. Since F is nilpotent, $F^n = 0$, we can make $|F|$ as small as we want by an appropriate choice of the norm; cf. D.IV and D.VII in the Appendix. Hence we can use (7) for $B = F/\lambda$ and prove that $X = g(F/\lambda) + I \log \lambda$ has the desired property:

$$e^X = \lambda e^{g(F/\lambda)} = \lambda(I + F/\lambda) = \lambda I + F = C.$$

∎

We conclude this section with a theorem about the

VII. Growth of Solutions in the neighborhood of a singular point. Here we will restrict ourselves to the case where $A(z)$ has a pole at $z = 0$. Let K_r^- be a disk with a cut along the negative real axis; that is, K_r^- is the set of all z with $|z| < r$ for which $\operatorname{Re} z \leq 0$ implies $\operatorname{Im} z \neq 0$.

Theorem. *Let $A(z) \in H(K_r^0)$ have a pole of order $m \geq 1$ at $z = 0$. If $\mathbf{w}(z)$ is a solution of (1) that is single-valued and holomorphic in K_r^-, then there exist positive constants a, b such that*

$$|\mathbf{w}(z)| \leq \begin{cases} a|z|^{-b}, & \text{for } m = 1, \\ ae^{b|z|^{1-m}}, & \text{for } m > 1, \end{cases} \qquad \text{for } z \in K_{r/2}^-. \tag{8}$$

Remark. If $\mathbf{w}(z)$ is a single-valued solution in K_r^0, then the estimate (8) holds in $K_{r/2}^0$. If $\mathbf{w}(z)$ is many-valued, then (8) holds for every branch of $\mathbf{w}$ belonging to $H(K_r^-)$. However, the constants a, b may depend on the chosen branch.

Proof. By hypothesis, there exists a constant c with

$$|A(z)| \leq c|z|^{-m} \qquad \text{for} \qquad 0 < |z| \leq \alpha,$$

$\alpha = r/2$. We now apply Corollary 21.III in $G = K_r^-$ and note that G is simply connected. We can set $p(z) = c|z|^{-m}$ (the values of $p(z)$ for $|z| > \alpha$ play no role in the following argument). Let $z_0 = \alpha$. If z is a point in $K_{r/2}^-$, then we connect α and z using a path C that goes along the real axis from α to $|z|$, and then along a circular arc from $|z|$ to z. Then, in the notation of 21.II, we have

$$P(z) \leq \int_C p(\zeta(s))\, ds \leq \int_{|z|}^{\alpha} ct^{-m} dt + c|z|^{-m}\pi|z|.$$

The last term comes from the fact that the circular arc from $|z|$ to z has length at most $\pi|z|$. Thus

$$P(z) \leq \begin{cases} c \log \dfrac{\alpha}{|z|} + \pi c & \text{for} \quad m = 1, \\ c|z|^{1-m}(m - 1 + \pi) & \text{for} \quad m > 1. \end{cases}$$

Now the assertion follows from the estimate of Corollary 21.III,

$$|\mathbf{w}(z) \leq 2|\mathbf{w}(\alpha)|e^{2P(z)}.$$

■

VIII. Exercise. Determine a fundamental system of solutions for the system

$$w_1' = w_2, \qquad w_2' = \frac{\alpha w_1}{z^2}.$$

For which values of α are all solutions rational functions?

§ 23. Weakly Singular Points. Equations of Fuchsian Type

I. Definition. Let the matrix $A(z)$ be single-valued and holomorphic for $0 < |z - z_0| < r$ $(r > 0)$. A point $z = z_0$ is called a *weakly singular point* of the differential equation

$$\mathbf{w}' = A(z)\mathbf{w} \tag{1}$$

if $A(z)$ has a pole of the first order at the point z_0. Thus, restricting again to the case $z_0 = 0$, $A(z)$ can be represented in the form

$$A(z) = \frac{1}{z}\sum_{k=0}^{\infty} A_k z^k, \quad \text{with} \quad A_0 \neq 0, \tag{2}$$

where the power series on the right-hand side converges in a disk $|z| < r$ $(r > 0)$. Here A_0, A_1,... are constant matrices. Formula (2) can be interpreted as matrix notation for n^2 power series for the components $a_{ij}(z)$, whose radii of convergence are all $\geq r$. The condition $A_0 \neq 0$ means that at least one of the functions $a_{ij}(z)$ has a pole at $z = 0$.

If $A_0 = 0$, then $A(z)$ is holomorphic at 0, and the point $z = 0$ is called a *regular point* of (1). We dealt with this case in §21. If $z = 0$ is neither regular nor weakly singular, then it is called *strongly singular*. The latter occurs if and only if at least one of the functions $a_{ij}(z)$ has a pole of order ≥ 2 or an essential singularity at $z = 0$.

The conclusion of Theorem 22.V can be significantly sharpened for weakly singular points. In this case the function $U(z)$ that arises there is holomorphic in the whole disk.

II. Theorem. *If $A(z)$ is holomorphic for $0 < |z| < r$ and if $z = 0$ is a weakly singular point of the differential equation* (1), *then every fundamental matrix has the form*

$$W(z) = U(z)z^B, \tag{3}$$

where $U(z)$ is a single-valued, holomorphic function in $K_r : |z| < r$ and B is a constant matrix.

Remark. A representation of the form (3) is not unique. In fact, the relation $e^{\alpha I} = e^{\alpha} I$ for $\alpha = k \log z$ (k a whole number) implies that $z^{kI} = z^k I$, and hence also $z^k z^{-kI} = I$. It follows then from (3) that

$$W(z) = (U(z)z^k)z^{B-kI}, \tag{4}$$

which is again a representation of the form (3).

Therefore, it is sufficient to show that a representation of the form (3) exists where the function $U(z)$ has, at worst, a pole at the point $z = 0$. If k is the order of the pole, then $U(z)z^k$ is holomorphic at $z = 0$, and the representation (4) has the property required by the theorem.

Theorem II goes back to Sauvage (1886). More detailed historical information on singular point theory can be found in the book by Hartman (1964), pages 91–92.

Proof. Let $W(z)$ be a fundamental matrix of (1). By Theorem 22.V (with remark) $W(z)$ has the form (3), where $U(z)$ is holomorphic in K_r^0. We consider $W(z)$ in K_r^-, where K_r^- is the disk with a cut along the negative real axis, and in particular the branch corresponding to the principal value of the logarithm. On this branch we have the estimate

$$|\log z| \le \log \frac{1}{|z|} + \pi \qquad (|z| < 1).$$

Further, from the series expansion of the exponential function one obtains $|e^{Bs}| \le e^{|B||s|}$, and hence

$$|z^{-B}| \le e^{|B||\log z|} \le c|z|^{-\beta} \quad \text{with} \quad \beta = |B|.$$

By 22.VII, the estimate

$$|W(z)| \le a|z|^{-b}$$

holds for $W(z)$ in $K_{r/2}^-$ with positive constants a, b, and therefore

$$|U(z)| = |W(z)z^{-B}| \le a|z|^{-b}c|z|^{-\beta}.$$

This implies that $U(z)$ has at most a pole at $z = 0$. ∎

Remark on Nomenclature. In the literature one finds different names for a weakly singular point such as simple singularity (Hartman), singularity of the first kind (Coddington–Levinson), and regular singular point (often in connection with second order differential equations). Some authors define a regular singular point by the properties of solutions expressed in Theorem II. The strongly singular point shares the same fate.

III. Singularities at Infinity. A function $f(z)$ that is holomorphic for $|z| > r$ is said to have a zero or a pole of order k at $z = \infty$ if the same statement applies to the function $g(z) = f(1/z)$ at $z = 0$. And $f(z)$ is said to be holomorphic at the point $z = \infty$ if $g(z)$ is holomorphic at $z = 0$.

We consider now the case where $A(z)$ in (1) is holomorphic for $|z| > r$. If $\mathbf{w}(z)$ is a solution of (1), then the function $\mathbf{v}(\zeta) = \mathbf{w}(1/\zeta)$ satisfies

$$\mathbf{v}'(\zeta) = -\frac{1}{\zeta^2} A\left(\frac{1}{\zeta}\right)\mathbf{v}(\zeta). \tag{5}$$

The point $z = \infty$ is called (i) regular or (ii) weakly singular or (iii) strongly singular for (1) if the point $\zeta = 0$ has this property with respect to equation (5). This is the case if and only if in the Laurent expansion

$$A\left(\frac{1}{\zeta}\right) = \sum_{k=-\infty}^{\infty} B_k \zeta^k,$$

(i) all B_k with $k \leq 1$ vanish or (ii) all B_k with $k \leq 0$ vanish and $B_1 \neq 0$ or (iii) not all B_k with $k \leq 0$ vanish.

IV. Theorem. *Let $A(z)$ be holomorphic for $|z| > r$. The point $z = \infty$ is weakly singular or regular if and only if $A(z)$ has a zero of first order or higher order at infinity, respectively, i.e., if $A(z)$ has an expansion of the form*

$$A(z) = \frac{B_1}{z} + \frac{B_2}{z^2} + \frac{B_3}{z^3} + \cdots \quad (|z| > r)$$

with $B_1 \neq 0$ or $B_1 = 0$, respectively.

Every fundamental matrix of (1) *has the form*

$$W(z) = U(z)z^B \quad \text{if } z = \infty \text{ is weakly singular,}$$

$$W(z) = U(z) \quad \text{if } z = \infty \text{ is regular,}$$

where $U(z)$ is a single-valued, holomorphic function for $|z| > r$ and $z = \infty$, and B is a constant matrix.

These statements are essentially proved by the discussion in III. Note that since $(1/z)^B = z^{-B}$, a fundamental matrix $V(\zeta) = U(\zeta)\zeta^B$ $(|\zeta| < 1/r)$ of equation (5) gives rise to a fundamental matrix $W(z) = V(1/z)$ of equation (1) of the form given above.

Of particular interest are the

V. Equations of Fuchsian Type. Equation (1) is called a differential equation of Fuchsian type if it has finitely many weakly singular points and if every other point of $\mathbb{C} \cup \{\infty\}$ is regular.

Theorem. *Equation* (1) *is a differential equation of Fuchsian type with weak singularities at the (pairwise distinct) points $z_1, \ldots, z_k \in \mathbb{C}$ and possibly also at ∞ if and only if*

$$A(z) = \sum_{j=1}^{\infty} \frac{1}{z - z_j} R_j, \tag{6}$$

where R_j are constant matrices $\neq 0$. If $\sum_{j=1}^{k} R_j = 0$, then ∞ is a regular point; if the sum is $\neq 0$, it is a weakly singular point.

Proof. The expansion of $A(z)$ about a weakly singular point z_j begins with $A(z) = (z - z_j)^{-1}R_j + \cdots$, $R_j \neq 0$. Therefore,

$$B(z) = A(z) - \sum_{j=1}^{k} \frac{1}{z - z_j} R_j$$

is holomorphic in $\mathbb{C}$. Since ∞ is either regular or weakly singular, we have $A(\infty) = 0$ by Theorem IV, and hence also $B(\infty) = 0$. Thus $B(z)$ is bounded and therefore by Liouville's theorem a constant. Then clearly, $B(z) \equiv 0$, and (6) follows. By Theorem IV, ∞ is regular or weakly singular if and only if the limit of $zA(z)$ as $z \to \infty$ is 0 or $\neq 0$, respectively. The assertion about the point ∞ follows from this. ∎

Remark. The theorem shows that except for the trivial case $A(z) \equiv 0$, every differential equation of Fuchsian type has at least two weak singularities.

§ 24. Series Expansion of Solutions

In this section we investigate series expansions of solutions of the differential equation

$$\mathbf{w}' = A(z)\mathbf{w} \tag{1}$$

in the neighborhood of a weakly singular point z_0. In the process, we obtain not only an algorithm for the computation of solutions, but we also gain a deeper insight into the structure of solutions in view of Theorem 23.II. We again assume without loss of generality that $z_0 = 0$.

First we consider power series, that is, vector-valued holomorphic functions of the form

$$\mathbf{u}(z) = \sum_{k=0}^{\infty} \mathbf{u}_k z^k \quad \text{with} \quad \mathbf{u}_k \in \mathbb{C}^n. \tag{2}$$

We use functional analytic methods to deal with the question of convergence and to this end introduce a new Banach space.

I. The Banach Space H_δ. The set of all sequences $\mathbf{u} = (\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \ldots) = (\mathbf{u}_k)_0^\infty$, $\mathbf{u}_k \in \mathbb{C}^n$, with finite norm

$$\|\mathbf{u}\| = \sum_{k=0}^{\infty} |\mathbf{u}_k| \delta^k \quad (\delta > 0 \text{ fixed})$$

forms a Banach space, which we denote by H_δ.

It is easy to see that H_δ is a vector space and $\|\cdot\|$ a norm. For example, the triangle inequality follows from

$$\|\mathbf{u} + \mathbf{v}\| = \sum |\mathbf{u}_k + \mathbf{v}_k|\delta^k \leq \sum |\mathbf{u}_k|\delta^k + \sum |\mathbf{v}_k|\delta^k = \|\mathbf{u}\| + \|\mathbf{v}\|.$$

To prove completeness, let $\mathbf{u}^1, \mathbf{u}^2, \ldots$ with $\mathbf{u}^n = (\mathbf{u}_k^n)_{k=0}^\infty$ be a Cauchy sequence in H_δ; that is, $\|\mathbf{u}^m - \mathbf{u}^n\| \to 0$ as $m, n \to \infty$. Because

$$|\mathbf{u}_k^m - \mathbf{u}_k^n|\delta^k \le \|\mathbf{u}^m - \mathbf{u}^n\| \to 0 \quad \text{for} \quad k = 0, 1, 2, \ldots,$$

there exists $\mathbf{v}_k \in \mathbb{C}^n$ such that $\mathbf{u}_k^n \to \mathbf{v}_k$ as $n \to \infty$. Let $\sum'$ denote an arbitrary finite sum. Then

$$\sum_k{}' |\mathbf{u}_k^m - \mathbf{u}_k^n|\delta^k \le \|\mathbf{u}^m - \mathbf{u}^n\| < \varepsilon \quad \text{for} \quad m, n > n_0,$$

and hence

$$\sum_k{}' |\mathbf{u}_k^m - \mathbf{v}_k|\delta^k \le \varepsilon \quad \text{for} \quad m > n_0.$$

The same then also holds for the sum running from $k = 0$ to ∞; i.e., if we set $\mathbf{v} = (\mathbf{v}_k)$, then we have $\|\mathbf{v} - \mathbf{u}^m\| \le \varepsilon$ and, in particular, $\mathbf{v} - \mathbf{u}^m \in H_\delta$. Thus it is also true that $\mathbf{v} = (\mathbf{v} - \mathbf{u}^m) + \mathbf{u}^m \in H_\delta$ and $\mathbf{u}^m \to \mathbf{v}$ in H_δ.

By formula (2), an element $\mathbf{u} = (\mathbf{u}_k)_0^\infty$ of H_δ generates a holomorphic function $\mathbf{u}(z)$ on $K_\delta : |z| < \delta$. Furthermore, the power series (2) is absolutely convergent for $z = \delta$ and hence absolutely and uniformly convergent in the closed disk $|z| \le \delta$. Conversely, if $\mathbf{u}(z)$ is a holomorphic function in K_δ whose power series (2) is absolutely convergent for $z = \delta$, then the sequence of coefficients $(\mathbf{u}_k)_0^k$ belongs to H_δ. In this sense, elements of H_δ can be identified with the functions $\mathbf{u}(z)$ generated by them.

II. Power Series Ansatz. Formal Solution. To obtain a solution of (1), we make the ansatz

$$\mathbf{w}(z) = \sum_{k=0}^\infty \mathbf{w}_k z^k, \quad \mathbf{w}_k \in \mathbb{C}^n.$$

If $A(z)$ has the form

$$A(z) = \frac{1}{z} \sum_{k=0}^\infty A_k z^k \quad (0 < |z| < r), \tag{3}$$

then (1) implies

$$z \sum_{k=0}^\infty k \mathbf{w}_k z^{k-1} = \left(\sum_{k=0}^\infty A_k z^k \right) \left(\sum_{i=0}^\infty \mathbf{w}_i z^i \right).$$

Forming the Cauchy product and equating coefficients of like terms, one obtains

$$k \mathbf{w}_k = \sum_{j=0}^k A_{k-j} \mathbf{w}_j \quad (k = 0, 1, 2, \ldots), \tag{4}$$

or, when the summand $A_0\mathbf{w}_k$ is brought to the left-hand side,

$$\begin{aligned} -A_0\mathbf{w}_0 &= \mathbf{0}, \\ (I - A_0)\mathbf{w}_1 &= A_1\mathbf{w}_0, \\ &\vdots \\ (kI - A_0)\mathbf{w}_k &= A_k\mathbf{w}_0 + \cdots + A_1\mathbf{w}_{k-1}, \\ &\vdots \end{aligned} \tag{4'}$$

A *formal solution* is by definition a sequence $(\mathbf{w}_k)_0^\infty$ that satisfies (4). If the formal solution $(\mathbf{w}_k)$ belongs to H_δ, then the function $\mathbf{w}(z) = \sum \mathbf{w}_k z^k$ is holomorphic in K_δ and it is a solution of (1), since term-by-term differentiation is allowed.

III. Convergence Theorem. *Let the series* (3) *converge for* $0 < |z| < r$. *Then every formal solution of* (1) *is a convergent power series for* $|z| < r$ *and hence is a holomorphic solution for* $|z| < r$.

The essence of the proof rests on an investigation of the following

IV. Two Operators in H_δ. Let two linear operators A and J_m from H_δ into itself be defined by

$$\mathbf{v} = A\mathbf{u} \iff \mathbf{v}_k = \sum_{j=0}^{k} A_{k-j}\mathbf{u}_j, \tag{5}$$

$$\mathbf{v} = J_m\mathbf{u} \iff \mathbf{v}_k = \begin{cases} \mathbf{0} & \text{for } k < m, \\ \dfrac{\mathbf{u}_k}{k} & \text{for } k \geq m. \end{cases} \tag{6}$$

Here the A_k are the matrices appearing in (3). Clearly, (5) is merely a restatement of the multiplication $\mathbf{v}(z) = zA(z)\mathbf{u}(z)$ in terms of the sequence of coefficients. We will prove the following two propositions:

(a) *If* $C := \sum_{k=0}^{\infty} |A_k|\delta^k < \infty$, *then* $\|A\| \leq C$; *i.e.,* $\|A\mathbf{u}\| \leq C\|\mathbf{u}\|$ *for* $\mathbf{u} \in H_\delta$.

(b) *The inequality* $\|J_m\| \leq \dfrac{1}{m}$ *holds for* $m = 1, 2, \ldots$

Whereas (b) is obvious, proposition (a) requires a short calculation:

$$\begin{aligned} C\|\mathbf{u}\| &= \sum_{i=0}^{\infty} |A_i|\delta^i \sum_{j=0}^{\infty} |\mathbf{u}_j|\delta^j = \sum_{k=0}^{\infty} \delta^k \sum_{j=0}^{k} |A_{k-j}||\mathbf{u}_j| \\ &\geq \sum_{k=0}^{\infty} \delta^k |\mathbf{v}_k| = \|\mathbf{v}\|. \end{aligned}$$

∎

V. Proof of the Convergence Theorem. Let $\bar{\mathbf{w}} = (\bar{\mathbf{w}}_k)$ be a formal solution of (1) and let $\delta \in (0, r)$ and $m > C$, where C is the constant in IV.(a). We claim that the equation

$$\mathbf{w} = J_m A\mathbf{w} + (\bar{\mathbf{w}}_0, \bar{\mathbf{w}}_1, \ldots, \bar{\mathbf{w}}_{m-1}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \ldots) \tag{7}$$

has exactly one solution in H_δ. Indeed, $J_m A$ is a linear operator, and by IV.(a), (b),

$$\|J_m A\mathbf{w}\| \le \frac{1}{m}\|A\mathbf{w}\| \le \frac{C}{m}\|\mathbf{w}\| \quad \text{with} \quad \frac{C}{m} < 1.$$

The assertion now follows from the fixed point theorem 5.IX.

The solution $\mathbf{w}$ of (7) satisfies the relations

$$\mathbf{w}_k = \bar{\mathbf{w}}_k \quad \text{for} \quad k = 0, 1, \ldots, m-1,$$

since the operator J_m sets the first m terms equal to zero. For $k = m$, we have

$$\mathbf{w}_m = \frac{1}{m}(A\mathbf{w})_m \iff m\mathbf{w}_m = \sum_{j=0}^{m} A_{m-j}\mathbf{w}_j,$$

and corresponding results hold for $k > m$; i.e., $\mathbf{w} = (\mathbf{w}_k)$ satisfies (4) and is therefore also a formal solution.

The matrix $\lambda I - A_0$ is singular for at most finitely many λ (the eigenvalues); thus $kI - A_0$ is nonsingular for large k, say $k \ge k_0$. This means that for $k \ge k_0$, the equations (4′) can be uniquely solved for $\mathbf{w}_k$, or equivalently, the $\mathbf{w}_k$ with indices $k \ge k_0$ are uniquely determined by the $\mathbf{w}_k$ with indices $k < k_0$. Thus, if one chooses $m \ge k_0$, then the relations $\mathbf{w}_k = \bar{\mathbf{w}}_k$ for $k < m$ imply that $\mathbf{w} = \bar{\mathbf{w}}_k$ for all k. Therefore, the formal solution $\bar{\mathbf{w}}$ belongs to H_δ, and the power series converges for $|z| \le \delta$. Since δ can be chosen arbitrarily close to r, proof of the convergence theorem is complete. ∎

VI. Discussion of the Results. (a) Suppose $\lambda = 0$ is an eigenvalue of A_0. Then the first equation of (4), $A_0\mathbf{w}_0 = \mathbf{0}$, has a nontrivial solution $\mathbf{w}_0$ (the eigenvector corresponding to the eigenvalue $\lambda = 0$). If $kI - A_0$ is nonsingular for $k \in \mathbb{N}$, i.e., if the numbers $1, 2, 3, \ldots$ are not eigenvalues, then (4) can be solved uniquely; there exists a formal solution $(\mathbf{w}_k)$ and it is a holomorphic solution in K_r.

(b) If the assumptions in (a) hold and if there exist several, let us say p, linearly independent eigenvectors corresponding to the eigenvalue $\lambda = 0$, then p solutions can be calculated from the p eigenvectors, and these solutions are likewise linearly independent, since their values at the point $z = 0$ are the linearly independent eigenvectors just mentioned.

(c) Now let $\lambda \ne 0$ be an eigenvalue of A_0. We make the ansatz

$$\mathbf{w}(z) = z^\lambda \mathbf{u}(z)$$

to obtain a solution of (1). Then the equation

$$\mathbf{w}' = \lambda z^{\lambda-1}\mathbf{u}(z) + z^\lambda \mathbf{u}'(z) = A(z) z^\lambda \mathbf{u}(z)$$

holds if and only if

$$\mathbf{u}' = \left(A(z) - \frac{\lambda I}{z}\right)\mathbf{u}. \tag{8}$$

In other words:

$\mathbf{u}(z) = z^{-\lambda}\mathbf{w}(z)$ *satisfies the same differential equation as* $\mathbf{w}(z)$, *but with* A_0 *replaced by* $A_0 - \lambda I$.

(d) Thus if $\lambda \in \mathbb{C}$ is an eigenvalue of A_0 but $\lambda + k$ is not an eigenvalue of A_0 for any $k \in \mathbb{N}$, then there exists a solution of the form

$$\mathbf{w}(z) = z^{\lambda} \sum_{k=0}^{\infty} \mathbf{w}_k z^k \tag{9}$$

of (1), where $\mathbf{w}_0 \neq \mathbf{0}$ is an eigenvector corresponding to λ and the $\mathbf{w}_k$ can be uniquely determined from the equations

$$((\lambda + k)I - A_0)\mathbf{w}_k = \sum_{j=0}^{k-1} A_{k-j}\mathbf{w}_j \tag{10}$$

$(k = 1, 2, 3, \ldots)$. If several linearly independent eigenvectors are associated with λ, then the solutions obtained from them are also linearly independent.

VII. Generalized Series Expansions. In order to obtain series expansions of solutions that have logarithmic components and hence are not of the form (9), we transform the differential equations as in §22 using $z = \mathrm{e}^s$. For $\mathbf{v}(s) := \mathbf{w}(\mathrm{e}^s)$ we obtain from (1) the differential equation

$$\frac{d\mathbf{v}(s)}{ds} = \mathrm{e}^s A(\mathrm{e}^s)\mathbf{v}(s) = \left(\sum_{k=0}^{\infty} A_k \mathrm{e}^{ks}\right)\mathbf{v}(s). \tag{11}$$

Let P_q be the space of all vector polynomials of degree $\leq q$,

$$\mathbf{p}(s) = \mathbf{p}_0 + \mathbf{p}_1 s + \cdots + \mathbf{p}_q s^q \quad \text{with} \quad \mathbf{p}_k \in \mathbb{C}^n.$$

If we make the ansatz

$$\mathbf{v}(s) = \sum_{k=0}^{\infty} \mathbf{v}_k(s)\mathrm{e}^{ks} \quad \text{with} \quad \mathbf{v}_k(s) \in P_q, \tag{12}$$

then (11) reads

$$\sum_{k=0}^{\infty} (\mathbf{v}_k' + k\mathbf{v}_k)\mathrm{e}^{ks} = \left(\sum_{k=0}^{\infty} A_k \mathrm{e}^{ks}\right)\left(\sum_{l=0}^{\infty} \mathbf{v}_l \mathrm{e}^{ls}\right). \tag{13}$$

Carrying out the multiplication on the right-hand side and equating "coefficients" of the terms e^{ks}, we obtain the system of equations

$$\mathbf{v}_k' + k\mathbf{v}_k = \sum_{j=0}^{k} A_{k-j}\mathbf{v}_j \quad (k = 0, 1, 2, \ldots). \tag{14}$$

In the case $q = 0$, we get the power series expansion considered earlier, and (14) reduces to (4).

As in II, we call a sequence $\mathbf{v} = (\mathbf{v}_k)_{k=0}^{\infty}$ with $\mathbf{v}_k \in P_q$ a formal solution of (11) if (14) holds. Our next objective is to prove the following

VIII. Convergence Theorem. *Let the series* (3) *for* $A(z)$ *be convergent for* $0 < |z| < r$. *Then every formal solution of* (11) *is a holomorphic solution in the half-plane* $R_{\log r} : \operatorname{Re} s < \log r$.

More precisely: If one constructs a formal solution $\mathbf{v}(s)$ *from* $(\mathbf{v}_k)$, *then the resulting series is absolutely convergent in* $R_{\log r}$ *and uniformly convergent in every compact subset of* $R_{\log r}$. *Moreover, the same holds for the series obtained by termwise differentiation and for the series expansion of* $e^s A(e^s)$, *and hence for all the series that appear in* (13). *It follows then from* (14) *that* $\mathbf{v}$ *is a solution of* (11).

The proof that follows is similar to the earlier convergence proof.

IX. The Banach Space H_δ^q. We introduce a norm $|\cdot|_q$ in P_q. If $\mathbf{p}(s) = \mathbf{p}_0 + \cdots + \mathbf{p}_q s^q$, we define

$$|\mathbf{p}(s)|_q := |\mathbf{p}_0| + \cdots + |\mathbf{p}_q|.$$

The space H_δ^q is defined to be the set of all sequences $\mathbf{u} = (\mathbf{u}_k)_0^\infty$, $\mathbf{u}_k \in P_q$, with finite norm

$$\|\mathbf{u}\| := \sum_{k=0}^{\infty} |\mathbf{u}_k|_q \delta^k \quad (\delta > 0 \text{ fixed}).$$

A polynomial $\mathbf{p}(s) \in P_q$ can be identified with the $n \times (q+1)$ matrix of its coefficients $(\mathbf{p}_0, \ldots, \mathbf{p}_q)$. Thus the space H_δ^q is the same as the earlier space H_δ, where, however, here the $\mathbf{u}_k$ are matrices (or $n(q+1)$-dimensional vectors). In particular, by the considerations in I, H_δ^q is a Banach space.

X. The Operators A and J_m. The operator A is defined as in (5),

$$\mathbf{v} = A\mathbf{u} \Longleftrightarrow \mathbf{v}_k = \sum_{j=0}^{k} A_{k-j}\mathbf{u}_j,$$

where, naturally, $\mathbf{u} \in H_\delta^q$ and hence $\mathbf{u}_k, \mathbf{v}_k \in P_q$. The definition of J_m is altered as follows

$$\mathbf{v} = J_m\mathbf{u} \Longleftrightarrow \mathbf{v}_k = \begin{cases} \mathbf{0} & \text{for} \quad k < m, \\ \frac{1}{k} D_k \mathbf{u}_k & \text{for} \quad k \geq m, \end{cases} \tag{15}$$

with

$$D_k\mathbf{p} = \mathbf{p} - \frac{\mathbf{p}'}{k} + \frac{\mathbf{p}''}{k^2} - + \cdots \quad (\mathbf{p} \in P_q). \tag{16}$$

The series has at most $q+1$ summands that are $\neq \mathbf{0}$. The significance of the operator D_k is that it solves a linear differential equation, namely

$$\mathbf{y}(s) := \frac{1}{k} D_k \mathbf{p} \Longrightarrow \mathbf{y}' + k\mathbf{y} = \mathbf{p}, \tag{17}$$

as is easily verified.

(a) The inequality $\|A\mathbf{u}\| \leq C\|\mathbf{u}\|$ holds with the constant C given in IV.(a). The proof is exactly as in IV.(a). Note that if $\mathbf{p} \in P_q$ and B is a constant matrix and if one takes the norm given above for P_q, then

$$B\mathbf{p} = B\mathbf{p}_0 + \cdots + B\mathbf{p}_q s^q \Longrightarrow |B\mathbf{p}|_q = |B\mathbf{b}_0| + \cdots + |B\mathbf{p}_q| \leq |B|\,|\mathbf{p}|_q.$$

(b) It is easy to see that $|\mathbf{p}'|_q \leq q|\mathbf{p}|_q$; hence there exists a constant C_1 such that

$$|D_k\mathbf{p}|_q \leq C_1|\mathbf{p}|_q \quad \text{for all} \quad k \in \mathbb{N},\ \mathbf{p} \in P_q.$$

It follows immediately that

$$\|J_m\mathbf{u}\| \leq \frac{C_1}{m}\|\mathbf{u}\|.$$

(c) *Convergence Proof.* Let $\bar{\mathbf{v}} = (\bar{\mathbf{v}}_k)$ be a formal solution and $0 < \delta < r$. For sufficiently large m, the equation

$$\mathbf{v} = J_m A\mathbf{v} + (\bar{\mathbf{v}}_0, \ldots, \bar{\mathbf{v}}_{m-1}, \mathbf{0}, \mathbf{0}, \ldots)$$

has exactly one solution $\mathbf{v} \in H_\delta^q$. This follows from the fixed point theorem 5.IX if we choose $m > CC_1$, because $\|J_m A\mathbf{v}\| \leq \dfrac{C_1}{m}\|A\mathbf{v}\| \leq \dfrac{CC_1}{m}\|\mathbf{v}\|$.

We will show that $\mathbf{v}$ is likewise a formal solution. First, as before, $\mathbf{v}_k = \bar{\mathbf{v}}_k$ holds for $k < m$. For $k \geq m$,

$$\mathbf{v}_k = \frac{1}{k} D_k \mathbf{p} \quad \text{with} \quad \mathbf{p} = (A\mathbf{v})_k.$$

Therefore, by (17),

$$\mathbf{v}_k' + k\mathbf{v}_k = \mathbf{p} = (A\mathbf{v})_k = \sum_{j=0}^{k} A_{k-j}\mathbf{v}_j,$$

i.e., $\mathbf{v} = (\mathbf{v}_k)$ is likewise a formal solution. It follows now from Lemma XI, given below, that if m is chosen sufficiently large, then the elements $\mathbf{v}_k$ with indices $k \geq m$ are uniquely determined by the $\mathbf{v}_k$ with indices $k < m$ (in the lemma $\mathbf{p} = \mathbf{v}_k$ and $B = kI - A_0$; hence B is nonsingular for large k). Thus $\mathbf{v}_k = \bar{\mathbf{v}_k}$ holds for all k, and hence $\bar{\mathbf{v}} \in H_\delta^q$.

Now suppose $M \subset R_{\log r}$ is a compact set. Clearly, there exists a constant C_M such that

$$\max_M |\mathbf{p}(s)| \leq C_M |\mathbf{p}|_q.$$

Further, there exist constants γ, δ with $0 < \gamma < \delta < r$ such that $M \subset R_{\log\gamma}$. Thus, we have

$$\max_M |k\mathbf{v}_k(s)\mathrm{e}^{ks}| \leq C_M k|\mathbf{v}_k|_q \gamma^k \leq C_M |\mathbf{v}_k|_q \delta^k \quad (k \geq k_0).$$

The uniform convergence of the series $\sum k\mathbf{v}_k(s)\mathrm{e}^{ks}$ in M follows from this inequality. Naturally, then, the series $\sum \mathbf{v}_k \mathrm{e}^{ks}$ and $\sum \mathbf{v}_k' \mathrm{e}^{ks}$ are also uniformly convergent in M, the latter because $|\mathbf{p}'|_q \leq q|\mathbf{p}|_q$. ∎

XI. Lemma. *Let B be a constant $n \times n$ matrix and $\mathbf{q}(s)$ a polynomial vector of degree m. If B is nonsingular, then the differential equation*

$$\mathbf{p}'(s) + B\mathbf{p}(s) = \mathbf{q}(s)$$

has exactly one polynomial solution $\mathbf{p}(s)$, and it has degree m. If B is singular, then there are several polynomial solutions; all have degree $\leq m + r_0$, where r_0 is the multiplicity of the zero $\lambda = 0$ of the characteristic polynomial of B.

Proof. Setting $\mathbf{p}(s) = T\mathbf{u}(s)$, where T is a nonsingular matrix, one obtains the differential equation for $\mathbf{u}$

$$\mathbf{u}' + C\mathbf{u} = \mathbf{h} \quad \text{with} \quad C = T^{-1}BT,\ \mathbf{h} = T^{-1}\mathbf{q}. \tag{18}$$

Let T be chosen such that C has Jordan normal form. If the first Jordan block of C is $J = \lambda I + F$ with r rows and columns (notation as in 18.V), then the corresponding equations read

$$\begin{aligned} u_1' + \lambda u_1 &= h_1 - u_2 \\ &\ \vdots \\ u_{r-1}' + \lambda u_{r-1} &= h_{r-1} - u_r \\ u_r' + \lambda u_r &= h_r. \end{aligned} \tag{18'}$$

These equations can be solved one by one, starting with the last equation.

If $\lambda = 0$, then it follows that $\operatorname{degree} u_r = 1 + \operatorname{degree} h_r \leq 1 + m$, $\operatorname{degree} u_{r-1} \leq 2 + m$, ..., $\operatorname{degree} u_1 \leq r + m$.

If $\lambda \neq 0$, then the homogeneous equation $y' + \lambda y = 0$ has no polynomial solutions, and therefore the nonhomogeneous equation $y' + \lambda y = h$ (h polynomial) has exactly one polynomial solution, namely

$$y = \frac{1}{\lambda} D_\lambda h = \frac{h}{\lambda} - \frac{h'}{\lambda^2} + \frac{h''}{\lambda^3} - + \cdots,$$

cf. (16), (17), and degree y = degree h. Applying this result to (18′) shows that (18′) has exactly one polynomial solution $(u_1, \ldots, u_r)^\top$ and the maximal degree of the u_i is equal to the maximal degree of the h_i.

If B is nonsingular, then all of the Jordan blocks have the form $J = \lambda I + F$ with $\lambda \neq 0$; if B is singular, then there exists at least one block where $\lambda = 0$. Using the solutions of (18′) for the individual blocks, we construct a solution $\mathbf{u}$ of (18) and then obtain $\mathbf{p}$ from the transformation $\mathbf{p} = T\mathbf{u}$. ∎

XII. Construction of a Fundamental System. As in VI.(c) we get

(a) $\mathbf{v}(s)$ is a solution to (11) if and only if $\mathbf{u}(s) := e^{-\lambda s}\mathbf{v}(s)$ is a solution to the corresponding differential equation with $A_0 - \lambda I$ in place of A_0.

(b) As a result, the ansatz

$$\mathbf{v}(s) = e^{\lambda s} \sum_{k=0}^{\infty} \mathbf{v}_k(s) e^{ks} \quad \text{with} \quad \mathbf{v}_k \in P_q, \tag{19}$$

for a solution of (11) leads to a system of equations of the form (14) with $A_0 - \lambda I$ instead of A_0, i.e.,

$$\begin{aligned} \mathbf{v}_0' + (\lambda I - A_0)\mathbf{v}_0 &= \mathbf{0} \\ \mathbf{v}_1' + ((\lambda+1)I - A_0)\mathbf{v}_1 &= A_1\mathbf{v}_0 \\ &\vdots \\ \mathbf{v}_k' + ((\lambda+k)I - A_0)\mathbf{v}_k &= A_k\mathbf{v}_0 + \cdots + A_1\mathbf{v}_{k-1} \\ &\vdots \end{aligned} \tag{20}$$

The first of these equations holds if and only if the function

$$\mathbf{y}(s) := e^{\lambda s}\mathbf{v}_0(s) \tag{21}$$

is a solution of the differential equation

$$\mathbf{y}'(s) = A_0\mathbf{y}(s). \tag{22}$$

By 17.VIII, equation (22) has a fundamental system of solutions of the form (21), where λ is an eigenvalue of A_0 and $\mathbf{v}_0(s)$ is a polynomial. For every such solution $\mathbf{y}(s)$, the function $\mathbf{v}_0(s)$ satisfies the first equation in (20), and the remaining equations of (20) can be solved using Lemma XI. In this way, one obtains a formal solution $(\mathbf{v}_k)$ of the differential equation with $A_0 - \lambda I$ in place of A_0. By the convergence theorem VII, the formal solution is a solution; i.e., the function $\mathbf{v}(s)$ given by (19) is a solution of (11).

(c) *The* n *solutions of* (11) *obtained in this manner form a fundamental system.*

To prove the linear independence of these solutions, we first let $s = ct$ with t real and $\operatorname{Re} c = -1$ and show that every solution $\mathbf{v}(s)$ of the form (19) can be written in the form

$$\mathbf{v}(s) = \mathrm{e}^{\lambda s}(\mathbf{v}_0(s) + o(1)) \quad \text{as } t \to \infty, \quad \text{where } s = ct. \tag{23}$$

This is true, since there exists $t_0 > 1$ such that for every $\mathbf{p}(s) \in P_q$,

$$|\mathbf{p}(ct)| \le |\mathbf{p}|_q |ct|^q \le |\mathbf{p}|_q \mathrm{e}^{t/2} \quad \text{for} \quad t \ge t_0.$$

Thus, because $\operatorname{Re} c = -1$,

$$|\mathbf{v}_k(ct)\mathrm{e}^{kct}| \le |\mathbf{v}_k|_q \mathrm{e}^{(-k+\frac{1}{2})t} \le \mathrm{e}^{-\frac{1}{2}t}|\mathbf{v}_k|_q \delta^{k-1}$$

with $\delta = \mathrm{e}^{-t_0} < 1$, and hence

$$\left|\sum_{k=1}^{\infty} \mathbf{v}_k(ct)\mathrm{e}^{kct}\right| \le \mathrm{e}^{-\frac{1}{2}t} \sum_{k=1}^{\infty} |\mathbf{v}_k|_q \delta^{k-1} = o(1).$$

Now let λ be an eigenvalue of A_0 and $\mathbf{y}^j(s) = \mathrm{e}^{\lambda s}\mathbf{v}_0^j(s)$, $j = 1, \ldots, m$, the m corresponding linearly independent solutions of (22). Obviously, the polynomials $\mathbf{v}_0^j$ are also linearly independent. If the corresponding solutions of (11) are denoted by $\mathbf{v}^j(s)$, then by (23),

$$\sum_{j=1}^{m} c_j \mathbf{v}^j(s) = \mathrm{e}^{\lambda s}(\mathbf{p}(s) + o(1)) \quad \text{with} \quad \mathbf{p}(s) = \sum_{j=1}^{m} c_j \mathbf{v}_0^j(s). \tag{24}$$

If this linear combination is equal to zero, then $\mathbf{p}(s) + o(1) = \mathbf{0}$. Therefore, $\mathbf{p}(s) = \mathbf{0}$, and consequently, $c_j = 0$ for $j = 1, \ldots, m$. This proves the linear independence of solutions of (11) corresponding to a fixed eigenvalue.

Let $\lambda_1, \ldots, \lambda_r$ be the distinct eigenvalues of A_0. A linear combination of all n solutions of (11) can be written in the form

$$\mathrm{e}^{\lambda_1 s}(\mathbf{p}_1(s) + o(1)) + \cdots + \mathrm{e}^{\lambda_r s}(\mathbf{p}_r(s) + o(1)), \tag{25}$$

where the combination of solutions corresponding to a fixed λ_i are summed as in (24). We assume that this linear combination (25) is identically zero and have to show that $\mathbf{p}_1(s) = \cdots = \mathbf{p}_r(s) = \mathbf{0}$. It then follows from what was proved for (24) that all of the coefficients of the linear combination vanish. To show that the $\mathbf{p}_j(s)$ $(1 \le s \le r)$ are zero, it is sufficient to determine c with $\operatorname{Re} c = -1$ such that the numbers $\alpha_j = \operatorname{Re} \lambda_j c$ $(j = 1, \ldots, r)$ are all different. This is certainly possible, since for each j, $y = \operatorname{Re} \lambda_j(-1 + \mathrm{i}x)$ is the equation of a line in the real xy-plane, and these r lines have at most a finite number of points of intersection. Thus we choose an x that is not the first coordinate of a point of intersection and set $c = -1 + \mathrm{i}x$. Further, we number the eigenvalues such that $\alpha_1 > \alpha_2 > \cdots > \alpha_r$. Then $|e^{\lambda_j ct}| = e^{\alpha_j t}$. Thus, setting $s = ct$ in (25) and multiplying by $e^{-\lambda_1 ct}$, we obtain

$$\mathbf{p}_1(ct) + o(1) + e^{(\lambda_2 - \lambda_1)ct}(\mathbf{p}_2(ct) + o(1)) + \cdots = \mathbf{0}.$$

It follows that $\mathbf{p}_1(ct) + o(1) \to \mathbf{0}$ as $t \to +\infty$, because $\operatorname{Re}(\lambda_2 - \lambda_1)c = \alpha_2 - \alpha_1 < 0, \ldots,$ and hence $\mathbf{p}_1(s) = \mathbf{0}$. This same argument gives, one after another, $\mathbf{p}_2 = \cdots = \mathbf{p}_r = \mathbf{0}$. This proves the linear independence. ∎

(d) Suppose $\mathbf{v}(s)$ is a solution of (19). Using the inverse transformation $\mathbf{w}(z) := \mathbf{v}(\log z)$ and rearranging in terms of powers of $\log z$, we get

$$\mathbf{w}(z) = z^\lambda\{\mathbf{h}_0(z) + (\log z)\mathbf{h}_1(z) + \cdots + (\log z)^q \mathbf{h}_q(z)\}, \tag{26}$$

where $\mathbf{h}_j(z)$ are holomorphic functions for $|z| < r$. If the Jordan normal form of A_0 is known, it is easy to estimate how large the powers of $\log z$ are.

As an example, consider

XIII. The case $n = 2$. Suppose λ, μ are the two zeros of the characteristic polynomial $\det(A_0 - \lambda I)$ and $\operatorname{Re}\lambda \le \operatorname{Re}\mu$. Further, let $\mathbf{h}(z)$, $\mathbf{h}_1(z)$, ...denote holomorphic functions for $|z| < r$. Then by VI.(d), there exists a solution $\mathbf{w}(z)$ of the differential equation (1) of the form

$$\mathbf{w}(z) = z^\mu \mathbf{h}(z).$$

Further, there exists a second solution, linearly independent of the first, of the form

$$\tilde{\mathbf{w}} = z^\lambda\{\mathbf{h}_1(z) + (\log z)\mathbf{h}_2(z)\}. \tag{27}$$

If $\lambda - \mu$ is not a whole number, then $\mathbf{h}_2 = \mathbf{0}$. If $\lambda = \mu$ and there exist two linearly independent eigenvectors for this eigenvalue, then again $\mathbf{h}_2 = \mathbf{0}$. This follows from VI.(d).

Thus a logarithmic part arises only if λ is a double zero of the characteristic polynomial with only one eigenvector or if $\mu = \lambda + m$ with $m \in \mathbb{N}$. In the second case, $\mathbf{v}_0$ is constant in equation (20), namely, $\mathbf{v}_0$ is the eigenvector of A_0 corresponding to λ. The solution of (20) gives constant vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{m-1}$. If $(\lambda + m)I - A_0$ is singular, then in general, the mth equation of (20) is not satisfied by a constant $\mathbf{v}_m$; i.e., if $k \ge m$, $\mathbf{v}_k(s)$ is a linear polynomial ($\mathbf{h}_2(z)$ starts with the power z^m). In the case of a double zero $\lambda = \mu$, $\mathbf{v}_0(s)$ is already a linear polynomial, and all remaining $\mathbf{v}_k(s)$ can be uniquely determined from (20) as linear functions of s; cf. Lemma XI.

The function $z^\lambda \mathbf{h}_2(z)$, which multiplies $\log z$ in the formula for $\tilde{\mathbf{w}}$, is itself a solution of (1), hence a multiple of $\mathbf{w}$. Therefore, the second solution $\tilde{\mathbf{w}}$ has the form

$$\tilde{\mathbf{w}}(z) = z^\lambda \mathbf{h}_1(z) + c\mathbf{w}(z)\log z \quad (c \in \mathbb{C}). \tag{27$'$}$$

The proof is left as an exercise: Show that $\tilde{\mathbf{w}}' - A(z)\tilde{\mathbf{w}} = z^{\lambda-1}(\mathbf{h}_3 + \mathbf{h}_4 \log z) = \mathbf{0}$ with $\mathbf{h}_i$ holomorphic, and deduce from this that $\mathbf{h}_4 = \mathbf{0}$.

§ 25. Second Order Linear Equations

As an application of our theory we consider the linear differential equation

$$u'' + a(z)u' + b(z)u = 0 \tag{1}$$

for a scalar function $u(z)$. The usual transformation yields an equivalent first order system

$$\mathbf{w}' = \begin{pmatrix} 0 & 1 \\ -b & -a \end{pmatrix} \mathbf{w} \quad \text{for} \quad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} u \\ u' \end{pmatrix}; \tag{2}$$

cf. (19.2). It is evident from the equation that if the functions $a(z)$, $b(z)$ have at most a pole of first order at z_0, then the point z_0 is either regular or weakly singular.

However, a more general result is obtained if we transform (1) into a system for $w_1 := u(z)$, $w_2 := (z - z_0)u'(z)$. After a simple calculation, we get

$$\mathbf{w}' = \begin{pmatrix} 0 & \dfrac{1}{z - z_0} \\ -(z - z_0)b(z) & \dfrac{1}{z - z_0} - a(z) \end{pmatrix} \mathbf{w} \quad \text{for} \quad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \tag{3}$$

For this system, the point z_0 is still a weakly singular point if $b(z)$ has a pole of second order and $a(z)$ has a pole of at most first order at z_0.

I. Classification of Singularities. Let $a(z)$, $b(z)$ be single-valued holomorphic functions in a punctured neighborhood of $z_0 \in \mathbb{C}$, say $0 < |z - z_0| < r$. The point z_0 is called a *regular point* of (1) if $a(z)$ and $b(z)$ are both holomorphic at z_0. If z_0 is not regular and if $a(z)$ has at most a pole of first order and $b(z)$ has at most a pole of second order, then z_0 is a *weakly singular point* of (1). If neither of the previous cases holds, then the point z_0 is *strongly singular.*

If (1) is transformed into a system of the form (3), then the above classification agrees with the one given in 23.I for systems.

The classification of (1) at the point $z_0 = \infty$ is carried out as in §23 by the change of variables $\zeta = 1/z$. The result is the differential equation

$$\frac{d^2v(\zeta)}{d\zeta^2} + \frac{dv(\zeta)}{d\zeta}\left\{\frac{2}{\zeta} - \frac{1}{\zeta^2}\,a\left(\frac{1}{\zeta}\right)\right\} + \frac{1}{\zeta^4}\,b\left(\frac{1}{\zeta}\right)v(\zeta) = 0 \tag{4}$$

for $v(\zeta) = u(1/z)$. As with systems in 23.III, we make the following definitions: The point $z = \infty$ is called regular or weakly singular or strongly singular for the differential equation (1) if the point $\zeta = 0$ is of the corresponding type for the differential equation (4).

II. Theorem. *Let $a(z)$, $b(z)$ be holomorphic for $|z| > r$. The point $z = \infty$ is regular or weakly singular for* (1) *if and only if $a(z)$ has a zero and $b(z)$ has a multiple zero at $z = \infty$, i.e., if the expansions*

$$a(z) = \frac{a_1}{z} + \frac{a_2}{z^2} + \cdots, \quad b(z) = \frac{b_2}{z^2} + \frac{z_3}{z^3} + \cdots \qquad (|z| > r)$$

hold, or equivalently, if $za(z)$ and $z^2 b(z)$ have finite limits as $z \to \infty$. In particular, the regular case holds if and only if

$$a_1 = 2 \quad \text{and} \quad b_2 = b_3 = 0.$$

This result follows immediately from a classification of the point $\zeta = 0$ with respect to the differential equation (4) with coefficients

$$\tilde{a}(\zeta) = \frac{2}{\zeta} - \frac{1}{\zeta^2} a\left(\frac{1}{\zeta}\right), \qquad \tilde{b}(\zeta) = \frac{1}{\zeta^4} b\left(\frac{1}{\zeta}\right).$$

III. Examples. (a) Consider the differential equation

$$(z+2)z^2 u'' + (z+2)u' - 4zu = 0;$$

the coefficients $a(z)$ and $b(z)$ are the functions

$$a(z) = \frac{1}{z^2}, \qquad b(z) = -\frac{4}{z(z+2)}.$$

The point $z = 0$ is strongly singular, $z = -2$ is weakly singular, $z = \infty$ is weakly singular, all other points of $\mathbb{C}$ are regular.

(b) In the equation

$$(\sin z)u'' - zu' + (e^z - 1)u = 0$$

the coefficients are

$$a(z) = -\frac{z}{\sin z}, \qquad b(z) = \frac{e^z - 1}{\sin z}.$$

Clearly, $z = 0$ is regular. The points $z = k\pi$ with $k = \pm 1, \pm 2, \ldots$ are weakly singular, since $\sin z$ has a simple zero at these points. The point $z = \infty$ is an accumulation point of singularities. Since it is not an isolated singularity, it cannot be classified.

IV. Series Expansions. The Indicial Equation. Let $z = 0$ be a weakly singular point of (1), i.e., let

$$a(z) = \frac{1}{z} \sum_{k=0}^{\infty} a_k z^k, \quad b(z) = \frac{1}{z^2} \sum_{k=0}^{\infty} b_k z^k \quad (0 < |z| < r).$$

Then, in the notation of §24, the matrix in (3) is given by

$$A(z) = \frac{1}{z} \begin{pmatrix} 0 & 1 \\ -z^2 b(z) & 1 - za(z) \end{pmatrix}, \qquad A_0 = \begin{pmatrix} 0 & 1 \\ -b_0 & 1 - a_0 \end{pmatrix}. \tag{5}$$

The characteristic polynomial of A_0 is

$$\boxed{P(\lambda) = \det(A_0 - \lambda I) = \lambda(\lambda + a_0 - 1) + b_0.} \tag{6}$$

The equation $P(\lambda) = 0$ is called the *indicial equation* for (1). Its roots λ_1, λ_2 (the eigenvalues of A_0) are called the *exponents* with respect to equation (1). Suppose $\operatorname{Re}\lambda_2 \le \operatorname{Re}\lambda_1$. Then we know from 24.XIII that there exists a fundamental system of the form

$$u_1(z) = z^{\lambda_1} h(z), \quad u_2(z) = z^{\lambda_2}(h_1(z) + h_2(z)\log z), \tag{7}$$

where h, h_i, are holomorphic functions for $|z| < r$. A logarithmic part arises only if $\lambda_1 - \lambda_2$ is an integer.

In the logarithmic case, the factor $z^{\lambda_2} h_2(z)$ that multiplies the logarithm is itself a solution to (1) and hence is a multiple of u_1. Thus one can write (7) in the form

$$u_1(z) = z^{\lambda_1} h(z), \qquad u_2(z) = u_1(z)\log z + z^{\lambda_2} h_1(z) \tag{7'}$$

in this case; cf. the exercise of 24.XIII.

In calculating the series expansion one can use the recursion formula from §24. However, it is often more convenient to substitute the corresponding series into the equation and calculate the coefficients of powers of z directly. We consider as an example

V. Bessel's Equation

$$\boxed{z^2 u'' + z u' + (z^2 - \alpha^2) u = 0,}$$

where $\alpha \in \mathbb{C}$ is a parameter with $\operatorname{Re}\alpha \ge 0$ (the latter is not a restriction, since α can be replaced by $-\alpha$). The results in I and II imply that the point $z = 0$ is weakly singular and the point $z = \infty$ is strongly singular. Since $a_0 = 1$ and $b_0 = -\alpha^2$, the indicial equation is

$$P_\alpha(\lambda) = \lambda^2 - \alpha^2 = 0 \Longrightarrow \lambda_1 = \alpha, \quad \lambda_2 = -\alpha.$$

The ansatz

$$u(z) = z^\lambda \sum_{k=0}^{\infty} u_k z^k \quad \text{with} \quad u_0 \ne 0$$

leads, after the coefficients of like terms are set equal, to the formulas

$$P_\alpha(\lambda + k) u_k + u_{k-2} = 0 \quad \text{for} \quad k \ge 0 \quad (\text{with } u_{-1} = u_{-2} = 0). \tag{8}$$

The requirement $u_0 \ne 0$ leads to the indicial equation $P_\alpha(\lambda) = 0$.

(a) $\boxed{\lambda = \lambda_1 = \alpha}$ If one sets, for instance, $u_0 = 1$, then the equations (8) are uniquely solvable, since $P_\alpha(\alpha + k) \neq 0$ for $k \geq 1$. The condition $u_{-1} = 0$ implies that $u_1 = u_3 = u_5 = \cdots = 0$. For even $k = 2m$, (8) reads

$$4m(m + \alpha)u_{2m} + u_{2m-2} = 0 \Longrightarrow u_{2m} = \frac{(-1)^m}{4^m m!(\alpha + 1)_m}. \tag{9}$$

Here we have used the notation

$$(x)_m := x(x+1)(x+2)\cdots(x+m-1), \quad (x)_0 = 1,$$

in particular, $(1)_m = m!$. This leads to a solution u_α, defined by

$$u_\alpha(z) = \sum_{m=0}^{\infty} \frac{(-1)^m z^{\alpha+2m}}{4^m m!(\alpha + 1)_m}. \tag{10}$$

(b) $\boxed{\lambda = \lambda_2 = -\alpha, \text{ where } \lambda_1 - \lambda_2 = 2\alpha \text{ is not an integer}}$ This is the "normal case," where no logarithmic part appears. Since $P_\alpha(-\alpha + k) \neq 0$ for $k \in \mathbb{N}$, (8) is again uniquely solvable. The relation (9) holds with $-\alpha$ instead of α, and one obtains the solution $u_{-\alpha}(z)$.

(c) $\boxed{\lambda = \lambda_2 = -\alpha \text{ with } \alpha = n + \tfrac{1}{2} \ (n \geq 0 \text{ an integer})}$ In this case too, the series for $u_{-\alpha} = u_{-n-\frac{1}{2}}$ is well-defined and represents a solution. Thus (contrary to expectation) no logarithmic part arises. In order to clarify matters, we consider the recursion formula (8) with $\lambda = -\alpha$. For the critical index $k = 2\alpha = 2n + 1$ we have $P_\alpha(\lambda + k) = 0$, and hence (8) is given by

$$0 \cdot u_{2n+1} + u_{2n-1} = 0.$$

While this equation is not uniquely solvable for u_{2n+1}, it is nevertheless satisfied, since all u_k with odd index vanish. For even k, (8) is uniquely solvable, and one obtains the above mentioned solution $u_{-\alpha}$.

(d) $\boxed{\lambda = \lambda_2 = -\alpha \text{ with } \alpha = n \in \mathbb{N}}$ In this case the recursion formula (8) breaks down for $k = 2n$, and a logarithmic part appears. To determine the logarithmic part, we apply the procedure in 24.VII. We consider $v(s) := u(\mathrm{e}^s)$, where $u(z)$ is a solution of Bessel's equation, and obtain the equation

$$\frac{d^2 v}{ds^2} + (\mathrm{e}^{2s} - \alpha^2)v = 0$$

for $v(s)$. The ansatz

$$v(s) = \mathrm{e}^{\lambda s} \sum_{k=0}^{\infty} v_k(s)\mathrm{e}^{ks} \qquad (v_k(s) \quad \text{linear})$$

leads to the recursion formula

$$P'_\alpha(\lambda + k)v'_k + P_\alpha(\lambda + k)v_k + v_{k-2} = 0 \ \text{ for } \ k \geq 0, \tag{11}$$

where $v_{-1} = v_{-2} = 0$ and $P'_\alpha(\lambda) = 2\lambda$. Again, it is easily seen that $v_1 = v_3 = v_5 = \cdots = 0$. Beginning as before with $v_0 = 1$, we see that (11) has constant solutions for $0 < k = 2m < 2n$. Indeed, equation (11) reads (with $\alpha = -\lambda = n$)

$$4m(m-n)v_{2m} + v_{2m-2} = 0 \Longrightarrow v_{2m} = \frac{(-1)^m}{4^m m!(1-n)_m} \quad (0 \le m < n).$$

For $k = 2n$, one obtains

$$2nv'_{2n} + 0 \cdot v_{2n} + v_{2n-2} = 0,$$

which has the solution

$$v_{2n} = \alpha_0(s + \beta_0) \text{ with } \alpha_0 = -\frac{v_{2n-2}}{2n} = \frac{2}{4^n n!(n-1)!} \tag{12}$$

and β_0 arbitrary. For $k > 2n$ we write $k = 2n + 2m$ and v_k as

$$v_{2n+2m} = a_m(s + \beta_m) \quad (m \ge 1).$$

Since $P'_n(-n+k) = 2(2m+n)$, $P_n(-n+k) = 4m(m+n)$, (11) shows that

$$4m(m+n)\alpha_m + \alpha_{m-1} = 0 \Longrightarrow \alpha_m = \frac{(-1)^m \alpha_0}{4^m m!(n+1)_m}$$

and

$$2(2m+n)\alpha_m + 4m(m+n)\alpha_m\beta_m + \alpha_{m-1}\beta_{m-1} = 0.$$

If we now replace α_{m-1} by $-4m(m+n)\alpha_m$ and divide by this same number, then

$$\beta_m = \beta_{m-1} - \frac{2m+n}{2m(m+n)} = \beta_{m-1} - \frac{1}{2}\left(\frac{1}{m} + \frac{1}{m+n}\right),$$

and hence

$$\beta_m = \beta_0 - \frac{1}{2}\sum_{j=1}^{m}\left(\frac{1}{j} + \frac{1}{n+j}\right).$$

Denote the partial sums of the harmonic series by

$$H_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}$$

and set $\beta_0 = -\frac{1}{2}H_n$. Then $\beta_m = -\frac{1}{2}(H_m + H_{m+n})$. If we set $\mathrm{e}^s = z$, we obtain the following solution of Bessel's equation for the case $\alpha = n$:

$$\begin{aligned}\tilde{u}_n(z) \quad = \quad & \alpha_0 \log z \sum_{m=0}^{\infty} \frac{(-1)^m z^{2m+n}}{4^m m!(n+1)_m} + \sum_{m=0}^{n-1} \frac{(-1)^m z^{2m-n}}{4^m m!(1-n)_m} \\ & - \frac{\alpha_0}{2} \sum_{m=0}^{\infty} \frac{(-1)^m (H_m + H_{n+m}) z^{2m+n}}{4^m m!(n+1)_m},\end{aligned}$$

where α_0 is given by (12).

(e) $\boxed{\alpha = 0}$ The solution that was just constructed is also valid for $\alpha = 0$. In this case the finite sum disappears. Note that the first equation (11) reads

$$0 \cdot v_0' + 0 \cdot v_0 = 0,$$

while the remaining equations are uniquely solvable. The choice $v_0 = 1$ gives $u_0(z)$, while the choice $v_0 = \alpha_0 s$ gives $\tilde{u}_0$; cf. (a) and (d).

VI. Bessel Functions. Certain linear combinations of the above solutions are considered on technical and historical grounds. The solutions

$$J_\alpha(z) = \frac{u_\alpha(z)}{2^\alpha \Gamma(\alpha+1)} = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\Gamma(m+\alpha+1)} \left(\frac{z}{2}\right)^{2m+\alpha}$$

($\alpha \neq -1, -2, \ldots$) are called *Bessel functions of the first kind*, and

$$\begin{aligned} Y_n(z) &= -\frac{(n-1)!2^n}{\pi} \tilde{u}_n(z) - \frac{\ln 2}{\pi 2^{n-1} n!} u_n(z) \\ &= \frac{2}{\pi} \log \frac{z}{2} J_n(z) - \frac{1}{\pi} \sum_{m=0}^{n-1} \frac{(-1)^m (n-1)!}{m!(1-n)_m} \left(\frac{z}{2}\right)^{2m-n} \\ &\quad - \frac{1}{\pi} \sum_{m=0}^{\infty} \frac{(-1)^m (H_m + H_{n+m})}{m!(n+m)!} \left(\frac{z}{2}\right)^{2m+n} \end{aligned}$$

($n = 0, 1, 2, \ldots$), are *Bessel functions of the second kind.* Note that the functional equation of the gamma function, $z\Gamma(z) = \Gamma(z+1)$, yields $\Gamma(\alpha+m+1) = \Gamma(\alpha+1)(\alpha+1)_m$ in the series for J_α.

A fundamental system for Bessel's equation in terms of Bessel functions is given by

$$\begin{array}{ll} J_\alpha(z),\ J_{-\alpha}(z) & \text{if } \alpha \text{ is not an integer}, \\ J_n(z),\ Y_n(z) & \text{if } \alpha = n \text{ is a nonnegative integer}. \end{array}$$

The linear independence of these solutions follows from 24.XII, but it can also be verified directly without difficulty.

VII. Differential Equations of Fuchsian Type. As in 23.V, equation (1) is called a differential equation of Fuchsian type if finitely many points of $\mathbb{C} \cup \{\infty\}$ are weakly singular and all the remaining points are regular.

Theorem. *Equation* (1) *is a differential equation of Fuchsian type with m singularities $z_1, \ldots, z_m \in \mathbb{C}$ if and only if*

$$a(z) = \sum_{k=1}^{m} \frac{r_k}{z - z_k}, \qquad b(z) = \sum_{k=1}^{m} \left(\frac{s_k}{(z-z_k)^2} + \frac{t_k}{z - z_k} \right), \tag{13}$$

where r_k, s_k, t_k are constants with $|r_k| + |s_k| + |t_k| \neq 0$ and

$$\sum_{k=1}^{m} t_k = 0. \tag{14}$$

The point ∞ is regular if and only if

$$\sum_{k=1}^{m} r_k = 2 \quad \text{and} \quad \sum_{k=1}^{m} (s_k + z_k t_k) = \sum_{k=1}^{m} (2 z_k s_k + z_k^2 t_k) = 0. \tag{15}$$

Proof. If (1) is of Fuchsian type and z_k is a weakly singular point, we have

$$a(z) = \frac{r_k}{z - z_k} + h_k(z), \quad b(z) = \frac{s_k}{(z - z_k)^2} + \frac{t_k}{z - z_k} + l_k(z),$$

where the functions h_k, l_k are holomorphic at the point z_k. Thus, if we denote by $A(z)$, $B(z)$ the differences between the left and right sides of (13), then $A(z)$ and $B(z)$ are holomorphic in $\mathbb{C}$. Further, Theorem II implies that $a(z) \to 0$ and $b(z) \to 0$ as $z \to \infty$, hence also $A(z) \to 0$ and $B(z) \to 0$ as $z \to \infty$. Thus, by Liouville's theorem, the functions $A(z)$, $B(z)$ are constant, in fact, are equal to zero. This proves (13).

By (13), $zb(z) \to \sum t_k$ as $z \to \infty$; however, we also have $zb(z) \to 0$ as $z \to \infty$, and hence (14) holds. The equations (15) also follow from Theorem II if one takes into account the following expansions about the point ∞:

$$\frac{1}{z - z_k} = \frac{1}{z} + \frac{z_k}{z^2} + \frac{z_k^2}{z^3} + \cdots, \qquad \frac{1}{(z - z_k)^2} = \frac{1}{z^2} + \frac{2 z_k}{z^3} + \frac{3 z_k^2}{z^4} + \cdots.$$

■

VIII. Exercise. Show that every Fuchsian differential equation with at most one finite singular point $z_0 \in \mathbb{C}$ has the form

$$u'' + \frac{r}{z - z_0} u' + \frac{s}{(z - z_0)^2} u = 0$$

and give a fundamental system. Find the special cases in which (a) a logarithmic term appears, (b) there is only one singular point.

IX. The Hypergeometric Equation

$$z(z-1)u'' + \{(\alpha + \beta + 1)z - \gamma\}u' + \alpha\beta u = 0 \quad (\alpha, \beta, \gamma \in \mathbb{C})$$

is a differential equation of Fuchsian type with three weak singularities at the points 0, 1, ∞ (compare the exercise for exceptions). The indicial equation corresponding to the point 0 is given by

$$\lambda(\lambda + \gamma - 1) = 0 \Longrightarrow \lambda_1 = 0, \ \lambda_2 = 1 - \gamma.$$

The power series ansatz leads to the *hypergeometric function*

$$F(z;\alpha,\beta,\gamma)=\sum_{k=0}^{\infty}\frac{(\alpha)_k(\beta)_k}{k!(\gamma)_k}z^k \qquad (\gamma\neq 0,-1,-2,\dots)$$

as a solution. Here, $(x)_k := x(x+1)\cdots(x+k-1)$. This series converges for $|z|<1$. If γ is not an integer, then there exists a second solution of the form $u=z^{1-\gamma}h(z)$ with h holomorphic at $z=0$. It is easy to show that

$$u(z)=z^{1-\gamma}F(z;\alpha-\gamma+1,\beta-\gamma+1,2-\gamma)$$

is such a solution. If γ is an integer, then $\lambda_1-\lambda_2$ is also an integer, i.e., we are in the exceptional case, which will not be discussed here.

Exercise. For which values of the parameters is 0 or 1 a regular point?

X. The Legendre Equation

$$(z^2-1)u''+2zu'-\alpha(\alpha+1)u=0 \qquad (\alpha\in\mathbb{C})$$

is likewise a differential equation of Fuchsian type. The points $+1$, -1, ∞ are singular. Thus there exists a series expansion about the point ∞ of the form $u(z)=\sum_{k=0}^{\infty}u_k s^{\lambda-k}$. In this way, one obtains the solution

$$u_\alpha(a)=\sum_{k=0}^{\infty}(-1)^k\frac{\binom{\alpha}{2k}\binom{\alpha}{k}}{\binom{2\alpha}{2k}}z^{\alpha-2k} \qquad (2\alpha\neq 1,3,5,\dots).$$

The series converges for $|z|>1$.

A second solution is $u_{-\alpha-1}$, since only $\alpha(\alpha+1)=(-\alpha-1)(-\alpha)$ appears in the differential equation, where now it must be assumed that $2\alpha\neq -3,-5,-7,\dots$. These two functions form a fundamental system for $2\alpha\neq \pm1,\pm3,\pm5,\dots$; they are, up to a constant factor, the *Legendre functions* of the first and second kind. If $\alpha=n$ is a natural number, then the series for u_α terminates, and one obtains polynomial solutions, the so-called *Legendre polynomials.*

XI. The Confluent Hypergeometric Equation

$$zu''+(\beta-z)u'-\alpha u=0 \qquad (\alpha,\beta\in\mathbb{C})$$

has the point 0 as a weak singularity and the point ∞ as a strong singularity. For $\beta\neq 0,-1,-2,\dots$, the *confluent hypergeometric function* (or *Kummer's function*)

$$K(z;\alpha,\beta)=\sum_{k=0}^{\infty}\frac{(\alpha)_k}{(\beta)_k k!}z^k$$

is a solution. Together with the function $z^{1-\beta}K(z;\alpha-\beta+1,2-\beta)$ it forms a fundamental system if β is not an integer.

The proof is an exercise.

XII. Exercise. Classify the singularities of the differential equation

$$z^2u'' + (3z+1)u' + u = 0.$$

Use the fact that the left side of the differential equation is a total differential in order to find a fundamental system of solutions. Show that every solution can be written in the form $u = h_1 + h_2 \log z$, where the h_i are single-valued and holomorphic in $\mathbb{C} \setminus \{0\}$ but have an essential singularity at the point $z = 0$; cf. in this regard Theorems 22.V and 23.II and the nature of the singularity at $z = 0$.

Transform the differential equation using the change of variables $\zeta = 1/z$, solve the corresponding differential equation by series expansions, and compare the two results.

XIII. Exercise. *Legendre Polynomials.* If $\alpha = n$ is a nonnegative integer, then the solution u_α of the Legendre differential equation given in X is a polynomial of degree n. Show that up to a constant factor, this polynomial is equal to the nth Legendre polynomial

$$P_n(z) = \frac{1}{n!2^n}\frac{d^n}{dz^n}(z^2-1)^n.$$

(Expand the power using the binomial formula and then differentiate.) Give $P_0, \ldots, P_4$ explicitly.

XIV. Exercise. *Linear Differential Equations of Higher Order.* (a) Using

$$w_1 = u,\ w_2 = zu',\ w_3 = z^2u'', \ldots,\ w_n = z^{n-1}u^{(n-1)},$$

transform the differential equation

$$u^{(n)} + a_{n-1}(z)u^{(n-1)} + \cdots + a_0(z)u = 0 \tag{16}$$

into a first order system

$$\mathbf{w}' = A(z)\mathbf{w}.$$

Show that the point $z = 0$ is regular or weakly singular for this system if and only if the functions $z^{n-k}a_k(z)$ are holomorphic in a neighborhood of the origin ($k = 0, \ldots, n-1$). If this is the case and if the a_i are not all holomorphic in a neighborhood of the origin, then, in analogy to I, the point $z = 0$ is called weakly singular for equation (16).

(b) Give a corresponding definition for regularity or weak singularity at a point $z_0 \in \mathbb{C}$.

(c) Classify the singularities of (16) at ∞ as in I by means of the transformation $\zeta = 1/z$. Formulate the conditions on the coefficients for regularity and weak singularity at ∞ in the special case $n = 3$.

Chapter VI Boundary Value and Eigenvalue Problems

§ 26. Boundary Value Problems

I. General. In a boundary value problem for an nth order differential equation

$$u^{(n)} = f(x, u, \ldots, u^{(n-1)}),$$

the n additional conditions that (we expect to) define a solution uniquely are not prescribed at a single point a, as in the case of the initial value problem, but at two points a and b that are the endpoints of the interval $a \le x \le b$ where the solution is considered. Boundary value problems for (real) linear second order equations

$$u'' + a_1(x)u' + a_0(x)u = g(x) \quad \text{for} \quad a \le x \le b \tag{1}$$

are particularly important because of numerous applications in science and technology.

Boundary conditions. The three most common types of boundary conditions for (1) are the *boundary conditions of the*

$$\begin{array}{lrr} \textit{first kind:} & u(a) = \eta_1, & u(b) = \eta_2, \\ \textit{second kind:} & u'(a) = \eta_1, & u'(b) = \eta_2, \\ \textit{third kind:} & \alpha_1 u(a) + \alpha_2 u'(a) = \eta_1, & \beta_1 u(b) + \beta_2 u'(b) = \eta_2. \end{array}$$

Obviously, the first two conditions are special cases of the third, which is also called a *Sturmian boundary condition.* There are also other boundary conditions such as

$$u(a) - u(b) = \eta_1, \qquad u'(a) - u'(b) = \eta_2.$$

When $\eta_1 = \eta_2 = 0$, these are called *periodic boundary conditions* for the following reason: If the coefficients are continuous, periodic functions with period

$l = b - a$ and if $u(x)$ is a solution to (1), then $v(x) := u(x+l)$ is also a solution of the differential equation (it follows from Theorem 19.I that every solution exists in $\mathbb{R}$). If u satisfies the periodic boundary conditions described above, then $v(a) = u(a)$ and $v'(a) = u'(a)$. This implies $u \equiv v$ by the uniqueness theorem for the initial value problem. In other words, u is a periodic function with period l.

Nonexistence and Nonuniqueness. In contrast to the initial value problem, where general existence and uniqueness theorems are available, cases of nonuniqueness or nonexistence arise in very simple boundary value problems. Consider the simplest example: $u'' = 0$. The solutions are linear functions $u(x) = a + bx$. A boundary value problem of first kind is always uniquely solvable, while one of the second kind has no solution if $\eta_1 \neq \eta_2$ and infinitely many solutions if $\eta_1 = \eta_2$.

Exercise. For which values of α_i, β_i is the third boundary value problem for the equation $u'' = 0$ in $[0,1]$ uniquely solvable? In the literature on boundary value problems it is customary to denote the independent variable by x, since in most applications it is a spatial variable. Although we will occasionally encounter complex-valued solutions in the discussion that follows, the independent variable x is always real.

II. Boundary Value Problems of Sturmian Type. We now consider the boundary value problem

$$\boxed{Lu := (p(x)u')' + q(x)u = g(x) \quad \text{in} \quad J = [a,b]} \tag{2}$$

$$\boxed{\begin{aligned} R_1 u &:= \alpha_1 u(a) + \alpha_2 p(a) u'(a) = \eta_1, \\ R_2 u &:= \beta_1 u(b) + \beta_2 p(b) u'(b) = \eta_2 \end{aligned}} \tag{3}$$

under the assumption

$$\text{(S)} \qquad \boxed{\begin{array}{l} p \in C^1(J) \text{ and } q, g \in C^0(J) \text{ are real-valued functions,} \\ p(x) > 0 \text{ in } J \quad \text{and} \quad \alpha_1^2 + \alpha_2^2 > 0, \quad \beta_1^2 + \beta_2^2 > 0. \end{array}}$$

The corresponding *homogeneous boundary value problem* is given by

$$Lu = 0 \ \text{ in } \ J, \qquad R_1 u = R_2 u = 0. \tag{4}$$

Remarks. 1. Self-adjointness. We have written the leading terms in the differential operator L in the so-called *self-adjoint* form $(pu')'$ instead of the form $u'' + a_1 u'$ that appears in (1). The reason for this becomes apparent when one considers the Lagrange identity (5). Using the notation of 28.II, formula

(6) can be expressed in the form $(Lu, v) = (u, Lv)$, which is analogous to the relation $(Ax, y) = (x, Ay)$ for symmetric matrices in $\mathbb{R}^n$. This relation plays a fundamental role in our treatment of the eigenvalue problem in § 28.

Equation (2) is equivalent to the following first order system for (y_1, y_2),

$$y_1' = y_2/p,\ y_2' = -qy_1 - g, \quad \text{where} \quad y_1 = u,\ y_2 = pu', \tag{2'}$$

which is also occasionally used.

This form of (2) also explains the appearance of the factors $p(a)$, $p(b)$ in (3).

2. The relationship between equations (1) *and* (2). Equation (1) can always be transformed into the self-adjoint form (2) by multiplying by the positive factor $p(x) := \exp\left(\int a_1(x)dx\right)$:

$$p(u'' + a_1u' + a_0u) = (pu')' + pa_0u.$$

Conversely, if $p > 0$ and $p \in C^1$, then equation (2) can be written in the form (1) by expanding the derivative

$$(pu')' + qu = pu'' + p'u' + qu.$$

3. Problem (2), (3) is named after Jacques Charles François Sturm (1803–1855), who was born in Geneva and spent most of his life in Paris, where he was professor at the Ecole Polytechnique. Sturm developed the theory of this boundary value problem, partly in collaboration with J. Liouville (1809–1882). The problem also goes under the name *Sturm–Liouville boundary value problem.*

The Lagrange Identity. The great French mathematician Joseph Louis Lagrange (1736–1813) discovered the identity

$$\boxed{vLu - uLv = \{p(x)(u'v - v'u)\}'} \qquad \textit{Lagrange identity} \tag{5}$$

$(u, v \in C^2(J))$ that carries his name. An important consequence of (5) is the relation

$$\int_a^b (vLu - uLv)\, dx = 0 \ \text{ if } \ R_iu = R_iv = 0 \quad (i = 1, 2). \tag{6}$$

Equation (5) is easily verified, and then (6) follows from (5) and the observation that $(u'v - v'u)$ vanishes at both endpoints a and b. It is sufficient to consider the point a. In the case $\alpha_2 = 0$ one has $u(a) = v(a) = 0$; if $\alpha_2 \neq 0$, then $u'(a) = \delta u(a)$, $v'(a) = \delta v(a)$, where $\delta = -\alpha_1/(\alpha_2 p(a))$. In either case, $(u'v - uv')(a) = 0$.

Note that (6) also holds for the periodic boundary condition $u(a) = u(b)$, $u'(a) = u'(b)$ if $p(a) = p(b)$ (Exercise!).

Consequences of Linearity. Since the boundary value problem is linear, the following simple propositions hold:

(a) *A (finite) linear combination* $\sum c_i u_i$ *of solutions* u_i *of the homogeneous problem* (4) *is again a solution of that problem.*

(b) *The difference* $v_1 - v_2$ *of two solutions* v_1, v_2 *of the inhomogeneous problem* (2), (3) *is a solution of the homogeneous problem* (4).

(c) *If* u *is a solution of the homogeneous problem* (4) *and* v *a solution of the inhomogeneous problem* (2), (3), *then the sum* $u + v$ *is a solution of the inhomogeneous problem* (2), (3).

(d) *Let* v^* *be a fixed solution of the inhomogeneous problem* (2), (3). *Then every solution* v *of the inhomogeneous problem can be written in the form*

$$v = v^* + u,$$

where u *runs through all solutions of the homogeneous problem.*

III. Theorem. *Let* $u_1(x)$, $u_2(x)$ *be a fundamental system of solutions to the homogeneous differential equation* $Lu = 0$. *The inhomogeneous boundary value problem* (2), (3) *is uniquely solvable if and only if the homogeneous problem* (4) *has only the zero solution* $u \equiv 0$. *The latter is true if and only if the determinant*

$$\begin{vmatrix} R_1 u_1 & R_1 u_2 \\ R_2 u_1 & R_2 u_2 \end{vmatrix} \neq 0. \tag{7}$$

It follows that condition (7) *is independent of the choice of fundamental system.*

Proof. The first part is an immediate consequence of II.(b). For the second part involving (7) we choose a solution v^* of (2). Then the general solution of this differential equation is given by

$$v = v^* + c_1 u_1 + c_2 u_2 \qquad (c_1, c_2 \in \mathbb{R}).$$

The two boundary conditions (3) lead to a system of two linear equations for c_1, c_2,

$$R_i v = R_i v^* + c_1 R_i u_1 + c_2 R_i u_2 = \eta_i \qquad (i = 1, 2).$$

This system is uniquely solvable if and only if (7) holds. ∎

Theorem III shows that a linear boundary value problem can be easily solved when a fundamental system for $Lu = 0$ is known. When this is the case, a solution v^* of the inhomogeneous equation can be constructed (see 19.VII) and the problem then reduces to the solution of a linear system of two algebraic equations.

Example. (a) $u'' + u = g(x)$ for $0 \le x \le \pi$,

$$R_1 u := u(0) + u'(0) = \eta_1, \qquad R_2 u := u(\pi) = \eta_2.$$

This problem is uniquely solvable for arbitrary η_1, η_2, $g(x)$ because for the fundamental system $u_1 = \cos x$, $u_2 = \sin x$ the determinant in (7) has the value

$$\begin{vmatrix} R_1(\cos x) & R_1(\sin x) \\ R_2(\cos x) & R_2(\sin x) \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ -1 & 0 \end{vmatrix} = 1.$$

(b) In the special case $g(x) = 1$ we may choose $v^*(x) = 1$. Then the general solution of the differential equation is given by

$$v(x) = 1 + c_1 \cos x + c_2 \sin x.$$

Consider the case where $\eta_1 = \eta_2 = 0$. It follows from the boundary values

$$R_1 v = 1 + c_2 + c_1 = 0, \qquad R_2 v = 1 - c_1 = 0$$

that $c_1 = 1$, $c_2 = -2$. Hence the solution of the boundary value problem is

$$v(x) = 1 + \cos x - 2\sin x.$$

(c) If the boundary condition in (a) is changed to

$$R_1 u := u(0) = \eta_1, \qquad R_2 u := u(\pi) = \eta_2,$$

then the determinant in (7) vanishes. Now the homogeneous boundary value problem has infinitely many solutions $u = c \sin x$, while the inhomogeneous problem $u'' + u = 0$, $u(0) = 0$, $u(\pi) = 1$ has no solution.

IV. Fundamental Solutions. Let $J = [a, b]$, let Q be the square $J \times J$ in the $x\xi$-plane, and let

$$\begin{aligned} &Q_1 \quad \text{be the triangle} \quad a \le \xi \le x \le b, \\ &Q_2 \quad \text{be the triangle} \quad a \le x \le \xi \le b. \end{aligned}$$

Note that both triangles are closed and that the diagonal $x = \xi$ belongs to both triangles. A function $\gamma(x, \xi)$ defined in Q is called a *fundamental solution* of the homogeneous differential equation (2) $Lu = 0$ if it has the following properties (recall that $p > 0$):

(a) $\gamma(x, \xi)$ is continuous in Q.

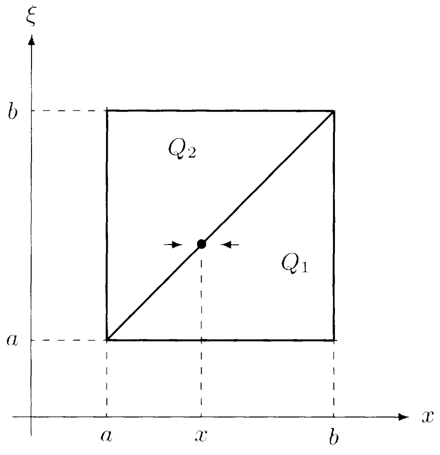
(b) The partial derivatives γ_x, γ_{xx} exist and are continuous in Q_1 and Q_2 (on the diagonal one has to take the one-sided derivatives from the corresponding triangle).

(c) Let $\xi \in J$ be fixed. Then $\gamma(x, \xi)$, considered as a function of x, is a solution to $L\gamma = 0$ for $x \ne \xi$, $x \in J$.

(d) On the diagonal $x = \xi$ the first derivative makes a jump of magnitude $1/p$; i.e.,

$$\gamma_x(x+, x) - \gamma_x(x-, x) = \frac{1}{p(x)} \quad \text{for } a < x < b.$$

Here, $\gamma_x(x+, x)$ is the right-sided derivative of $\gamma(x, \xi)$ with respect to x at the point (x, x) (or, equivalently, the limit of γ_x when the point (x, x) is approached from the right); the left-sided derivative $\gamma_x(x-, x)$ is defined similarly.



Lemma. *Under the assumption* (S) *a fundamental solution exists, but it is not unique.*

Proof. Let $u(x;\xi)$ be the solution to the initial value problem

$$Lu = 0, \quad u(\xi) = 0, \quad u'(\xi) = \frac{1}{p(\xi)} \quad (\xi \in J).$$

Then

$$\gamma(x,\xi) = \begin{cases} 0 & \text{for} \quad a \le x \le \xi \le b, \\ u(x;\xi) & \text{for} \quad a \le \xi \le x \le b \end{cases}$$

is a fundamental solution. Properties (a)–(d) are easily verified.

Now let $g \in C^2(J)$ be a solution of $Lg = 0$, and let $h \in C^0(J)$. Then the function $\gamma_1(x,\xi) = \gamma(x,\xi) + g(x)h(\xi)$ is also a fundamental solution. ∎

Two Examples. Using the notation $a_+ = \max(0, a)$, one obtains

$$u'' = 0 \implies \gamma(x,\xi) = (x-\xi)_+,$$

$$u'' + \lambda^2 u = 0 \implies \gamma(x,\xi) = \frac{1}{\lambda}\sin\lambda(x-\xi)_+ \quad (0 \neq \lambda \in \mathbb{R}).$$

Given a fundamental solution, it is easy to construct a solution of the inhomogeneous differential equation, as the following theorem shows.

V. Theorem. *Let* (S) *from* II *hold. If* $\gamma(x,\xi)$ *is a fundamental solution, then*

$$v(x) = \int_a^b \gamma(x,\xi) g(\xi)\, d\xi \tag{8}$$

belongs to $C^2(J)$ *and is a solution of the inhomogeneous equation* $Lv = g(x)$.

Proof. If the integral (8) is divided into two parts, an integral from a to x and an integral from x to b, and each part is then differentiated, then one obtains

$$\begin{aligned} v'(x) &= \gamma(x,x)g(x) + \int_a^x \gamma_x(x,\xi)g(\xi)\,d\xi \\ &\qquad - \gamma(x,x)g(x) + \int_x^b \gamma_x(x,\xi)g(\xi)\,d\xi \\ &= \int_a^b \gamma_x(x,\xi)g(\xi)\,d\xi. \end{aligned}$$

Applying the same procedure to this last integral and taking into account property IV.(d) of the fundamental solution, one is led to

$$\begin{aligned} v''(x) &= \gamma_x(x+,x)g(x) + \int_a^x \gamma_{xx}(x,\xi)g(\xi)\,d\xi \\ &\qquad - \gamma_x(x-,x)g(x) + \int_x^b \gamma_{xx}(x,\xi)g(\xi)\,d\xi \\ &= \int_a^b \gamma_{xx}(x,\xi)g(\xi)\,d\xi + \frac{g(x)}{p(x)}. \end{aligned}$$

Since $L\gamma = 0$ by IV.(c), it follows that

$$Lv = pv'' + p'v' + qv = \int_a^b L\gamma(x,\xi)g(\xi)\;d\xi + g(x) = g(x).$$

∎

VI. Green's Function. Green's function $\Gamma(x,\xi)$ for the Sturmian boundary value problem (4) $Lu = 0$, $R_1 u = R_2 u = 0$ is defined by the following two properties:

(a) $\Gamma(x,\xi)$ is a fundamental solution of $Lu = 0$.

(b) $R_1\Gamma = R_2\Gamma = 0$ for each $\xi \in J^0 = (a,b)$.

It is assumed that the homogeneous problem (4) has only the trivial solution.

Our construction of Green's function is based on determining two solutions u_1, u_2 of the homogeneous equation $Lu = 0$ satisfying the conditions

$$R_1 u_1 = 0, \quad R_2 u_2 = 0. \tag{9}$$

The function u_1 can be determined as the solution of $Lu = 0$ with the initial values $u(a) = \lambda$, $p(a)u'(a) = \mu$, where $(\lambda,\mu) \neq 0$ satisfies the equation $\alpha_1\lambda + \alpha_2\mu = 0$. Constants λ, μ with this property are easily found. The function u_2 is constructed in a similar manner. If u_1, u_2 are linearly dependent, that is, if

$u_1 = \gamma u_2$, then it follows that u_1 also satisfies the second boundary condition $R_2 u_1 = 0$ and therefore is a nontrivial solution of the homogeneous problem (4). This case has been excluded. Therefore, (u_1, u_2) is a fundamental system of solutions of $Lu = 0$. By Lagrange's identity, the expression

$$c = p(u_1 u_2' - u_1' u_2) \quad \text{is constant and } \neq 0,$$

since the left side of (5) equals 0, and $u_1 u_2' - u_1' u_2$ is the Wronskian of (u_1, u_2), hence $\neq 0$ by the results in 19.II. Green's function can now be determined. It is given by

$$\Gamma(x,\xi) = \frac{1}{c} \cdot \begin{cases} u_1(\xi)u_2(x) & \text{in} \quad Q_1 : a \leq \xi \leq x \leq b, \\ u_1(x)u_2(\xi) & \text{in} \quad Q_2 : a \leq x \leq \xi \leq b. \end{cases} \tag{10}$$

The properties (a), (b), (c) of IV are obvious for Γ. The jump relation (d), $\Gamma_x(x+,x) - \Gamma_x(x-,x) = 1/p(x)$, also follows without difficulty from the relations

$$c\Gamma_x(x+,x) = u_1(x)u_2'(x), \quad c\Gamma_x(x-,x) = u_1'(x)u_2(x).$$

∎

VII. Theorem. *Let* (S) *hold. If the homogeneous boundary value problem* (4) *has only the trivial solution, i.e., if* (7) *holds, then Green's function for* (4) *exists and is unique. It is explicitly given by* (10) *and is symmetric,*

$$\Gamma(x,\xi) = \Gamma(\xi,x). \tag{11}$$

The solution of the "semihomogeneous" boundary value problem

$$Lv = g(x), \qquad R_1 v = R_2 v = 0,$$

which is unique by Theorem III, *is given by*

$$v(x) = \int_a^b \Gamma(x,\xi) g(\xi)\, d\xi. \tag{12}$$

Proof. Theorem V shows that v satisfies the equation $Lv = g$. Since Γ satisfies the homogeneous boundary conditions, this is also true of v, because $v'(x)$ can be obtained by differentiating under the integral sign (see the proof of V), and therefore $R_i \int_a^b \Gamma g\, d\xi = \int_a^b R_i \Gamma g\, d\xi = 0$.

To prove uniqueness of Γ, we consider a second Green's function Γ'. Let

$$v(x) = \int_a^b \Gamma(x,\xi) g(\xi)\, d\xi, \quad w(x) = \int_a^b \Gamma'(x,\xi) h(\xi)\, d\xi$$

with continuous functions g, h. Since $R_i v = R_i w = 0$ $(i = 1, 2)$, equation (6) holds,

$$\int_a^b (vLw - wLv)\, dx = 0.$$

Inserting the expressions for v and w and observing that $Lv = g$, $Lw = h$, one obtains

$$\int_a^b \int_a^b h(x)\Gamma(x,\xi)g(\xi)\,dx\,d\xi = \int_a^b \int_a^b g(x)\Gamma'(x,\xi)h(\xi)\,d\xi\,dx$$

and then, by interchanging x and ξ in the second integral,

$$\iint_Q \{\Gamma(x,\xi) - \Gamma'(\xi,x)\}g(\xi)h(x)\,d\xi\,dx = 0.$$

Since this relation holds for arbitrary continuous functions g, h, the expression in curly braces vanishes; i.e., $\Gamma(x,\xi) = \Gamma'(\xi,x)$. Now set $\Gamma' = \Gamma$. Then this relation shows that Γ is symmetric (this also follows from (10)). But then the same relation implies that $\Gamma = \Gamma'$. ∎

Example. Consider the problem

$$Lu = u'' = 0 \text{ in } [0,1], \quad R_1 u = u(0) = 0, \quad R_2 u = u(1) = 0.$$

Equations (9) are satisfied by $u_1 = x$, $u_2 = x - 1$, for which $c = 1$. Hence

$$\Gamma(x,\xi) = \begin{cases} \xi(x-1) & \text{for} \quad 0 \le \xi \le x \le 1, \\ x(\xi - 1) & \text{for} \quad 0 \le x \le \xi \le 1 \end{cases}$$

is Green's function for this boundary value problem. Since all solutions of $Lu = 0$ are linear functions, the homogeneous problem (4) has only the zero solution.

VIII. Linear and Nonlinear Boundary Value Problems. Theorem VII gives an explicit formula for the solution to a semihomogeneous boundary value problem. Green's function for problem (4) can also be used to solve

(a) *Inhomogeneous boundary value problems.* Given an inhomogeneous problem (2), (3), one looks first for a function $\varphi \in C^2(J)$ that satisfies the boundary conditions $R_i\varphi = \eta_i$ $(i = 1,2)$. This is easily accomplished. Then the ansatz $u = v + \varphi$ in the inhomogeneous boundary value problem leads to the equations

$$Lu = L\varphi + Lv = g, \quad R_i u = R_i\varphi + R_i v = \eta_i \qquad (i = 1,2)$$

which are satisfied if v is a solution to the semi-homogeneous problem

$$Lv = h, \quad R_1 v = R_2 v = 0, \ \text{ with } \ h = g - L\varphi.$$

Assuming (7) holds, the solution v can be found using Theorem VII. Then $u = v + \varphi$ is the solution of the given problem.

The importance of Green's function goes further yet.

(b) *Nonlinear boundary value problems.* Consider the boundary value problem

$$Lu = f(x,u) \text{ in } J = [a,b] \text{ with } R_1 u = R_2 u = 0, \tag{13}$$

where f is continuous in $J \times \mathbb{R}$. We can transform this boundary problem into an integral equation using Green's function, as the following theorem shows.

Theorem. *The function u is of class $C^2(J)$ and a solution of problem* (13) *if and only if u is continuous in J and satisfies the integral equation*

$$u(x) = \int_a^b \Gamma(x,\xi) f(\xi, u(\xi))\, d\xi \quad \text{in } J. \tag{14}$$

The *proof* follows immediately from Theorem VII using $g(x) := f(x, u(x))$. ∎

IX. Existence and Uniqueness Theorem. *Assume that the function $f(x,y)$ is continuous in $[0,1] \times \mathbb{R}$ and satisfies a Lipschitz condition*

$$|f(x,y) - f(x,z)| \le L|y-z| \quad \text{with } L < \pi^2.$$

Then the boundary value problem

$$u'' = f(x,u) \quad \text{for } 0 \le x \le 1, \qquad u(0) = u(1) = 0$$

has a unique solution.

The restriction on L in this theorem is sharp, i.e., the theorem becomes false if $L = \pi^2$. To see this, we consider the examples $f(x,u) = -\pi^2 u$ and $f(x,u) = -\pi^2(u+1)$. In the first case, an infinity of solutions $u(x) = C \sin \pi x$ exists; in the second case there is no solution (proof?).

Proof. We consider (14) as a fixed point equation of the form $u = Tu$, where T is defined by

$$(Tu)(x) = \int_0^1 \Gamma(x,\xi) f(\xi, u(\xi))\, d\xi.$$

Green's function Γ is taken from the example given in VII. We consider this equation in the Banach space $B = C[0,1]$ and apply the contraction principle. If the norm in B is the maximum norm, then T satisfies a Lipschitz condition with Lipschitz constant $L/8$, since $\displaystyle\int_0^1 |\Gamma(x,\xi)|\, dx \le 1/8$. The theorem now follows from the contraction principle 5.IX for $L < 8$.

In order to obtain the general result, we consider the space B^* of all functions $u \in B$ that satisfy an estimate of the form $|u(x)| \le C \sin \pi x$ (this implies in particular $u(0) = u(1) = 0$). In B^* we use a weighted maximum norm

$$||u||^* := \sup_{0<x<1} \frac{|u(x)|}{\sin \pi x} < \infty.$$

The proof that $(B^*, ||\cdot||^*)$ is a Banach space is left to the reader as an exercise. If $u, v \in B^*$, then

$$|f(\xi, u(\xi)) - f(\xi, v(\xi))| \le L|u(\xi) - v(\xi)| \le L||u-v||^* \sin \pi\xi.$$

Therefore,

$$|(Tu - Tv)(x)| \leq L||u - v||^* \int_0^1 |\Gamma(x, \xi)| \sin \pi\xi \, d\xi.$$

We denote the integral term on the right-hand side by $w(x)$. It follows from $\Gamma \leq 0$ and Theorem VII that $w'' = -\sin \pi x$ and $w(0) = w(1) = 0$, and hence $w(x) = (\sin \pi x)/\pi^2$. This implies the estimate

$$|(Tu - Tv)(x)| \leq \frac{L}{\pi^2} ||u - v||^* \sin \pi x.$$

Dividing by $\sin \pi x$ leads to the inequality $||Tu - Tv||^* \leq \dfrac{L}{\pi^2}||u - v||^*$. This shows that T is a contraction for $L < \pi^2$. ∎

Example. We consider the boundary value problem

$$u'' = g(x)\mathrm{e}^u \quad \text{in} \quad [0, 1], \qquad u(0) = u(1) = 0,$$

where $g \in C[0, 1]$, and claim that there is one and only one solution if $0 \leq g(x) \leq L < \pi^2$ in $[0, 1]$.

For a *proof*, we note that $u'' \geq 0$, i.e., u is convex and therefore $u \leq 0$. If $f(x, u) = g(x)\mathrm{e}^u$, then $f_u = g(x)\mathrm{e}^u \leq g(x) \leq L < \pi^2$, since $u \leq 0$ can be assumed. Hence f satisfies a Lipschitz condition with constant $L < \pi^2$. ∎

X. General Linear Boundary Value Problems. The theory developed so far carries over without difficulty to linear differential equations of order n and, more generally, to linear systems of first order. We consider the boundary value problem

$$L\mathbf{y} = \mathbf{f}(x) \quad \text{in} \quad J = [a, b], \qquad R\mathbf{y} = \boldsymbol{\eta}, \tag{15}$$

where

$$L\mathbf{y} := \mathbf{y}' - A(x)\mathbf{y}, \qquad R\mathbf{y} := C\mathbf{y}(a) + D\mathbf{y}(b). \tag{16}$$

Here, A, C, D are $n \times n$ matrices, and $\mathbf{f}$, $\boldsymbol{\eta}$ are n-vectors; A and $\mathbf{f}$ are continuous in J, while C, D, and $\boldsymbol{\eta}$ are constant. All these entities are allowed to be complex-valued.

Three examples. (a) Letting $C = I$, $D = 0$, gives the initial value problem (14.8).

(b) For $n = 2$, the Sturmian boundary value problem (2), (3) is obtained when $\mathbf{y}(x)$ is taken to be $(u, p(x)u')^\top$ and

$$A(x) = \begin{pmatrix} 0 & 1/p \\ -q & 0 \end{pmatrix}, \quad f(x) = \begin{pmatrix} 0 \\ g \end{pmatrix}, \quad C = \begin{pmatrix} \alpha_1 & \alpha_2 \\ 0 & 0 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 0 \\ \beta_1 & \beta_2 \end{pmatrix}.$$

Observe that the system (2′), which is equivalent to (2), has been used here.

(c) The *periodic boundary condition* $\mathbf{y}(a) = \mathbf{y}(b)$ is obtained by putting $C = -D = I$, $\boldsymbol{\eta} = \mathbf{0}$.

XI. Theorem. *Let $Y(x) = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ be a fundamental system ofsolutions of the differential equation $L\mathbf{y} = \mathbf{0}$, where $A(x) \in C(J)$ is complex-valued. Then the following properties are equivalent:*

(a) *The homogeneous problem $L\mathbf{y} = \mathbf{0}$, $R\mathbf{y} = \mathbf{0}$ has only the trivial solution $\mathbf{y} = \mathbf{0}$.*

(b) *The matrix $R(Y) = CY(a) + DY(b)$ is nonsingular,* $\det R(Y) \neq 0$.

(c) *For given $\mathbf{f} \in C^0(J)$, $\boldsymbol{\eta} \in \mathbb{C}^n$, the boundary value problem* (15) *has a unique solution.*

Proof. The general solution of the inhomogeneous equation $Ly = f$ is

$$\mathbf{y} = \mathbf{z} + c_1\mathbf{y}_1 + \cdots + c_n\mathbf{y}_n = \mathbf{z} + Y\mathbf{c}, \tag{17}$$

where $\mathbf{z}(x)$ is a solution of the inhomogeneous equation $L\mathbf{y} = \mathbf{f}$ and $\mathbf{c} = (c_1, \ldots, c_n)^\top \in \mathbb{C}^n$ is arbitrary. The boundary condition reads

$$R(\mathbf{y}) = R(\mathbf{z}) + R(Y)\mathbf{c} = \boldsymbol{\eta}. \tag{18}$$

Hence problem (15) has a unique solution if and only if equation (18) has a unique solution. The rest is simple linear algebra; notice that $\mathbf{z} = 0$ in case (a). ∎

XII. Green's Operator and Green's Function. Let C^1_R be the vector space of all $\mathbf{y} \in C^1(J)$ satisfying the homogeneous boundary condition $R\mathbf{y} = \mathbf{0}$. The semihomogeneous problem

$$L\mathbf{y} = \mathbf{f}, \quad R\mathbf{y} = \mathbf{0} \tag{19}$$

can be formulated in another way: Look for a $\mathbf{y} \in C^1_R$ such that $L\mathbf{y} = \mathbf{f}$. According to Theorem XI, this problem has a unique solution if and only if $\det R(Y) \neq 0$. Under this assumption,

$$L : C^1_R \to C^0(J)$$

becomes a bijective linear map between the two spaces, and the inverse map

$$L^{-1} = G : C^0(J) \to C^1_R$$

is likewise linear and bijective. Expressed in terms of the operator G, the function $\mathbf{y} := G\mathbf{f}$ is the unique solution of the semihomogeneous problem (19). We look for an integral representation of G, i.e., an $n \times n$ matrix $\Gamma(x, \zeta)$ such that

$$\mathbf{y}(x) = (G\mathbf{f})(x) = \int_a^b \Gamma(x, \xi)\mathbf{f}(\xi)\, d\xi \quad \text{for} \quad \mathbf{f} \in C^0(J). \tag{20}$$

The operator G is called *Green's operator* and Γ correspondingly *Green's function* or *Green's matrix* for problem (19).

To find Green's function, we look back to the proof in XI. According to (16.3),

$$\mathbf{z}(x) = \int_a^x Y(x)Y^{-1}(\xi)\mathbf{f}(\xi)\,d\xi$$

is a solution of $L\mathbf{z} = \mathbf{f}$, and the boundary operator R is given by

$$R(\mathbf{z}) = C\mathbf{z}(a) + D\mathbf{z}(b) = D\mathbf{z}(b) = D\int_a^b Y(b)Y^{-1}(\xi)\mathbf{f}(\xi)\,d\xi.$$

Solving (18) with $\boldsymbol{\eta} = 0$ for $\mathbf{c}$ leads to a solution $\mathbf{y}$ of (19) of the form

$$\mathbf{y}(x) = G\mathbf{f} = \mathbf{z}(x) - Y(x)R(Y)^{-1}R(\mathbf{z}).$$

If the above expresssions for $\mathbf{z}(x)$ and $R(\mathbf{z})$ are inserted into this equation, we get an integral representation for G. Since the integral for $\mathbf{z}(x)$ runs only from a to x, we multiply the integrand by

$$c(x,\xi) = \begin{cases} 1 & \text{for} \quad \xi \le x, \\ 0 & \text{for} \quad \xi > x, \end{cases}$$

which allows us to integrate from a to b. In this way an integral representation (20) is obtained, where Green's function is given by

$$\begin{aligned} \Gamma(x,\xi) &= c(x,\xi)Y(x)Y^{-1}(\xi) - Y(x)R(Y)^{-1}DY(b)Y^{-1}(\xi) \\ &= Y(x)\{c(x,\xi) - R(Y)^{-1}DY(b)\}Y^{-1}(\xi). \end{aligned} \tag{21}$$

XIII. Theorem. *Let $Y(x)$ be a fundamental matrix for $L\mathbf{y} = \mathbf{0}$, where A is continuous in J. If* $\det R(Y) \ne 0$, *then Green's function $\Gamma(x,\xi)$ for the boundary value problem* (19) *is unique and given by* (21). *It has the following properties (recall that Q_1, Q_2 are the closed triangles $a \le \xi \le x \le b$ and $a \le x \le \xi \le b$, resp.):*

(a) *If $\Gamma(x,\xi)$ is defined to be $\Gamma(x+,x)$, the limit from the interior of Q_1 on the diagonal $x = \xi$, then Γ is continuous on Q_1; it is continuous Q_2 if it is defined to be $\Gamma(x-,x)$, the limit from the interior of Q_2 on the diagonal $x = \xi$. These limits satisfy the jump relation*

$$\Gamma(x+,x) - \Gamma(x-,x) = I.$$

(b) *For fixed $\xi \in J$, $L\Gamma = 0$ in $J \setminus \{\xi\}$.*

(c) *For fixed $\xi \in J^0 = (a,b)$, $R(\Gamma) = C\Gamma(a,\xi) + D\Gamma(b,\xi) = 0$.*

The properties (a)–(c) *characterize $\Gamma(x,\xi)$ uniquely.*

Proof. The properties in (a) follow immediately from formula (21). The second summand is continuous in the set $Q = J \times J$, while putting $x = \xi$ in the first gives $c(x,x)Y(x)Y^{-1}(x) = I$. Property (b) follows similarly, since in each of the two triangles Q_1, Q_2, Green's function has the form $\Gamma(x,\xi) = Y(x)S$, where the matrix S is independent of x; cf. 15.II.(h). To verify (c), we evaluate $R(\Gamma(x,\xi))$,

$$\begin{aligned} R(\Gamma(x,\xi)) &= R(c(x,\xi)Y(x)Y^{-1}(\xi)) - R\big(Y(x)R(Y)^{-1}DY(b)Y^{-1}(\xi)\big) \\ &= DY(b)Y^{-1}(\xi) - R(Y)R(Y)^{-1}DY(b)Y^{-1}(\xi) = 0. \end{aligned}$$

Hence Γ satisfies (a)–(c). The proof of the uniqueness of Γ uses the fact that a piecewise continuous function $h(\xi)$ that satisfies

$$\int_a^b h(\xi)f(\xi)\,d\xi = 0 \quad \text{for all real-valued} \quad f \in C^0(J)$$

also vanishes in J (except for the points of discontinuity). This remains true if h is complex-valued. One can show after a little thought that a piecewise continuous matrix $H(\xi)$ satisfying

$$\int_a^b H(\xi)\mathbf{f}(\xi)\,d\xi = \mathbf{0} \;\; \text{for all} \;\; \mathbf{f} \in C^0(J) \tag{$\star$}$$

again vanishes in J. Now assume that Γ' is another Green's function. The function $H(\xi) = \Gamma(x,\xi) - \Gamma'(x,\xi)$, with x fixed, satisfies ($\star$). Therefore, since $x \in J$ is arbitrary, $\Gamma = \Gamma'$ holds at all points of continuity of both functions. It remains to show that the function Γ given by (21) is the only function with the properties (a)–(c). If Γ' is another function with these properties and if $\xi \in (a,b)$ is fixed, then $V(x) := \Gamma(x,\xi) - \Gamma'(x,\xi)$ is continuous in J (in particular at $x = \xi$). The equation $LV = 0$, which holds for $x \neq \xi$ by (b), is also true for $x = \xi$ (for a proof let $x \to \xi$ and use Lemma 6.VI). Hence V (that is, each column of V) is a solution of the homogeneous boundary value problem and therefore vanishes identically. ∎

XIV. Remarks. (a) Clearly, the theory developed in X–XIII applies also to the real case, where $A(x)$, C, D, $\mathbf{f}(x)$ are real-valued and L and $G = L^{-1}$ are bijections between the real spaces C^1_R and $C^0(J)$.

(b) The above theory contains the earlier results obtained for the Sturmian boundary value problem. To see this, define $A(x)$, $\mathbf{f}(x)$, C, D as in X.(b). Let L and (Γ_{ij}) $(i,j = 1,2)$ correspond to this matrix problem, while L_s, Γ_s denote the Sturmian operator (2) and its (scalar) Green's function. If $\mathbf{y}$ satisfies $L\mathbf{y} = \mathbf{f}$ and $u := y_1$, then $y_2 = pu'$ and $L_s u = g$. The solution of the semihomogeneous problem (19) is given by

$$\begin{pmatrix} u(x) \\ p(x)u'(x) \end{pmatrix} = \int_a^b \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \begin{pmatrix} 0 \\ g(\xi) \end{pmatrix} d\xi,$$

which implies

$$u(x) = \int_a^b \Gamma_{12}\, g\, d\xi, \qquad pu' = \int_a^b \Gamma_{22}\, g\, d\xi.$$

A comparison with Theorem VII shows that $\Gamma_{12} = \Gamma_s$, $\Gamma_{22} = p\,\partial\Gamma_s/\partial x$. According to the jump relation XIII.(a), Γ_{12} is continuous in Q, while Γ_{22} makes a jump of magnitude 1 on the diagonal. This coincides with the properties IV.(a) and (d) of Γ_s.

XV. Boundary Value Problems with Parameters. Holomorphy in $\boldsymbol{\lambda}$. We consider the operator L_λ depending on a complex parameter λ,

$$L_\lambda \mathbf{y} := \mathbf{y}' - (A(x) + \lambda B(x))\,\mathbf{y},$$

where A and B are continuous in J, together with the earlier boundary operator (without parameter)

$$R\mathbf{y} := C\mathbf{y}(a) + D\mathbf{y}(b).$$

By Theorem 13.III, the solution $\mathbf{y}(x,\lambda)$ to an initial value problem $L_\lambda \mathbf{y} = \mathbf{0}$, $\mathbf{y}(a) = \boldsymbol{\eta}$ ($\boldsymbol{\eta}$ independent of λ) is an entire holomorphic function of $\lambda \in \mathbb{C}$. If n such initial value problems with linearly independent initial values are solved, we obtain a fundamental matrix $Y(x,\lambda)$ that for every $x \in J$ is an entire holomorphic function of λ. It follows that $Y^{-1}(x,\lambda)$, the Wronskian $\det Y(x,\lambda)$, and the boundary operator

$$R(Y(x,\lambda)) = CY(a,\lambda) + DY(b,\lambda)$$

are also entire holomorphic functions of λ. Hence either

(a) $\det R(Y(x,\lambda)) = 0$ for all $\lambda \in \mathbb{C}$; or

(b) $\det R(Y(x,\lambda)) \neq 0$ with the possible exception of a finite or countable set of values $\lambda = \lambda_k$. In the latter case $\lim |\lambda_k| = \infty$.

In case (b), Green's function $\Gamma(x,\xi,\lambda)$ exists for $\lambda \neq \lambda_k$, and it is a holomorphic function of λ. This can be seen from the representation (21), since $R(Y(x,\lambda))^{-1}$ is holomorphic for $\lambda \neq \lambda_k$.

XVI. Exercise. Prove that Green's function for the boundary value problem

$$u'' + \lambda u = 0 \ \text{ in } \ [0,1], \quad u(0) = 0, \quad u(1) = 0$$

is given, for $\lambda > 0$, $\lambda \neq n^2\pi^2$ ($n = 1, 2, 3, \ldots$), by

$$\Gamma(x,\xi;\lambda) = \frac{1}{\sqrt{\lambda}\sin\sqrt{\lambda}} \cdot \begin{cases} \sin\sqrt{\lambda}\,\xi \cdot \sin\sqrt{\lambda}\,(x-1), & 0 \le \xi \le x \le 1, \\ \sin\sqrt{\lambda}\,x \cdot \sin\sqrt{\lambda}\,(\xi-1), & 0 \le x \le \xi \le 1 \end{cases}$$

and that the numbers $\lambda = n^2\pi^2$ belong to the exceptional case, where (4) has a nontrivial solution.

The function $S(z) = \sum\limits_0^\infty (-1)^k z^k/(2k+1)!$ is holomorphic in $\mathbb{C}$, and $\sin z = zS(z^2)$. Show that Γ can be written as

$$\Gamma(x,\xi;\lambda) = \frac{1}{S(\lambda)} \cdot \begin{cases} S(\lambda\xi^2)S(\lambda(x-1)^2)\xi(x-1), & 0 \le \xi \le x \le 1, \\ S(\lambda x^2)S(\lambda(\xi-1)^2)x(\xi-1), & 0 \le x \le \xi \le 1, \end{cases}$$

and that in this form it represents Green's function for all complex $\lambda \neq n^2\pi^2$ ($n = 1, 2, \ldots$). For $\lambda = 0$ the example from VII is obtained. Express Green's function for real $\lambda < 0$ using the hyperbolic sine.

XVII. Exercise. (a) Solve the inhomogeneous boundary value problem

$$u'' + u = \mathrm{e}^x \text{ in } [0,1], \qquad u(0) = u(1) = 0;$$

(a_1) using a fundamental system of the homogeneous and a special solution of the inhomogeneous differential equation; (a_2) using Green's function.

(b) Determine Green's function for the boundary value problem

$$u'' + \frac{1}{4x^2}\, u = 0 \text{ in } [1,2], \qquad u(1) = u(2) = 0.$$

Hint: The substitution $x = \mathrm{e}^t$ helps.

(c) Prove that the boundary value problem

$$u'' = g(x) \sin u \text{ in } [0,1], \quad u(0) = u(1) = 0,$$

where $g \in C[0,1]$ satisfies the inequality $|g(x)| < \pi^2$, has a unique solution.

(d) Find Green's function for $u'' = 0$ in $[0,1]$, $u'(0) = u(1) = 0$.

Supplement I: Maximum and Minimum Principles

Generally speaking, a maximum or minimum principle is a proposition on maxima or minima of solutions of a differential equation or inequality. We consider theorems of this kind for the operator $Lu = (pu')' + qu$. Instead of assuming $p \in C^1$, $u \in C^2$, we make weaker assumptions. If u and pu' belong to $C^1(I)$, where I is an interval, we write $u \in C_p(I)$. As before, $J = [a,b]$ and $J^\circ = (a,b)$.

XVIII. Strong Minimum Principle. (a) *Let $p, q \in C^0(J^\circ)$ and $p > 0$, $q \leq 0$ in J°. Suppose that $u \in C^0(J) \cap C_p(J^\circ)$ satisfies*

$$Lu = (pu')' + qu \leq 0 \quad \text{in } J^\circ, \quad u(a) \geq 0, \quad u(b) \geq 0. \tag{22}$$

Then (i) $u \equiv 0$ *or* (ii) $u > 0$ *in* J°.

(b) *Suppose the assumptions regarding p, q, and u hold in J rather than in J°, in particular $p > 0$ in J. If $u > 0$ in J° and $u(a) = 0$ or $u(b) = 0$, then $u'(a) > 0$ or $u'(b) < 0$, respectively.*

(c) *If $u \geq 0$, then* (a) *and* (b) *remain true without the assumption $q \leq 0$.*

Proof. (a) If $\min u < 0$, then there is an interval $I = [\alpha, \beta]$ with $u(\alpha) = u(\beta) = 0$ and $u < 0$ in I°. Since $qu \geq 0$ in I, we have $(pu')' \leq Lu \leq 0$, i.e., pu' is decreasing in I. On the other hand, there are points in I close to α where u', and hence pu', is negative, and points in I close to β where pu' is positive. This is a contradiction; it shows that $u \geq 0$ in J.

Now assume that $u \not\equiv 0$ is not positive in J°. Then there is a point $\alpha \in J^\circ$ such that $u(\alpha) = 0$ and $u > 0$ in an interval to the right or left of α. We consider only the first case and assume that $I = [\alpha, \alpha + \varepsilon] \subset J^\circ$, $u > 0$ in I°, and $p > \delta > 0$, $q \geq -K$ in I. It follows from $u(\alpha) = 0$, $u'(\alpha) = 0$, and $(pu')' \leq Ku$ in I that

$$u(x) = \int_\alpha^x u'(\xi)\, d\xi \quad \text{and} \quad (pu')(x) \leq \int_\alpha^x Ku(t)\, dt.$$

Let $U(x) = \max\{u(t) : \alpha \leq t \leq x\}$. Then the last inequality yields $p(x)u'(x) \leq K(x-\alpha)U(x)$ and hence $u'(x) \leq c(x-\alpha)U(x)$, with $c = K/\delta$, and so the preceding equation leads to

$$u(x) \leq c(x-\alpha)^2 U(x).$$

Since there are points x with the property $u(x) = U(x)$ arbitrarily close to α, we have again arrived at a contradiction.

(b) The above proof applies also at $\alpha = a$ and shows that the assumption $u(a) = u'(a) = 0$ leads to a contradiction.

(c) Let $q^-(x) = \min\{q(x), 0\} \leq 0$ and $L^-u = (pu')' + q^-u$. Because of $u \geq 0$, we have $q^-u \leq qu$ and hence $L^-u \leq Lu \leq 0$. Now apply (a) to L^-. ∎

In the next theorem $q(x)$ is allowed to assume positive values.

XIX. Theorem. *Let $p, q \in C^0(J)$ and $p > 0$ in J. Assume that $u \in C_p(J)$ satisfies* (22) *and that an "auxiliary function" $h \in C_p(J)$ exists with the properties*

$$Lh \leq 0 \ \text{ in } J^\circ \ \text{ and } \ h < 0 \ \text{ in } J^\circ.$$

Then (i) *$u \equiv 0$, or* (ii) *$u > 0$ in J°, or* (iii) *$u = -\mu h$ with $\mu > 0$.*

Proof. The references to (a), (b), (c) refer to items in the preceding strong minimum principle. The function $d(x) = \operatorname{dist}(x, \partial J) = \min\{x - a, b - x\}$ satisfies $d'(a) = -d'(b) = 1$. If $u \geq 0$ then, according to (c), (i) or (ii) holds. If $\min u < 0$, then the Lipschitz continuity of u together with the inequalities $u(a), u(b) \geq 0$ leads to a lower bound $u(x) \geq -\gamma d(x)$ for some $\gamma > 0$. If (b) is now applied to h, one obtains $h(x) \geq \delta d(x)$ with $\delta > 0$ because $h(a) = 0$ implies $h'(a) > 0$. Both inequalities give $u + \lambda h \geq (-\gamma + \lambda\delta)d > 0$ for large λ. Let

$$\mu = \inf\{\lambda > 0 : u + \lambda h \geq 0 \ \text{ in } \ J\}.$$

Clearly, $\mu > 0$ and $v := u + \mu h \geq 0$. Since $Lv \leq 0$, (c) and (b) imply $v \equiv 0$ or $v \geq \gamma d(x)$, where $\gamma > 0$. The first case leads to (iii), while the second case is incompatible with the minimality of μ. ∎

Example and Remarks. 1. Let $u'' + q(x)u \leq 0$ in $(0, \pi)$, where $q(x) \leq 1$, and suppose $u(0), u(\pi) \geq 0$. Then $u \equiv 0$ or $u > 0$ in $(0, \pi)$ or $u(x) = -\gamma \sin x$, where $\gamma < 0$. Here, case (c) with $h(x) = \sin x$ applies. Note that $u > 0$ in $(0, \pi)$ unless we have equality in all four inequalities of the hypotheses.

2. The eigenvalue case. Note that case (iii) in (c) can occur only if all inequalities reduce to equations, i.e., $Lu = Lh = 0$ and $u(a) = h(a) = u(b) = h(b) = 0$. If just one of these equalities is violated, then $u \equiv 0$ or $u > 0$. The situation is more clearly described in the terminology of eigenvalues (cf. § 27). Case (iii) means that $\lambda_0 = 0$ is the first eigenvalue and $u_0 = h$ a corresponding eigenfunction to the eigenvalue problem $Lu + \lambda u = 0$ in J, $u(a) = u(b) = 0$.

3. Continuity of $q(x)$ is not needed in Theorem XVIII. The proof carries through if q is locally bounded in part (a) and bounded in part (b).

4. The above theorems have their counterparts for second order elliptic differential equations. The alternative $u \equiv 0$ or $u > 0$ in the interior is known as the strong minimum principle, while the statement about positivity of normal derivatives at the boundary, the analogue of XVIII.(b), is called Hopf's lemma. It was discovered in 1952 by Eberhard Hopf. A version of Theorem XIX under the stronger assumption that $h(x) \geq \delta > 0$, which excludes case (iii), has been known for a long time. In the sharper version given here ($h > 0$ only in the interior) it goes back to Walter (1990) and has since been extended to elliptic systems by various authors.

Exercises. (a) Suppose $p > 0$ and $q \leq 0$ in J°. Prove that if $u \in C_p(J^\circ)$ satisfies $Lu \geq 0$ in J° and has a positive maximum in J°, then u is constant. This remains true if u has a negative minimum and $Lu \leq 0$ in J°.

(b) Prove that the strong minimum principle XVIII remains valid for the operator $Mu = a_2(x)u'' + a_1(x)u' + q(x)u$ in non-self-adjoint form if a_2 has the properties of p, a_1 is continuous, and u is of class C^2 (in J° or J, resp.).

Hints. (a) Find an interval where pu' is monotone. (b) Use the transformation in Remark 2 of II.

Supplement II: Nonlinear Boundary Value Problems

In IX a nonlinear boundary value problem was solved using the contraction principle. Here we shall employ Schauder's fixed point theorem and other means to accomplish the same purpose. Since Schauder's theorem does not exclude the possibility of several fixed points, new methods for dealing with the uniqueness problem must be developed. First, we consider the boundary value problem

$$Lu = f(x, u, u') \quad \text{in} \quad J = [a, b], \qquad R_1 u = \eta_1, \qquad R_2 u = \eta_2, \tag{23}$$

where $Lu = (pu')' + qu$ and R_1, R_2 are defined as in (2), (3).

In 1935, the Italian mathematician G. Scorza Dragoni proved the following

XX. Existence Theorem. *Let $f = f(x, z, p)$ be continuous and bounded in $J \times \mathbb{R}^2$, and assume that the homogeneous problem (4) has only the trivial solution. Then the boundary value problem (23) has at least one solution.*

This result is a special case of the next theorem, which deals with the general nonlinear boundary value problem

$$\mathbf{y}'(x) = A(x)\mathbf{y} + \mathbf{f}(x, \mathbf{y}) \text{ in } J = [a, b], \quad R\mathbf{y} = \boldsymbol{\eta}. \tag{24}$$

The corresponding linear problem with $\mathbf{f} = \mathbf{0}$ was treated in Sections X–XIII, and we use the notation introduced there, in particular $R\mathbf{y} = C\mathbf{y}(a) + D\mathbf{y}(b)$. As before, $\mathbf{y}$, A, and $\mathbf{f}$ are allowed to be complex-valued.

Existence Theorem. *If* $\mathbf{f}(x, \mathbf{y})$ *is continuous and bounded in* $J \times \mathbb{C}^n$ *and* $\det R(Y) \neq 0$, *then the boundary value problem* (24) *has at least one solution.*

Proof. As in VIII.(a), one reduces the problem to the semihomogeneous case $\boldsymbol{\eta} = \mathbf{0}$ by writing the solution in the form $\mathbf{y} = \mathbf{y}_0 + \mathbf{z}$, where $\mathbf{y}_0 \in C^1(J)$ satisfies $R\mathbf{y}_0 = \boldsymbol{\eta}$, and considering the corresponding semihomogeneous problem for $\mathbf{z}$. According to XII, in particular equation (20), $\mathbf{y}$ is a solution of (24) with $\boldsymbol{\eta} = \mathbf{0}$ if and only if $\mathbf{y}$ is a continuous solution of the integral equation

$$\mathbf{y} = T\mathbf{y} \text{ with } (T\mathbf{y})(x) := \int_a^b \Gamma(x, \xi)\mathbf{f}(\xi, \mathbf{y}(\xi))\, d\xi. \tag{25}$$

Schauder's fixed point theorem is applied in the Banach space $B = C(J, \mathbb{C}^n)$ with the maximum norm $\|\mathbf{y}\| = \max\{|\mathbf{y}(x)| : x \in J\}$. For $\mathbf{y} \in B$, $\mathbf{f}(x, \mathbf{y}(x))$ is continuous in J; hence $\mathbf{v} = T\mathbf{y}$, which is a solution of

$$\mathbf{v}' = A(x)\mathbf{v} + \mathbf{f}(x, \mathbf{y}(x)), \tag{26}$$

belongs to $C_R^1 \subset B$; i.e., $T(B) \subset B$.

The functions $\mathbf{f}$, A, and Γ are bounded, say, $|\mathbf{f}|, |A|, |\Gamma| \le c$. Therefore, $\mathbf{v} = T\mathbf{y}$ satisfies by (25) and (26):

$$\begin{aligned} |\mathbf{v}(x)| &= |(T\mathbf{y})(x)| \le c^2(b-a) =: c_1, \\ |\mathbf{v}'(x)| &\le cc_1 + c. \end{aligned} \tag{27}$$

These two estimates show that $T(B)$ is bounded and equicontinuous and hence relatively compact in B. It remains to show that T is continuous. Let $(\mathbf{y}_k)$ be a sequence that converges in B, i.e., uniformly in J, to $\mathbf{z} \in B$. Then $\mathbf{f}(x, \mathbf{y}_k(x)) \to \mathbf{f}(x, \mathbf{z}(x))$ uniformly in J because $\mathbf{f}$ is uniformly continuous in bounded subsets of $J \times \mathbb{C}^n$. Since Γ is bounded, $T\mathbf{y}_k \to T\mathbf{z}$ uniformly in J, which shows that T is continuous. Now Schauder's theorem can be applied. ∎

XXI. Upper and Lower Solutions. If the nonlinearity f in the boundary value problem (23) is unbounded, but upper and lower solutions exist, then existence of a solution can be established by reduction to the case where f is bounded. We will demonstrate this important method by proving the existence of a solution to the first boundary value problem

$$Lu = f(x, u) \text{ in } J = [a, b], \quad u(a) = \eta_1, \quad u(b) = \eta_2, \tag{28}$$

where $Lu = (pu')' + qu$. Let $v, w \in C^2(J)$. The function v is a lower solution and w an upper solution for (28) if

$$Lv \geq f(x, v), \quad v(a) \leq \eta_1, \quad v(b) \leq \eta_2,$$
$$Lw \leq f(x, w), \quad w(a) \geq \eta_1, \quad w(b) \geq \eta_2;$$

note that in the differential inequality the direction of the inequality is reversed.

Existence Theorem. *Assume that p satisfies* (S) *and that v is a lower solution and w an upper solution with $v \leq w$. If $f(x, z)$ is continuous in the region $K = \{(x, z) : a \leq x \leq b,\ v(x) \leq z \leq w(x)\}$, then there exists a solution u of the boundary value problem* (28) *between v and w.*

Proof. We write the differential equation in the form $(pu')' = g(x, u)$, where $g(x, z) = f(x, z) - q(x)z$. We then extend g as a continuous function to the strip $J \times \mathbb{R}$ in such a way that $g(x, z)$ is constant in z outside K. Let G be the extension. By Theorem XX, the boundary value problem for $(pu')' = G(x, u)$ has a solution, since the homogeneous problem $(pu')' = 0$, $u(a) = u(b) = 0$ has only the zero solution. We have to show that $\operatorname{graph} u \subset K$. Assume to the contrary that, e.g., u is not $\leq w$. Then $\varphi = w - u$ is negative in an open interval I° and vanishes at its endpoints. For $x \in I^\circ$,

$$(p'\varphi')' = (pw')' - (pu')' \leq G(x, w) - G(x, u) = 0.$$

This is a contradiction to Theorem XVIII. Hence $u \leq w$, and a similar argument shows that $v \leq u$. It follows that $\operatorname{graph} \varphi \subset K$ and therefore $g(x, u(x)) = G(x, u(x))$; i.e., u is a solution to problem (28). ∎

Example. (cf. IX). The problem

$$u'' = f(x, u) \ \text{ for } \ 0 \leq x \leq 1, \quad u(0) = \eta_1, \quad u(1) = \eta_2$$

has at least one solution if f is continuous in $[0, 1] \times \mathbb{R}$ and satisfies

$$|f(x, u)| \leq A + B|u| \ \text{ with } \ B < \pi^2.$$

Proof. Let $w = -v = \alpha \cos \gamma \left(x - \frac{1}{2}\right)$, where $B < \gamma^2 < \pi^2$. Then

$$w(x) \geq w(0) = w(1) = \alpha \cos(\gamma/2) > 0 \ \text{ in } \ [0, 1].$$

Furthermore, $w'' \leq f(x, w)$ and $v'' \geq f(x, v)$ if

$$\left(x - \frac{1}{2}\right) - \gamma^2 w(x) \leq -A - Bw(x).$$

Choose $\alpha > 0$ so large that $A < (\gamma^2 - B)w(0)$ and $|\eta_i| \leq w(0)$. ∎

The following estimate uses a family $w_\lambda(x)$ of upper solutions that is increasing in $\lambda \in [\alpha, \beta]$. It was established by A. McNabb (1961) for elliptic equations and is related to "Serrin's sweeping principle."

XXII. Estimation Theorem. *Assume that $f(x,z)$ is continuous in $J \times \mathbb{R}$ and locally Lipschitz continuous in z; that $u \in C^2(J)$, $w_\lambda \in C^2(J)$ for $\lambda \in \Lambda = [\alpha, \beta]$; and that w_λ and w'_λ are continuous in $(x,\lambda) \in J \times \Lambda$. Suppose further that w_λ is increasing in λ. Let*

(a) *$Lu \geq f(x,u)$ in J and $Lw_\lambda \leq f(x,w_\lambda)$ in J for $\alpha \leq \lambda \leq \beta$,*

(b) *$u(a) \leq w_\lambda(a)$ and $u(b) \leq w_\lambda(b)$ for $\alpha \leq \lambda \leq \beta$,*

where we exclude the case that equality holds everywhere in (a) *and* (b)*. Then*

$$u(x) \leq w_\beta(x) \text{ in } J \text{ implies } u(x) \leq w_\alpha(x) \text{ in } J.$$

Proof. There is a minimal $\mu \in \Lambda$ such that $u(x) \leq w_\mu(x)$ in J. Assume that the assertion is false, i.e., that $\alpha < \mu$. The function $\varphi = w_\mu - u$ satisfies

$$L\varphi = Lw_\mu - Lu \leq f(x,w_\mu) - f(x,u) = c(x)\varphi.$$

Here $c(x) = [f(x,w_\mu) - f(x,u)]/(w_\mu - u)$ for $w_\mu \neq u$, and $c(x) = 0$ otherwise. Our assumption on f implies that c is bounded. Using the notation $L^*\varphi = (p\varphi')' + (q-c)\varphi$, we get

$$\varphi \geq 0 \text{ and } L^*\varphi \leq 0 \text{ in } J.$$

Theorem XVIII.(c) with Remark 3 in XIX implies $\varphi(x) > 0$ in J°, since the case $\varphi \equiv 0$ is excluded by our assumption in (a), (b). Now we shall derive a contradiction by showing that μ is not minimal, more precisely, that there is an index $\nu \in (\alpha, \mu)$ with

$$w_\nu \geq u \iff w_\mu - w_\nu \leq \varphi. \tag{$*$}$$

If $\varphi(a) > 0$, $\varphi(b) > 0$, then $\varphi(x) \geq \varepsilon > 0$ in J, and $(*)$ is easily established. Now let $\varphi(a) = \varphi(b) = 0$, whence $\varphi'(a) \geq \delta$, $\varphi'(b) \leq -\delta$ by XVIII.(b). The function $\psi = w_\mu - w_\nu \geq 0$ satisfies $\psi(a) = \psi(b) = 0$. If ν is sufficiently close to μ, then $\psi'(a) \leq \delta/2$, $\psi'(b) \geq -\delta/2$. Hence $\psi(x) \leq \varphi(x)$ for $a \leq x \leq a+\varepsilon$ and $b-\varepsilon \leq x \leq b$ ($\varepsilon > 0$ and small). In the interval $[a+\varepsilon, b-\varepsilon]$ we have $\varphi \geq \gamma > 0$. Moving ν still closer to μ if necessary, we may assume that $(*)$ holds in this interval and hence in all of J. The cases $\varphi(a) = 0 < \varphi(b)$ and $\varphi(a) > 0 = \varphi(b)$ are treated similarly. This reasoning shows that the inequality $(*)$ holds in all cases. The theorem follows. ∎

Remark. The example $Lu = u'' + u$, $J = [0,\pi]$, $f \equiv 0$, $u = \sin x$, $w_\lambda = \lambda \sin x$ for $0 \leq \lambda \leq 1$ shows that the theorem is false without the provision regarding equality in (a) and (b).

Example. The stationary Logistic Equation. The parabolic logistic equation for $u = u(t,x)$, $x \in \mathbb{R}^n$,

$$u_t = \Delta u + u(b - cu) \text{ in } (0,\infty) \times D,$$

models the density u of a population that is not evenly distributed in $D \subset \mathbb{R}^n$ and subject to diffusion. If u depends only on t, the equation reduces to the logistic equation (1.16) $u' = u(b - cu)$. If u depends only on x, then the stationary logistic equation $\Delta u + u(b - cu) = 0$ is obtained.

We consider the case $n = 1$ and look for *positive solutions* of the boundary value problem

$$u'' + u(b - cu) = 0 \text{ in } [0, 1], \quad u(0) = u(1) = 0,$$

where b and c are positive numbers (or functions of x). If $0 < b \leq \pi^2$, then the family $w_\lambda = \lambda \sin \pi x$ ($\lambda \geq 0$) satisfies

$$w''_\lambda + w_\lambda(b - cw_\lambda) = w_\lambda(b - \pi^2 - cw_\lambda) \leq 0.$$

A positive solution u satisfies $u(x) \leq w_\beta(x) = \beta \sin \pi x$ for some $\beta > 0$. The above theorem with $\Lambda = [0, \beta]$ shows that $u(x) \leq w_0 = 0$. Hence a positive solution does not exist. Now let $b > \pi^2$. It is easily seen that $v = \varepsilon \sin \pi x$ (ε small) is a lower solution and $w = \text{const.} \geq b/c$ an upper solution. Theorem XXI shows that there exists a positive solution u between v and w, i.e., $\varepsilon \sin \pi x \leq u \leq \max b/c$. The uniqueness of this solution follows from our next theorem.

XXIII. Uniqueness Theorem. *The function $f(x, z)$ is assumed to be continuous and nonnegative in $J \times [0, \infty)$, locally Lipschitz continuous in z, and such that $f(x, z)/z$ is strictly decreasing in $z > 0$ for each $x \in J$. Then the boundary value problem*

$$u'' + f(x, u) = 0 \quad in \quad J, \quad u(a) = \eta_1 \geq 0, \quad u(b) = \eta_2 \geq 0$$

has at most one solution that is positive in J°.

Proof. Let u, v be two positive solutions and $w_\lambda = \lambda u$, $\lambda \geq 1$. We show that $v \leq \beta u = w_\beta$ for some $\beta > 1$. If $\eta_1 > 0$, $\eta_2 > 0$, this is easily established. If $\eta_1 = 0$, then $u'(a) > 0$, since u is concave down ($f \geq 0$). It follows that $v'(a) < \beta u'(a)$ for β large. Similarly, $\eta_2 = 0$ implies that $|v'(b)| < \beta |u'(b)|$. A moment's reflection shows that $v \leq \beta u$ in all of J for large values of β. It follows from the monotonicity of $f(x, z)/z$ that the functions $w_\lambda = \lambda u$ satisfy the conditions of Theorem XXII for $\lambda \in \Lambda = [1, \beta]$. According to that theorem, $v \leq w_1 = u$. Since the inequality $u \leq v$ can be proved in exactly the same way, it follows that $u = v$. ■

We close this section with some considerations regarding

XXIV. Boundary Value Problems in the Sense of Carathéodory. The theory of the general linear boundary value problem developed in X–XV carries over to solutions in the sense of Carathéodory without change. It is assumed that the components of $A(x)$, $B(x)$, and $\mathbf{f}(x)$ belong to $L(J)$. A solution $\mathbf{z}(x)$ satisfies the differential equation $L\mathbf{y} = \mathbf{f}(x)$ a.e. in $J = [a, b]$. Theorem 10.XII.(b) guarantees existence and uniqueness for the initial value problem, in particular extistence of a fundamental matrix $Y(x)$. The special case X.(b) leads to the Sturmian problem (2), (3), where the general assumption (S) is now replaced by the weaker assumption

(S$_C$) $\boxed{p \text{ is measurable, } p(x) > 0 \text{ a.e. in } J, \text{ and } q, g, 1/p \in L(J).}$

A solution u of $Lu = g$ is now a function with the properties that $y_1 = u$ and $y_2 = pu'$ belong to $AC(J)$ and satisfy the system (2′) a.e. in J. Note that we do not require that p or u' be continuous, only that pu' be absolutely continuous. In the boundary condition (3), the value $p(a)u'(a)$ is understood to be the value of y_2 at the point a. Green's function Γ exists if (4) has only the zero solution; it can be constructed as in VI.

Example. $Lu = (\sqrt{x}\,u')'$. The functions $u = 1$ and $u = \sqrt{x}$ are solutions of $Lu = 0$ for $x \geq 0$. For example, the boundary value problem

$$Lu = 1 \text{ in } J = [0,1], \quad (\sqrt{x}\,u')(0) = 1, \quad u(1) = 1$$

has the solution $u = \frac{2}{3}\,x^{3/2} + 2\sqrt{x} - \frac{8}{3}$. Green's function for the corresponding homogeneous problem is constructed as in (9), (10) with $u_1(x) = 1$, $u_2(x) = 1 - \sqrt{x}$, $c = -1/2$:

$$\Gamma(x,\xi) = -\frac{1}{2}\begin{cases} 1 - \sqrt{x} \text{ in } Q_1: & 0 \leq \xi \leq x \leq 1, \\ 1 - \sqrt{\xi} \text{ in } Q_2: & 0 \leq x \leq \xi \leq 1. \end{cases}$$

Exercise. We consider the equation $Lu = (x^\alpha u')' = 0$ in $J = [0,1]$, where $\alpha < 1$ (note that for $\alpha \geq 1$ we have $1/p \notin L(J)$).

(a) Find the solution with boundary values (a$_1$) $u(0) = \eta_1$, $u'(1) = \eta_2$ and (a$_2$) $(x^\alpha u')(0) = \eta_1$, $u(1) = \eta_2$.

(b) Construct Green's function for both cases.

(c) Find the solution of the problem $Lu = 1$ in J, $(x^\alpha u')(0) = 1$, $u'(1) = 0$.

XXV. Strong Minimum Principle. *Let $p > 0$ and $q \leq 0$ a.e. in $J = [a,b]$. Suppose the function $u \in C^0(J)$ with $u, pu' \in AC_{\text{loc}}(J^\circ)$ satisfies*

$$Lu = (pu')' + qu \leq 0 \text{ a.e. in } J, \quad u(a) \geq 0, \quad u(b) \geq 0.$$

Then $u \geq 0$ in J. If $q \in L_{\text{loc}}(J^\circ)$, then $u \equiv 0$ or $u > 0$ in J°.

The method of proof used in XVIII carries over. If u is negative in an interval I°, then pu' is decreasing in I°, which easily leads to a contradiction. Hence $u \geq 0$. If $u(\alpha) = 0$, $u > 0$ in $[\alpha, \alpha + \varepsilon]$, then $(pu')(\alpha) = 0$, and as in the proof in XVIII, $p(x)u'(x) \leq U(x)Q(x)$, $Q(x) = \int_\alpha^x |q(t)|\,dt$, and, with $P(x) = \int_\alpha^x (1/p)\,dt$, $u(x) \leq U(x)Q(x)P(x)$. A contradiction is obtained as before, since $P(x)Q(x) \to 0$ as $x \to \alpha$. ∎

Remark. The assertion that $u \equiv 0$ or $u > 0$ in J° is false without further assumptions on p and q. A simple counterexample is $u = x^2$ in $[-1,1]$ with $p = 1$, $q = -2/x^2$ or with $p = x^2$, $q = -6$. But Theorems XVIII and XIX can be generalized considerably for C-solutions; cf. Walter (1992).

§ 27. The Sturm–Liouville Eigenvalue Problem

I. Formulation of the Problem. The Sturm–Liouville eigenvalue problem is of the form

$$Lu + \lambda r(x)u = 0 \text{ in } J = [a, b], \quad R_1 u = R_2 u = 0, \tag{1}$$

where L and R_1, R_2 are the operators defined in (26.2–3)

$$Lu := (p(x)u')' + q(x)u, \tag{2}$$

$$R_1 u := \alpha_1 u(a) + \alpha_2 p(a)u'(a), \quad R_2 u := \beta_1 u(b) + \beta_2 p(b)u'(b). \tag{3}$$

This is a homogeneous boundary value problem for the differential equation

$$(pu')' + (q + \lambda r)u = 0 \tag{4}$$

depending on a real parameter λ (all functions are real-valued). In the eigenvalue problem one is interested in those cases where (1) is *not* uniquely solvable, that is, where not only the trivial solution $u \equiv 0$, but also a nontrivial solution $u(x) \not\equiv 0$, exists. This exceptional case does not hold for all λ, but only for certain values of λ called the *eigenvalues* of the problem. Thus an eigenvalue is a number λ for which (1) has a nontrivial solution u; this solution is called an *eigenfunction* corresponding to the eigenvalue λ. An eigenfunction is determined only up to a constant factor, since obviously $c \cdot u(x)$ $(c \neq 0)$ is also an eigenfunction. If (1) has k, but not $k+1$, linearly independent eigenfunctions for a given eigenvalue, then the eigenvalue is said to have *multiplicity* k; if $k = 1$, the eigenvalue is called *simple*.

Example.

$$u'' + \lambda u = 0, \quad u(0) = u(\pi) = 0.$$

It is easy to see that in the cases $\lambda = 0$ (general solution $u = c_1 + c_2 x$) and $\lambda = -\mu^2 < 0$ (general solution $u = c_1 e^{\mu x} + c_2 e^{-\mu x}$) the boundary value problem does not have a nontrivial solution. In the case $\lambda = \mu^2 > 0$ (general solution $u = c_1 \cos \mu x + c_2 \sin \mu x$) the boundary conditions are satisfied if $c_1 = 0$ and $\sin \mu\pi = 0$. Thus we obtain the

$$\textit{eigenvalues} \qquad \lambda_n = n^2 \quad (n = 1, 2, 3, \ldots)$$

and the corresponding

$$\textit{eigenfunctions} \qquad u_n(x) = \sin nx.$$

An arbitrary function $\varphi \in C^1(J)$ with $\varphi(0) = \varphi(\pi)$ has an "eigenfunction expansion"

$$\varphi(x) = \sum_{n=1}^{\infty} a_n \sin nx.$$

This follows from a well-known theorem about Fourier series: If φ is extended to the interval $[-\pi, \pi]$ as an odd function, then the Fourier expansion of φ has only sine terms. This simple example introduces two fundamental questions in eigenvalue theory.

Existence of eigenvalues. Under what conditions do eigenvalues exist? Are there infinitely many eigenvalues? Is there an asymptotic growth law (such as $\lambda_n \sim n^2$ as $n \to -\infty$)?

Eigenfunction expansion. Under what conditions can an arbitrary function be expanded into a series in terms of the eigenfunctions,

$$\varphi(x) = \sum a_n u_n(x)?$$

We will develop a theory that gives a satisfactory answer to both questions under the following "Sturm–Liouville assumption":

$$\text{(SL)} \qquad \boxed{\begin{array}{l} p(x) \in C^1(J); \quad q(x), r(x) \in C^0(J); \\ p(x) > 0,\ r(x) > 0 \ \text{ in } J; \quad \alpha_1^2 + \alpha_2^2 > 0,\ \beta_1^2 + \beta_2^2 > 0. \end{array}}$$

II. Existence Theorem. *Under the assumption* (SL) *the eigenvalue problem* (1) *has infinitely many simple real eigenvalues*

$$\lambda_0 < \lambda_1 < \lambda_2 < \cdots, \qquad \lambda_n \to +\infty \ \text{ as } \ n \to \infty$$

and no other eigenvalues. The eigenfunction $u_n(x)$ corresponding to λ_n has exactly n zeros in the open interval $J^\circ = (a, b)$. Between two successive zeros of u_n and also between a and the first zero and between the last zero and b there is exactly one zero of u_{n+1} (zeros on the boundary of J, which occur if $\alpha_2 = 0$ or $\beta_2 = 0$, are not counted).

III. Expansion Theorem. *The eigenfunctions can be normalized in such a way that*

$$\int_a^b r(x) u_n^2(x)\, dx = 1 \qquad (n = 0, 1, 2, \ldots).$$

They then form an orthonormal set of functions, i.e., one has additionally that

$$\int_a^b r(x) u_m(x) u_n(x)\, dx = 0 \ \text{ for } \ m \neq n.$$

Each function $\varphi(x) \in C^2(J)$ that satisfies the homogeneous boundary conditions can be expanded in terms of the eigenfunctions in a series

$$\varphi(x) = \sum_{n=0}^{\infty} c_n u_n(x)$$

that converges absolutely and uniformly in J. This series is called the Fourier series of φ (with respect to the u_n). The Fourier coefficients c_n are given by

$$c_n = \int_a^b r(x)\varphi(x)u_n(x)\,dx.$$

There are several methods for proving these theorems. We will first present a method that goes back to Prüfer (1926). In this regard, we note that a solution $\mathbf{y}(x)$ of a nonautonomous system $\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$ can also be considered as a parametric representation of a solution curve (or trajectory) in the phase space $\mathbb{R}^n$; this point of view is adopted extensively for autonomous systems. In the present case of a second order equation, we have $n = 2$ for the equivalent system; hence trajectories are curves in the phase plane $\mathbb{R}^2$. The essence of Prüfer's method is to represent these curves in polar coordinates.

IV. The Prüfer Transform. The differential equation

$$Lu = (pu')' + qu = 0 \tag{5}$$

can be represented in the form of an equivalent first order system for $y_1 = u$, $y_2 = pu'$; cf. (26.2′). It is customary to interchange y_1 and y_2 and to look for a representation of the curve $(p(x)u'(x), u(x))$ in a $\xi\eta$-plane (the phase plane) in polar coordinates,

$$\xi(x) = p(x)u'(x) = \rho(x)\cos\varphi(x), \quad \eta(x) = u(x) = \rho(x)\sin\varphi(x). \tag{6}$$

If $u(x_0) = u'(x_0) = 0$, then by Theorem 19.I, $u \equiv 0$; therefore, the trajectory of a nontrivial solution never passes through the origin. The functions $\xi(x)$, $\eta(x)$ belong to $C^1(J)$, and it is not hard to see that there exist functions $\rho(x) > 0$, $\varphi(x)$ in $C^1(J)$ such that (6) holds. One begins by defining

$$\rho(x) = \sqrt{\xi^2(x) + \eta^2(x)}, \quad \varphi(x) = \arctan\frac{\eta(x)}{\xi(x)} = \operatorname{arccot}\frac{\xi(x)}{\eta(x)}.$$

To construct φ, we first fix $\varphi(a)$, say, by requiring $-\pi < \varphi(a) \le \pi$. When the solution curve is close to the ξ-axis, the arctan formula is used, and near the η-axis, the arccot formula. When changing from one formula to the other, one must choose a value of the (multivalued) arc function such that φ is continuous. The function φ is uniquely determined up to an additive constant $2k\pi$ (k an integer). Additional details are given in A.III.

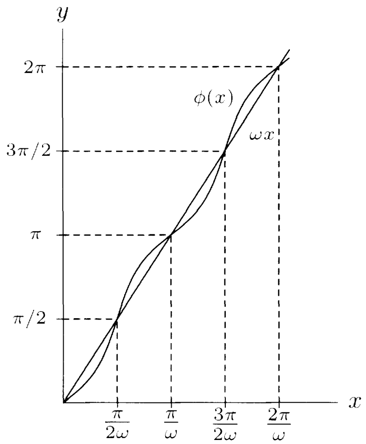
In complex notation, $\zeta(x) = \xi(x) + i\eta(x)$, the representation (6) reads simply $\zeta(x) = \rho(x)e^{i\varphi(x)}$. The function φ is denoted by $\varphi(x) = \arg\zeta(x)$ and is called the *argument function* belonging to the solution u.

From the equations

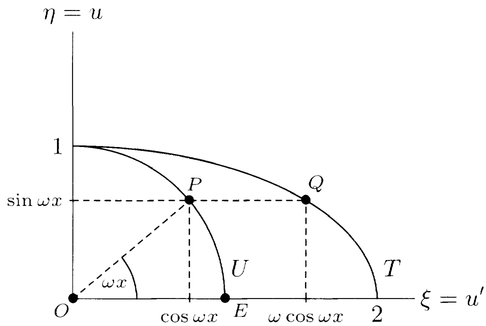
$$\xi' = \rho'\cos\varphi - \rho\varphi'\sin\varphi, \quad \eta' = \rho'\sin\varphi + \rho\varphi'\cos\varphi$$

one obtains

$$\eta'\cos\varphi - \xi'\sin\varphi = \rho\varphi',$$



The argument function of
$u(x) = \sin \omega x$



Its geometric construction
$(\omega = 2)$

and further, since $\xi' = (pu')' = -qu = -q\rho \sin \varphi$ and $\eta' = \xi/p = \rho \cos \varphi/p$,

$$\varphi' = \frac{1}{p} \cos^2 \varphi + q \sin^2 \varphi = \frac{1}{p} + \left(q - \frac{1}{p}\right) \sin^2 \varphi. \tag{7}$$

A similar argument yields

$$\rho' = \left(\frac{1}{p} - q\right) \rho \cos \varphi \sin \varphi. \tag{8}$$

Thus we have arrived—and here lies the significance of the Prüfer transform—at a *first order* differential equation for φ. Once φ is known, then ρ can be calculated explicitly by a quadrature.

Let φ be the argument function of a solution $u \not\equiv 0$. We make the following observations concerning ϕ (k is always an integer):

(a) $u(x_0) = 0 \iff \varphi(x_0) = k\pi$ and $u'(x_0) = 0 \iff \varphi(x_0) = k\pi + \frac{1}{2}\pi$.

(b) The function $\varphi_1(x) = \varphi(x) + \pi$ is an argument function for $-u$.

(c) $\varphi_2(x) = \varphi(x) + k\pi$ is also a solution of equation (7).

(d) *Example.* The function $u = \sin \omega x$ ($\omega > 0$) is a solution of the differential equation $u'' + \omega^2 u = 0$. Its trajectory in the $\xi\eta$-plane is an ellipse with semiaxes ω and 1, and its argument function $\varphi(x) = \arctan(\omega^{-1} \tan \omega x)$ satisfies the differential equation

$$\varphi' = \cos^2 \varphi + \omega^2 \sin^2 \varphi \quad \text{in } \mathbb{R}, \quad \varphi(0) = 0.$$

This follows from $\xi = u' = \omega \cos \omega x$, $\eta = u = \sin \omega x$, $(\xi/\omega)^2 + \eta^2 = 1$, and (7).

The relation $\varphi(x) = \omega x$ holds for $\omega x = \frac{1}{2}k\pi$. Since the functions $\varphi(x)$ and ωx are increasing,

$$|\varphi(x) - \omega x| < \frac{\pi}{2} \quad \text{in } \mathbb{R}.$$

In the special case $\omega = 1$, the trajectory is the unit circle and $\varphi(x) = x$.

The right-hand figure shows part of the unit circle U and the trajectory T in the $\xi\eta$-plane with two points $P = (\cos\omega x, \sin\omega x) \in U$ and $Q = (u'(x), u(x)) \in T$. Let $O = (0,0)$ and $E = (1,0)$. Then $\angle POE = \omega x$ and $\angle QOE = \phi(x)$.

(e) *Exercise.* Show that in the preceding example

$$\phi(x) < \omega x \text{ in } (0, \pi/2\omega), \quad \phi(x) > \omega x \text{ in } (\pi/2\omega, \pi/\omega)$$

for $\omega > 1$ and that for $0 < \omega < 1$ these inequalities are reversed. Show also that the function $\phi(x) - \omega x$ is periodic with period π/ω.

Hint. Since P and Q have the same η-coordinate, the first inequality can be read off the figure. Draw a similar figure for each of the other quadrants and for the case $\omega < 1$.

In the next theorems two operators L and L_0 with coefficients (p, q) and (p_0, q_0) are considered. It is assumed that the conditions in (SL) hold for L and L_0 (in particular, $p > 0$ and $p_0 > 0$ in J).

V. Lemma. *Let $p_0 \geq p$, $q_0 \leq q$ in $J = [a,b]$. Let u, v be nontrivial solutions of $Lu = 0$, $L_0 v = 0$ with the argument functions $\varphi = \arg u$, $\varphi_0 = \arg v$. Then $\varphi_0(a) \leq \varphi(a)$ implies $\varphi_0 \leq \varphi$ in J. More precisely:*

(a) *$\varphi_0(a) < \varphi(a)$ implies $\varphi_0 < \varphi$ in J.*

(b) *$\varphi_0(a) = \varphi(a)$ and $q_0 \not\equiv q$ implies $\varphi_0(b) < \varphi(b)$.*

Proof. Equation (7) is of the form $\varphi' = f(x, \varphi)$, where $f(x, y)$ and $\partial f/\partial y$ are continuous. Therefore, f is locally Lipschitz continuous in y, and Theorem 9.IX applies. The inequality $\varphi_0 \leq \varphi$ and (a) follow directly from 9.IX. Conclusion (b) also follows easily, since (i) $\sin^2\varphi(x) > 0$ if $u(x) \neq 0$ and (ii) u does not vanish in any interval. ∎

VI. The Location of the Zeros. Let J be an arbitrary interval. Assume that the coefficients p, q of L satisfy $0 < p \in C^1(J)$, $q \in C^0(J)$ and that the same is true for the coefficients p_0, q_0 of L_0.

(a) *A nontrivial solution u of $Lu = 0$ has only a finite or countable number of zeros, and they are all simple. In the second case, the zeros have no accumulation point in J. If v is another solution and $v(x_0) = u(x_0) = 0$, then $v = \text{const.} \cdot u$.*

Proof. Since $u(x_0) = u'(x_0) = 0$ implies $u \equiv 0$ by uniqueness (19.I), it follows that all zeros of u are simple. Let $u(x_k) = 0$ $(k = 1, 2, \ldots)$ and $\lim x_k = \xi \in J$. It is easily seen that this implies $u(\xi) = u'(\xi) = 0$, hence $u \equiv 0$ again, which is a contradiction. The proof of the last assertion is straightforward. ∎

We derive now two central theorems on the distribution of zeros.

Sturm Separation Theorem. *The zeros of two linearly independent solutions of $Lu = 0$ separate each other.*

This means that between two consecutive zeros of u there is a zero of v and vice versa.

Theorem of Sturm–Picone. *Let $p_0 \geq p$, $q_0 \leq q$ in J. Let v be a nontrivial solution of $L_0 v = 0$ with $v(\alpha) = v(\beta) = 0$ $(\alpha < \beta)$. If $q_0 \not\equiv q$ in (α, β), then every solution u of $Lu = 0$ has a zero in (α, β).*

Roughly speaking: If q is enlarged or p diminished, then the zeros come closer together. Consider as an example the equation $u'' + \lambda u = 0$ $(\lambda > 0)$ with the solutions $u = \sin \sqrt{\lambda}\,(x + c)$.

Proof. We may assume that α, β are consecutive zeros of v and $v > 0$ in (α, β) (v can be replaced by $-v$). Then $v'(\alpha) > 0$, $v'(\beta) < 0$, and the corresponding argument function φ_0 satisfies $\varphi_0(\alpha) = 0$, $\varphi_0(\beta) = \pi$ (consider the location of $(p_0 v', v)$ in the phase plane). In Sturm's theorem we have $L = L_0$ and $u(\alpha) \neq 0$, say, $u(\alpha) > 0$. Then the argument function φ of u satisfies $0 < \varphi(\alpha) < \pi$, and Lemma V.(a) gives $\varphi(\beta) > \pi$. Hence there is $x_0 \in (\alpha, \beta)$ with $\varphi(x_0) = \pi$, i.e. $u(x_0) = 0$; cf. IV.(a). This proves Sturm's theorem.

This proof works also for the Sturm–Picone theorem if $u(\alpha) \neq 0$. If $u(\alpha) = 0$ and, say, $u'(\alpha) > 0$, then $\varphi(\alpha) = 0$, and Lemma V.(b) implies again $\varphi(\beta) > \pi$. Hence u has a zero in (α, β). ■

Historical remark. In the special case $p = p_0$ the Sturm–Picone theorem was proved by Jacques Charles François Sturm in 1836. The general form goes back to Mauro Picone (1909). His proof is based on the Picone identity, which generalizes the Lagrange identity in 26.II.

VII. Preliminaries to the Eigenvalue Problem. We consider the solution $u = u(x, \lambda)$ of an initial value problem for equation (4),

$$Lu + \lambda r u = 0 \ \text{ in } \ J, \ u(a) = \sin \alpha, \ p(a)u'(a) = \cos \alpha \ (0 \leq \alpha < \pi), \tag{9}$$

under the assumption (SL). The solution is unique and, by Theorem 13.II, continuous in $(x, \lambda) \in J \times \mathbb{R}$ (it is even holomorphic in λ; cf. 13.III). The argument function $\varphi(x, \lambda)$ corresponding to $u(x, \lambda)$ is also continuous in (x, λ), and it satisfies equation (7) with $q + \lambda r$ in place of q, i.e.,

$$\varphi' = \frac{1}{p} + \left(q - \frac{1}{p} + \lambda r\right) \sin^2 \varphi \tag{10}$$

and $\varphi(a, \lambda) = \alpha$. The argument function has the following properties:

(a) $\varphi(x, \lambda)$ is strongly increasing in $\lambda \in \mathbb{R}$ for $a < x \leq b$.

(b) $\varphi(b, \lambda) \to 0$ as $\lambda \to -\infty$.

(c) There exist positive constants δ, D, λ_0 such that

$$\delta\sqrt{\lambda} \leq \varphi(b, \lambda) \leq D\sqrt{\lambda} \ \text{ for } \ \lambda \geq \lambda_0.$$

(d) If $\varphi(x_0, \lambda_0) = k\pi$ $(k \in \mathbb{Z})$, then $\varphi'(x_0, \lambda_0) > 0$. This means that in the xy-plane (not in the phase plane) the curve $y = \varphi(x, \lambda_0)$ crosses the line $y = k\pi$ at most once, and in a strictly increasing manner.

Proof. (d) follows in an obvious way from (10), since $p > 0$.

(a) Let $\lambda_0 < \lambda$. We apply Lemma V with q_0, q replaced by $q + \lambda_0 r$, $q + \lambda r$ and $p_0 = p$. Since $q + \lambda_0 r < q + \lambda r$, the inequality $\varphi(x, \lambda_0) < \varphi(x, \lambda)$ for $a < x \le b$ follows.

(b) To construct an upper bound for φ, we use Theorem 9.IV on upper solutions. Thus we are looking for a function w with

$$w' > f(x, w) \quad \text{and} \quad w(a) > \alpha,$$

where $f(x, \varphi)$ denotes the right-hand side of the differential equation (10).

Let $w(x)$ be the linear function with $w(a) = \pi - \varepsilon$, $w(b) = \varepsilon$. Here $\varepsilon > 0$ is chosen so small that $\alpha < w(a)$. We have $\sin^2 w \ge \sin^2 \varepsilon$, $r_0 := \min r(x) > 0$, and hence for $\lambda < 0$,

$$f(x, w) \le \frac{1}{p} + \left(q + \frac{1}{p} + \lambda r_0 \right) \sin^2 \varepsilon \to -\infty \quad \text{for} \quad \lambda \to -\infty.$$

Since w' is constant, w satisfies the conditions for a supersolution if $\lambda \le \lambda^* < 0$. Hence $\varphi(x, \lambda) \le w(x)$, and in particular, $\varphi(b, \lambda) \le \varepsilon$ for $\lambda \le \lambda_0$. This proves (b).

(c) For purposes of comparison, we consider a problem with constant coefficients p_0, q_0, r_0,

$$p_0 u'' + (q_0 + \lambda r_0) u = 0, \quad u(a) = 0, \quad u'(a) > 0,$$

where $p_0 = \max p(x)$, $q_0 = \min q(x)$, $r_0 = \min r(x) > 0$. For $q_0 + \lambda r_0 > 0$, the function $u_0(x, \lambda) = \sin \omega_0 (x - a)$ with $\omega_0 = \sqrt{(q_0 + \lambda r_0)/p_0}$ is a solution of this problem. According to Example IV.(d), with x replaced by $x - a$, the argument function $\varphi_0(x, \lambda)$ satisfies

$$\varphi_0(b, \lambda) = \omega_0 (b - a) + c \quad \text{with} \quad |c| < \pi/2,$$

and this implies $\varphi_0(b, \lambda) \ge \delta(b - a)\sqrt{\lambda}$ for large λ and, e.g., $\delta = \sqrt{r_0/(2p_0)}$. Since $\varphi_0(a, \lambda) = 0 \le \varphi(a, \lambda)$, Lemma V shows that $\varphi_0(b, \lambda) \le \varphi(b, \lambda)$, which establishes the first inequality in (c).

For the proof of the second inequality, we consider again a problem with constant coefficients $p_1 = \min p(x) > 0$, $q_1 = \max q(x)$, $r_1 = \max r(x)$. The solution $u_1(x, \lambda) = \sin \omega_1 (x - a)$, $\omega_1 = \sqrt{(q_1 + \lambda r_1)/p_1}$, has an argument function $\varphi_1(x, \lambda)$ that satisfies $\varphi_1(a, \lambda) = 0$ and

$$\varphi_1(b, \lambda) = \omega_1 (b - a) + c, \quad |c| < \pi/2.$$

We again apply Lemma V, but with φ and φ_1 in place of φ_0 and φ. Since the initial condition $\varphi(a, \lambda) = \alpha \le \varphi_1(a, \lambda) = 0$ is not satisfied in general, we change over to the argument function $\varphi_1(x, \lambda) + 2\pi$ and obtain $\varphi(b, \lambda) < \varphi_1(b, \lambda) + 2\pi$. This proves the second inequality in (c). ∎

VIII. The Eigenvalue Problem. The first boundary condition

$$R_1 u = \alpha_1 u(a) + \alpha_2 p(a) u'(a) = 0$$

has a geometrical interpretation in the phase plane. It says that the vectors $(p(a)u'(a), u(a))$ and (α_2, α_1) are perpendicular (their inner product vanishes). Thus there is one and only one number α such that

$$\alpha_1 \sin\alpha + \alpha_2 \cos\alpha = 0, \quad 0 \le \alpha < \pi. \tag{11}$$

It is given by $\alpha = \dfrac{\pi}{2} + \arctan\dfrac{\alpha_1}{\alpha_2}$ ($\alpha = 0$ if $\alpha_2 = 0$, otherwise the principal value of arctan is used). In geometric terms, α is the angle between the positive ξ-axis and the straight line through 0 perpendicular to the vector (α_2, α_1). If $u(x, \lambda)$ is the solution of the initial value problem (9) with this value of α, then $R_1 u = 0$, and every solution of (4) satisfying $R_1 u = 0$ is a multiple of u.

Likewise (and with a similar interpretation in the phase plane), there is a unique number β satisfying

$$\beta_1 \sin\beta + \beta_2 \cos\beta = 0, \quad 0 < \beta \le \pi \tag{11'}$$

(note that if $\beta_2 = 0$ here, then we use the value $\beta = \pi$). The solution $u(x, \lambda)$ of the initial value problem (9) satisfies $R_2 u = 0$ if and only if $(p(b)u'(b), u(b)) = \text{const} \cdot (\cos\beta, \sin\beta)$ and hence if and only if $\varphi(b, \lambda) = \beta + n\pi$ $(n \in \mathbb{Z})$. By VII.(a)–(c) there is for each $n \ge 0$ exactly one $\lambda = \lambda_n$ such that

$$\varphi(b, \lambda_n) = \beta + n\pi, \quad n = 0, 1, 2, \ldots,$$

while for $n < 0$ no such λ exists (this statement would be false if we had taken $\beta = 0$ for the case $R_2 u = u(b) = 0$). The numbers λ_n are the eigenvalues we are seeking, and the functions

$$u_n(x) := u(x; \lambda_n)$$

are the corresponding eigenfunctions. Property VII.(c) shows that

$$\delta^2 \lambda_n \le (\beta + \pi n)^2 \le D^2 \lambda_n.$$

This inequality implies the following result on

Asymptotic Behavior. *There are positive constants c, C such that*

$$cn^2 \le \lambda_n \le Cn^2 \ \textit{ for large } \ n. \tag{12}$$

This inequality proves and sharpens the first part of Theorem II.

Distribution of Zeros. By IV.(a), u_n has a zero at x if and only if $\varphi(x, \lambda_n) = k\pi$. Now

$$0 \le \varphi(a; \lambda_n) = \alpha < \pi \ \text{ and } \ n\pi < \varphi(b; \lambda_n) = n\pi + \beta \le (n+1)\pi.$$

This inequality combined with VII.(d) shows that $\varphi(x; \lambda_n)$ assumes the values π, $2\pi, \ldots, n\pi$ exactly once in $J^\circ = (a, b)$ and no other values of the form $k\pi$. Hence

u_n has exactly n zeros in (a,b), which we label $a < x_1 < x_2 < \cdots < x_n < b$. The Sturm–Picone theorem tells us that between x_k and x_{k+1} $(k = 1, \ldots, n-1)$ there is a zero of u_{n+1}. Now, using VII.(a), we get

$$\alpha = \varphi(a, \lambda_n) = \varphi(a, \lambda_{n+1}) < \pi = \varphi(x_1, \lambda_n) < \varphi(x_1, \lambda_{n+1}).$$

Therefore, there is a zero of u_{n+1} between a and x_1. Similarly, one shows that u_{n+1} has a zero in (x_n, b). Since there are $n+1$ zeros, there must be exactly one zero of u_{n+1} in each of the $n+1$ intervals (a, x_1), (x_k, x_{k+1}), (x_n, b). This completes the proof of Theorem II. ∎

The expansion theorem III will be proved in the next section, §28. There is also a direct proof, which can be found in Kamke (1945) or Titchmarsh (1962).

IX. Comparison Theorem for Eigenvalues. *Consider two eigenvalue problems with data* $(p_0, q_0, r_0, \alpha_0, \beta_0)$ *and* (p, q, r, α, β) *satisfying* (SL)*; the numbers* α, α_0 *or* β, β_0 *determine the boundary conditions at* a *or* b *according to* (11) *or* (11′)*, resp. If the inequalities*

$$p_0 \geq p,\ q_0 \leq q,\ r_0 \leq r \ \text{ in } J \text{ and } \ 0 \leq \alpha_0 \leq \alpha < \pi,\ 0 < \beta \leq \beta_0 \leq \pi$$

hold with a strict inequality in at least one place (e.g., $r_0 \not\equiv r$ *or* $\alpha_0 < \alpha$*), then the corresponding eigenvalues satisfy*

$$\lambda_n^0 > \lambda_n \ \text{ for } \ n = 0, 1, 2, \ldots.$$

Proof. By Lemma V, we have $\varphi_0(b, \lambda) < \varphi(b, \lambda)$ for the corresponding argument functions. Since λ_n^0 is determined by $\varphi_0(b, \lambda_n^0) = \beta_0 + n\pi$ and λ_n by $\varphi(b, \lambda_n) = \beta + n\pi \leq \beta_0 + n\pi = \varphi_0(b, \lambda_n^0)$ and since $\varphi(b, \lambda)$ is strictly increasing in λ, the inequality $\lambda_n < \lambda_n^0$ follows. ∎

X. Oscillation. We consider (real-valued) solutions of $Lu = (pu')' + qu = 0$ in a noncompact interval J. A solution u is said to be *oscillating* (in J) if it is nontrivial and has an infinite number of zeros. The equation $Lu = 0$ is *oscillatory* (in J) if it has an oscillating solution, otherwise *nonoscillatory.* If the equation $Lu = 0$ is oscillatory, then by the Sturm separation theorem, every nontrivial solution oscillates.

Oscillation Theorem. *Consider the differential equation* $Lu = (pu')' + qu = 0$ *in* $J = [a, \infty)$*, where* $p > 0$ *and* q *are continuous in* J *and such that*

$$\int_a^\infty [1/p(x)]\,dx = \infty \ \text{ and } \ \int_a^\infty q(x)\,dx = \infty.$$

(a) *If* $q(x) \geq 0$*, then the differential equation is oscillatory.*

(b) *If for some* $\alpha > 0$ *the integral* $\displaystyle\int_a^\infty \left|\frac{1}{p} - \alpha q\right| dx < \infty$*, then the differential equation is oscillatory and all solutions are bounded.*

Proof. (a) We have to prove that $\varphi(x) = \arg u(x) \to \infty$ as $x \to \infty$. Equation (7) implies that φ is increasing. Assume that $\lim \varphi(x) = c < \infty$. Then either $\sin^2 c \geq \frac{1}{2}$ or $\cos^2 c \geq \frac{1}{2}$. In the first case there exists x_0 such that $\sin^2 \varphi(x) \geq \frac{1}{4}$ for $x > x_0$. From equation (7) one obtains $\varphi' \geq \frac{1}{4}q(x)$, which together with the divergence of the integral $\int_a^\infty q\,dx$ shows that $\lim \varphi(x) = \infty$, contrary to our assumption. A similar reasoning applies if $\cos^2 c \geq \frac{1}{2}$.

(b) If we consider, instead of L, the operator βL with the coefficients $\bar{p} = \beta p$, $\bar{q} = \beta q$ and choose $\beta = \sqrt{\alpha}$, then $\alpha q - 1/p = \beta(\bar{q} - 1/\bar{p})$. Thus we may assume that $\alpha = 1$ and that the integral of $|q - 1/p|$ is convergent.

The differential equations (7), (8) for φ and ρ imply that

$$\varphi' = \frac{1}{p} + h_1(x), \qquad \rho' = h_2(x)\rho,$$

where h_1 and h_2 are integrable over the interval $[a, \infty)$. It easily follows that $\varphi(x) \to \infty$ as $x \to \infty$ and $\rho(x)$ remains bounded. ∎

XI. Amplitude Theorem. *Let J be an arbitrary interval, $p, q \in C^1(J)$, and $p > 0$, $q > 0$. Further, let u be a nontrivial solution of $Lu = (pu')' + qu = 0$. Then every stationary point of u (point where u' vanishes) is an extremal point; i.e., u has a local maximum or minimum at that point. For two consecutive extremal points $x_k < x_{k+1}$,*

$$|u(x_k)| \geq |u(x_{k+1})| \quad \textit{if } pq \textit{ is weakly increasing}$$

and similarly,

$$|u(x_k)| \leq |u(x_{k+1})| \quad \textit{if } pq \textit{ is weakly decreasing.}$$

In short, the amplitudes are decreasing or increasing, when pq is increasing or decreasing. If pq is strictly monotone, then the inequalities are strict.

Proof. The assertion about extremal points follows easily from the equation $Lu = 0$, taking into account $q > 0$. Consider now the function

$$y(x) := u^2 + \frac{1}{pq}(pu')^2.$$

The derivative is

$$y' = 2uu' - \frac{(pq)'}{(pq)^2}(pu')^2 + \frac{2pu'}{pq}(-qu) = -(pq)'\left(\frac{u'}{q}\right)^2.$$

Thus, y is decreasing or increasing whenever $(pq)' \geq 0$ or ≤ 0, resp. Since $u'(x) = 0$ implies $y(x) = u^2(x)$ and $u'(x_k) = u'(x_{k+1}) = 0$, the conclusion follows. ∎

In the following two sections the Sturm–Picone theorem is used to study the oscillatory behavior of solutions and the asymptotic distribution of their zeros. The coefficients p, q of L and p_0, q_0 of L_0 are assumed to have the usual properties (see VI) in an interval $J = [a, \infty)$. Solutions are understood to be nontrivial solutions; the zeros of a solution u of $Lu = 0$ are denoted by $x_1 < x_2 < x_3 < \cdots$. An immediate consequence of the Sturm–Picone theorem is a

XII. Comparison Theorem. *Let $p_0 \geq p$ and $q_0 \leq q$ in J. If the equation $L_0 u = 0$ is oscillatory in J, then the same is true for the equation $Lu = 0$. Equivalently: If $Lu = 0$ is nonoscillatory, then so is $L_0 u = 0$.*

XIII. The Distribution of Zeros. The differential equation

$$p_0 v'' + q_0 v = 0 \quad (p_0 > 0,\ q_0 > 0 \ \text{ constant}) \tag{13}$$

has the solutions $v = \alpha \sin \omega_0 (x + \delta)$, $\omega_0 = \sqrt{q_0/p_0}$. The distance d_0 between consecutive zeros is constant, $d_0 = \pi\sqrt{p_0/q_0}$.

For $c > 0$ we define

$$p_c(x) = \min p(t), \quad P_c(x) = \max p(t), \quad \text{where} \ \ x \leq t \leq x + c,$$

and $q_c(x)$ and $Q_c(x)$ are defined similarly. As before, $Lu = (pu')' + qu$.

Theorem. (a) *Assume that q is positive in J and*

$$\liminf_{x\to\infty} \frac{P_c(x)}{q_c(x)} = A < \infty \ \ \textit{for} \ \ c = \pi\sqrt{A + \varepsilon}, \tag{14}$$

where $\varepsilon > 0$. Then equation $Lu = 0$ is oscillatory in $[a, \infty)$.

(b) *Moreover, if*

$$\lim_{x\to\infty} \frac{P_c(x)}{q_c(x)} = \lim_{x\to\infty} \frac{p_c(x)}{Q_c(x)} = A < \infty \ \ \textit{for} \ \ c = \pi\sqrt{A + \varepsilon}, \tag{15}$$

then $\lim(x_{k+1} - x_k) = \pi\sqrt{A}$ *for every solution of $Lu = 0$.*

Proof. (a) There is a sequence (a_k) tending to ∞ such that $P_c(a_k)/q_c(a_k) < A + \varepsilon$ for all k. Let v be a solution of (13) with coefficients $p_0 = P_c(a_k)$, $q_0 = q_c(a_k)$ and initial value $v(a_k) = 0$. Then $v(a_k + d_k) = 0$, where $d_k = \pi\sqrt{p_0/q_0} < c$. Since $p(x) \leq p_0$ and $q(x) \geq q_0$ in $J_k = [a_k, a_k + d_k]$, a solution u has a zero in J_k by the Sturm–Picone theorem ($k = 1, 2, \ldots$).

(b) It follows from (15) that $P_c(x_k)/q_c(x_k) < A + \varepsilon$ (k large), when (x_k) is the sequence of zeros of a solution u. Now the proof of (a) (with $a_k = x_k$) shows that $x_{k+1} \in J_k$ and hence $x_{k+1} - x_k < c$. If $A = 0$, then $\lim (x_{k+1} - x_k) = 0$ follows.

Now let $0 < B < A$. It follows from (15) that $p_c(x_k)/Q_c(x_k) > B$. We shall prove that $x_{k+1} - x_k > \pi\sqrt{B}$ ($k \geq k_0$). Assume to the contrary that $x_{r+1} - x_r \leq \pi\sqrt{B} < c$ ($r \geq k_0$). Let $p_0 = p_c(x_r)$, $q_0 = Q_c(x_r)$. Then the difference d_0 of consecutive zeros of a solution v of (13) equals $\pi\sqrt{p_0/q_0} > \pi\sqrt{B}$. But this is a contradiction to the Sturm–Picone theorem, because $p_0 \leq p(x)$ and

$q_0 \geq q(x)$ in J_r. It shows that $x_{k+1} - x_k > \pi\sqrt{B}$, as asserted. Since we can assume that B is arbitrarily close to A and $\varepsilon > 0$ in (15) can be chosen arbitrarily small, part (b) is proved. ∎

The class S. The function q belongs to the class S if it is continuous and positive in an interval $[a, \infty)$ and

$$\lim_{x\to\infty} \frac{Q_c(x)}{q_c(x)} = 1 \quad \text{for every} \quad c > 0.$$

Corollary. *If p and q belong to the class S and*

$$\lim_{x\to\infty} \frac{p(x)}{q(x)} = A, \tag{15'}$$

then (15) *holds. It follows that the equation $Lu = 0$ is oscillatory and that for every solution, $x_{k+1} - x_k \to \pi\sqrt{A}$ as $k \to \infty$.*

(c) *Properties of the class S.* (i) If f and g belong to S, then αf $(\alpha > 0)$ and fg also belong to S. (ii) If f is continuous and $f(x) \to \alpha > 0$ as $x \to \infty$, then f belongs to S. (iii) If f belongs to S and $h = o(f)$, i.e., $h(x)/f(x) \to 0$ as $x \to \infty$, then $f + h$ belongs to S.

(d) All functions x^α $(\alpha \in \mathbb{R})$ and all polynomials that are positive for large x belong to S, but $e^{\alpha x}$ does not $(\alpha \neq 0)$.

(e) If for some $c > 0$, $Q_c(x)/q_c(x) \to 1$ as $x \to \infty$, then q belongs to S.

Exercise. Prove the corollary and (c)–(e).

(f) Consider the equation

$$Lu = (x^\alpha u')' + g(x)x^\alpha u = 0 \quad \text{in} \quad [1, \infty),$$

where α is any real number and $g(x) \to \beta > 0$ as $x \to \infty$. Then p and q belong to S. Since $\lim p/q = 1/\beta$, the equation is oscillatory and $x_{k+1} - x_k \to \pi/\sqrt{\beta}$ as $k \to \infty$.

(g) *The Bessel equation*

$$x^2u'' + xu' + (x^2 - \alpha^2)u = 0$$

is oscillatory in $[1, \infty)$. Every solution u satisfies an inequality $|u(x)| \leq C/\sqrt{x}$, its amplitudes are strictly decreasing, and $\lim\,(x_{k+1} - x_k) = \pi$. These properties hold for all real α.

Exercise. Give a proof of (g). *Hints:* Write the equation in self-adjoint form and apply (f) and the amplitude theorem. Show that the function $z(x) = \sqrt{x}\,u(x)$ satisfies a differential equation $z'' + qz = 0$ and use Theorem IX.(b) to prove that z is bounded.

In order to apply the oscillation theorem to general second order equations, the following transformations are helpful.

XIV. Transformation Formulas. (a) The differential equation

$$u'' + a_1(x)u' + a_0(x)u = h(x)$$

is transformed by

$$v(x) := u(x)e^{A(x)}, \quad A(x) = \frac{1}{2}\int a_1(x)\,dx$$

into the following differential equation for v,

$$v'' + \left(a_0(x) - \frac{1}{2}a_1'(x) - \frac{1}{4}a_1^2(x)\right)v = h(x)e^{A(x)}.$$

(b) The differential equation

$$(p(x)u')' + q(x)u = h(x)$$

transforms, after a new independent variable

$$t = t(x) := \int \frac{dx}{p(x)}$$

with the inverse function $x(t)$ is introduced, into a differential equation for $v(t) := u(x(t))$,

$$\frac{d^2v}{dt^2} + p(x)q(x)v = p(x)h(x) \text{ with } x = x(t).$$

The inverse of $t(x)$ exists because $p > 0$. Proof as an exercise.

(c) Use (a) to transform the differential equation

$$xu'' - u' + x^3u = 0$$

and then write the equation in self-adjoint form (see 26.II) and transform it using (b).

XV. Exercise. For which values of $\alpha, \beta, \gamma \in \mathbb{R}$ is the differential equation

$$(e^{\alpha x}u')' + \gamma e^{\beta x}u = 0$$

oscillatory in $[0, \infty)$? Incidentally, the answer shows that Theorem XIII.(b) does not hold with the weaker assumption (15′). *Hint:* In the case $\alpha = \beta$ the solutions can be given explicitly.

XVI. Exercises. (a) Consider the eigenvalue problem

$$u'' + \lambda u = 0 \text{ for } 0 \le x \le 1, \quad u(0) = u'(0), \quad u(1) = 0.$$

Determine the eigenvalues and eigenfunctions and show that

$$\sqrt{\lambda_n} = \frac{\pi}{2} + n\pi + \beta_n \quad (n = 0, 1, 2, \ldots), \quad \text{where } \beta_n \downarrow 0 \quad (n \to \infty).$$

Draw a sketch of u_0 and u_1.

(b) Determine the eigenvalues and eigenfunctions if the boundary conditions in (a) are changed to

$$u(0) = u'(0), \quad u(1) = u'(1).$$

(c) Solve the eigenvalue problem

$$(xu')' + \frac{\lambda}{x}\, u = 0 \text{ in } [1, \pi], \quad u'(1) = 0, \quad u'(\mathrm{e}^{2\pi}) = 0.$$

Is $\lambda = 0$ an eigenvalue?

(d) Show that assertion (b) of Lemma V is true if $q \equiv q_0$ and $p < p_0$ unless $u = v = \text{const.}$

XVII. Exercise. Determine all solutions of the differential equation

$$u'' + \frac{\alpha}{x^2}\, u = 0 \qquad (\alpha \in \mathbb{R})$$

and all values α for which the differential equation is oscillatory (substitute $x = \mathrm{e}^t$). Using the Sturm–Picone theorem, prove the following

Oscillation Theorem. *The differential equation*

$$u'' + q(x)u = 0$$

($q(x)$ continuous for $x \ge a$) has the following properties in $[a, \infty)$:

(a) *It is oscillatory if* $\liminf\limits_{x\to\infty} x^2 q(x) > \frac{1}{4}$.

(b) *It is nonoscillatory if* $\limsup\limits_{x\to\infty} x^2 q(x) < \frac{1}{4}$.

Supplement: Rotation-Symmetric Elliptic Problems

We first investigate radial solutions of

$$\Delta u + \lambda u = 0 \text{ in } B, \quad u = 0 \text{ on } \partial B, \tag{16}$$

where B is the unit ball in $\mathbb{R}^n$, and then turn to the nonlinear boundary value problem $\Delta u = f(u)$ in B, $u = 0$ on ∂B. As in the supplement of §6, we use the operator L_α of 6.XII for real $\alpha \ge 0$ ($\alpha = n - 1$ gives the radial Δ-operator). Observe that

$$L_\alpha y = f(x, y) \iff (x^\alpha y')' = x^\alpha f(x, y).$$

XVIII. The Eigenvalue Problem. Let y be the solution of the initial value problem

$$(x^\alpha y')' + x^\alpha y = 0, \quad y(0) = 1, \quad y'(0) = 0 \quad (\alpha \geq 0). \tag{17}$$

It exists in $[0, \infty)$, is uniquely determined, and oscillates; cf. Theorem 6.XIII and Example XIII.(f). The zeros of y are denoted by $0 < \xi_0 < \xi_1 < \xi_2 < \cdots$. According to XIII.(f), $\xi_{k+1} - \xi_k \to \pi$, which implies $\xi_k/(k\pi) \to 1$ $(k \to \infty)$.

The eigenvalue problem

$$(x^\alpha u')' + \lambda x^\alpha u = 0, \quad u'(0) = 0, \quad u(1) = 0 \tag{18}$$

is easily solved, since $u(x) = y(\beta x)$ satisfies $(x^\alpha u')' + \beta^2 x^\alpha u = 0$, $u'(0) = 0$ and $u(1) = y(\beta)$. One obtains the

(a) eigenvalues $\lambda_n = \xi_n^2$ with eigenfunctions $u_n(x) = y(\xi_n x)$,

(b) asymptotic behavior of eigenvalues, $\lambda_n/\pi^2 n^2 \to 1$ as $n \to \infty$, and

(c) distribution of zeros of the eigenfunctions as described in Theorem II.

Exercise. Show that there are no other eigenvalues. *Hint.* Assume that λ and $u \not\equiv 0$ satisfy (18) and show: (i) $u(0) \neq 0$, hence one may assume that $u(0) = 1$; (ii) if $\lambda \leq 0$, then $u > 0$ implies $u' > 0$; if $\lambda = \beta^2 > 0$, then $u = y(\beta x)$.

XIX. The Boundary Value Problem. For the linear equation $L_\alpha y = f(x)$ the solution of the boundary value problem

$$(x^\alpha u')' = x^\alpha f(x) \text{ in } [0, 1], \quad u'(0) = 0, \quad u(1) = 0 \tag{19}$$

is given by (6.10–12):

$$y(x) = (I_\alpha f)(x) - (I_\alpha f)(1) = \int_0^1 \Gamma(x, \xi)\xi^\alpha f(\xi)\, d\xi.$$

Since

$$(I_\alpha f)(x) = \int_0^x [h(x) - h(\xi)]\xi^\alpha f(\xi)\, d\xi, \quad \text{where} \quad h(x) = \int_1^x s^{-\alpha}\, ds,$$

Green's function is given by $\Gamma(x, \xi) = h(x)$ for $\xi < x$ and $h(\xi)$ for $\xi > x$. The function $\xi^\alpha \Gamma(x, \xi)$ is continuous in the square $[0, 1]^2$. The three existence theorems in 26.IX, XX, and XXI now carry over to the present case:

Existence Theorem. (a) *If the function $f(x, z)$ is continuous in the strip $S = [0, 1] \times \mathbb{R}$ and satisfies a Lipschitz condition in z with a Lipschitz constant $L < \lambda_0 = \xi_0^2$, then the boundary value problem*

$$(x^\alpha u')' = x^\alpha f(x, u) \text{ in } [0, 1], \quad u'(0) = 0, \quad u(1) = \eta \tag{20}$$

has exactly one solution.

(b) *If f is continuous and bounded in S, then* (20) *has a solution.*

(c) *Let v be a subfunction and w a superfunction and $v \leq w$ in $[0, 1]$. If $f(x, z)$ is continuous in $K = \{(x, z) : 0 \leq x \leq 1,\ v(x) \leq z \leq w(x)\}$, then problem* (20) *has a solution between v and w.*

Sub- and superfunctions are defined as in 26.XXI,

$$\begin{aligned}(x^\alpha w')' &\le f(x,w),\ w'(0)=0,\ w(1)\ge\eta,\\ (x^\alpha v')' &\ge f(x,v),\ v'(0)=0,\ v(1)\le\eta.\end{aligned}$$

Proof. (a) As before, we may assume that $\eta = 0$. In this case, (20) is equivalent to the fixed point equation

$$u = Tu, \quad \text{where} \quad (Tu)(x) = \int_0^1 \Gamma(x,\xi)\xi^\alpha f(\xi, u(\xi))\, d\xi. \tag{21}$$

We consider the operator T in the space X consisting of all functions in $C^0([0,1])$ with a *finite* norm

$$\|v\| = \sup\left\{\frac{|v(x)|}{u_0(x)} : 0 \le x \le 1\right\};$$

recall that the eigenfunction u_0 is positive in $[0,1)$. For $v, w \in X$ the Lipschitz condition on f implies

$$|Tv - Tw|(x) \le L\|v-w\| \int_0^1 \Gamma(x,\xi)\xi^\alpha u_0(\xi)\, d\xi.$$

The eigenfunction u_0 is the solution of (20) with $\eta = 0$ and $f = -\lambda_0 u_0$; hence

$$u_0(x) = -\lambda_0 \int_0^1 \Gamma(x,\xi)\xi^\alpha u_0(\xi)\, d\xi.$$

Since $\Gamma \le 0$, it follows that

$$|Tw - Tv|(x) \le q\|w-v\|u_0(x), \quad q = L/\lambda_0 < 1.$$

Hence $\|Tv - Tw\| \le q\|v-w\|$, and the contraction principle applies. ∎

Remarks and Exercises. 1. Prove (b) and (c); the proofs from § 26 carry over.

2. Replace the second boundary condition in (19) by $R_2u = \beta_1 u(1) + \beta_2 u'(1) = 0$ and in (20) by $R_2 u = \eta$ and treat the linear and the nonlinear boundary value problems in a similar way.

3. It may seem surprising that the eigenvalues in (18) have the same asymptotic behavior for all $\alpha \ge 0$. But one should be aware that for large values of x the term $\alpha u'/x$ becomes small and the differential equation is essentially $u'' + \lambda u = 0$, which corresponds to the case $\alpha = 0$.

4. The book *Comparison and Oscillation Theory of Linear Differential Equations* by C.A. Swanson (Acad. Press 1968) contains many results on oscillation of solutions and asymptotic distribution of eigenvalues, together with historical remarks, mostly for the case $p(x) \equiv 1$.

XX. Exercise. The results of XVIII und XIX extend to more general equations of the form (1), where $p(0) = 0$ and $r(0) = 0$ are permitted. Assume that $p(x)$ and $r(x)$ are continuous in $J = [0, b]$ and positive in $J_0 = (0, b]$. Let $R(x) = \int_0^x r(t)\,dt$ and assume that $\int_0^b [R(t)/p(t)]\,dt < \infty$. We consider the operator $Lu = (pu')'$. Prove the following:

(a) The initial value problem $Ly = r(x)f(x)$, $y(0) = \eta$, $(py')(0) = 0$ has for $f \in C^0(J)$ the unique solution

$$y = \eta + Kf, \text{ where } (Kf)(x) = \int_0^x \frac{1}{p(s)} \int_0^s r(t)f(t)\,dt\,ds.$$

(b) The corresponding nonlinear initial value problem, where $f(x)$ is replaced by $f(x, y)$, has one and only one solution whenever $f(x, y)$ is continuous in $S = J \times \mathbb{R}$ and satisfies a Lipschitz condition with respect to y.

(c) The existence theorem II remains true for the eigenvalue problem

$$Lu + \lambda r(x)u = 0 \text{ in } J_0, \quad (pu')(0) = 0, \quad u(b) = 0; \tag{22}$$

furthermore, the estimation (12) holds and $\lambda_0 > 0$.

(d) The semihomogeneous boundary value problem

$$Lu = r(x)f(x),\ (pu')'(0) = 0,\ u(b) = 0 \text{ with } f \in C(J)$$

has the solution $u(x) = (Kf)(x) - (Kf)(b) = \int_0^b \Gamma(x, \xi)r(\xi)f(\xi)\,d\xi$ and no other solution. Green's function is defined as in XIX, but with $h(x) = -\int_x^b p^{-1}(t)\,dt$.

(e) For the corresponding nonlinear boundary value problem with $f(x, u)$ in place of $f(x)$, the three propositions of the existence theorem XIX remain true.

Hints. Reduce (b) to a fixed point equation $y = Ty$, using (a). In the space $C^0(J)$ with the maximum norm, T is a contraction if the interval J is small. The solution can be extended to a larger interval by solving a "normal" initial value problem.

(c) Consider as in VII the corresponding initial value problem with $u(0) = 1$, $(pu')(0) = 0$. The argument function $\varphi(x, \lambda)$ of the solution $u(x, \lambda)$ satisfies $\varphi(0, \lambda) = \pi/2$. The propositions VII.(a)–(d) follow as before, since $p(x) \geq p_0 > 0$ and $r(x) \geq r_0 > 0$ in $[\varepsilon, b]$.

(e) The problem with $\eta = 0$ is reduced to a fixed point equation that is similar to (21). It can be treated as before.

XXI. Eigenvalue Problems in the Sense of Carathéodory. The eigenvalue problem (2), (3) for C-solutions can be studied under the assumption

$$(\mathrm{SL_C}) \qquad \boxed{1/p, q, r \in L(J),\ p > 0,\ r > 0 \text{ a.e. in } J,}$$

which is significantly weaker than (SL); cf. 26.XXIV for the boundary value problem. If $u(x, \lambda)$ is a solution of (9), then we have $u, pu' \in AC(J)$, which implies that the argument function $\varphi(x, \lambda)$ belongs also to $AC(J)$, since the arctan function satisfies a Lipschitz condition.

The argument function satisfies (10) and has the properties of VII. For the proof, Theorem 10.XXI is again crucial. Lemma V and the separation theorems VI and VII.(a) follow as before, but also VII.(d), because the function $\psi(x) = k\pi$ is a subfunction for $x > x_0$ and a superfunction for $x < x_0$, which means that $\varphi(x, \lambda_0)$ is $< k\pi$ to the left and $> k\pi$ to the right of x_0. In VII.(b) and (c) the proofs require some modification. We sketch a proof of a weaker form of VII.(c), namely $\varphi(b, \lambda) \to \infty$ as $\lambda \to \infty$.

Let $I = [\alpha, \beta] \subset J$ be an (arbitrarily small) interval and let $\phi_0(x, \lambda)$ be the solution of (10) with $\phi_0(\alpha, \lambda) = 0$ and $\phi_1(x, \lambda)$ the solution with $\phi_1(\beta, \lambda) = \pi$. We show that $\phi_0(\beta, \lambda) \geq \pi$ for large values of λ. Let $\alpha < \gamma < \delta < \beta$. According to VII.(d), there exists $\varepsilon > 0$ such that

$$\phi_0(x, 0) \geq \varepsilon \text{ in } [\gamma, \beta] \text{ and } \phi_1(x, 0) \leq \pi - \varepsilon \text{ in } [\alpha, \delta],$$

and these inequalities hold also for $\phi_0(x, \lambda)$ and $\phi_1(x, \lambda)$, $\lambda > 0$. As long as $\varepsilon \leq \phi_0 \leq \pi - \varepsilon$, we have $\phi_0' \geq (q + \lambda r) \sin^2 \varepsilon$. Since the integral of r in the interval $[\gamma, \delta]$ is positive, there is $\lambda_0 > 0$ such that $\phi_0(x, \lambda_0)$ assumes the value $\pi - \varepsilon$ in $[\gamma, \delta]$. When x moves further to the right, $\phi_0(x, \lambda_0)$ remains above $\phi_1(x, \lambda_0)$, and $\phi_0(\beta, \lambda_0) \geq \pi$ follows.

The rest is simple. We take a partition $x_0 = a < x_1 < \cdots < x_p = b$ of J with subintervals $I_k = [x_{k-1}, x_k]$ and in each I_k a solution $\phi_k(x, \lambda_k)$ of (10) that vanishes at x_{k-1} and is $\geq \pi$ at x_k. Then Lemma V shows that the argument function $\phi(x, \lambda^*)$ with $\lambda^* = \max \lambda_k$ satisfies $\phi(x_1, \lambda^*) \geq \phi_1(x_1, \lambda_1) \geq \pi$ and, since $\phi_2 + \pi$ is also a solution of (10), $\phi(x_2, \lambda^*) \geq \pi + \phi_2(x_2, \lambda_2) \geq 2\pi, \ldots, \phi(b, \lambda^*) \geq p\pi$.

Now all prerequisites required for the proof of the existence theorem II in VIII have been assembled.

XXII. Exercise. Riccati Equations. (a) If u is a nonvanishing solution of equation (5) $(p(x)u')' + q(x)u = 0$, then the function $r(x) = p(x)u'/u$ satisfies the equation

$$r' + \frac{r^2}{p(x)} + q(x) = 0\ ; \tag{23}$$

it is called the Riccati equation of (5). Conversely, if $r(x)$ is a solution of (23), then $u(x) = \exp\left(\int (r/p)\, dx\right)$ is a nonvanishing solution of (5).

(b) The same connection exists between the linear equation (5′) $u'' + g(x)u' + q(x)u = 0$ and its Riccati equation

$$r' + r^2 + g(x)r + q(x) = 0\,, \tag{23′}$$

where $r(x) = u'/u$ and $u(x) = \exp\left(\int r\, dx\right)$.

These relations can be used to derive properties of solutions of equation (5) from those of equation (23) and vice versa. An example is given in

(c) *A blow-up problem.* Let y_α be the solution of the problem

$$y' = x^2 + y^2, \quad y(0) = \alpha\,. \tag{24}$$

It exists to the right in a maximal interval $[0, b_\alpha)$ and blows up at b_α, i.e., $y_\alpha(b_\alpha) = \infty$, and b_α is continuous and strictly decreasing in $\alpha \in \mathbb{R}$ with $b_\alpha \to 0$ as $\alpha \to \infty$. The case $\alpha = 1$ has been studied in 9.V.

Hints for (c). We have $p(x) \equiv 1$ and $q(x) = x^2$ in (5). Let v or w be the solution of (5) with $(u(0), u'(0)) = (1, 0)$ or $(0, 1)$, resp. Then $u_\alpha = v - \alpha w$ is the solution of (5) with initial values $u(0) = 1, u'(0) = -\alpha$, and $r_\alpha = u'_\alpha / u_\alpha$ satisfies (23), and $r_\alpha(0) = -\alpha$. Hence $y_\alpha = -r_\alpha$ is the solution of (24). The first positive zero b_α of u_α has the above properties, and furthermore, b_α tends to the first positive zero c of w as $\alpha \to -\infty$. Note that $u_\alpha(c) = v(c) < 0$ by 27.VI and hence $b_\alpha < c$.

Remark. Strict monotonicity of b_α implies that there is exactly one blow-up solution in $[0, b)$ for $0 < b < c$ and no such solution for $b \geq c$.

§ 28. Compact Self-Adjoint Operators in Hilbert Space

In this section, we first develop an eigenvalue theory for compact self-adjoint operators in a Hilbert space. The results are then applied to the Sturm–Liouville eigenvalue problem.

I. Inner Product. An inner product (scalar product) in a real or complex linear space H is a mapping $(\cdot, \cdot)$ of $H \times H$ to $\mathbb{R}$ or $\mathbb{C}$, respectively, with the following properties ($f, g, h \in H$; $\lambda, \mu \in \mathbb{R}$ or $\mathbb{C}$)

$$(\lambda f + \mu g, h) = \lambda(f, h) + \mu(g, h) \qquad \textit{linearity,}$$

$$(f, g) = \overline{(g, f)} \qquad \textit{symmetry,}$$

$$(f, f) > 0 \quad \text{for} \quad f \neq 0 \qquad \textit{definiteness.}$$

In the complex case it follows from the second property, which is called the *Hermite property*, that (f, f) is always real and that the inner product is "antilinear" in the second argument:

$$(f, \lambda g + \mu h) = \bar{\lambda}(f, g) + \bar{\mu}(f, h).$$

In the real case, the bars are all superfluous, and the inner product is bilinear.

The following statements hold, unless otherwise noted, in both the real and the complex case. The proofs will be given for the complex case; the arguments are also valid in the real case.

In the linear space H with inner product, the relation

$$\|f\| := \sqrt{(f, f)}$$

defines a norm. The norm properties follow from 5.II. The definiteness is immediately obvious, and the homogeneity follows from $(\lambda f, \lambda f) = \lambda\bar{\lambda}(f, f) =$

$|\lambda^2|\|f\|^2$. To prove the triangle inequality, consider, for arbitrary $f, g \in H$, the expression

$$0 \le (f+\lambda g, f+\lambda g) = (f,f) + \lambda(g,f) + \bar{\lambda}(f,g) + \lambda\bar{\lambda}(g,g).$$

Setting $\lambda = (f,g)/\|g\|^2$, one obtains, after a simple intermediate calculation,

$$|(f,g)| \le \|f\| \cdot \|g\| \qquad \textit{Schwarz inequality}$$

(it follows from $(g,0) = 0$ that the inequality is valid for $g = 0$). Therefore,

$$\begin{aligned}(f+g, f+g) &= (f,f) + (f,g) + (g,f) + (g,g)\\ &\le (f,f) + 2\|f\| \cdot \|g\| + (g,g) = (\|f\| + \|g\|)^2,\end{aligned}$$

and the triangle inequality

$$\|f+g\| \le \|f\| + \|g\|$$

follows. We note two additional simple propositions that can be immediately verified by a calculation:

$$\begin{aligned}\|f+g\|^2 + \|f-g\|^2 &= 2\|f\|^2 + 2\|g\|^2 && \textit{parallelogram identity,}\\ \|f+g\|^2 &= \|f\|^2 + \|g\|^2, \quad \text{if} \quad (f,g) = 0 && \textit{Pythagorean theorem.}\end{aligned}$$

II. Inner Product Space and Hilbert Space. A linear space with an inner product is called an *inner product space* or a *pre-Hilbert space*. It is a normed space in which the norm is induced by the inner product; cf. I. If an inner product space is complete as a normed space, i.e., a Banach space, then it is called a Hilbert space. These definitions hold in both the real and complex cases. Some examples:

(a) The space $\mathbb{R}^n$ with the inner product

$$(\mathbf{a}, \mathbf{b}) = a_1 b_1 + \cdots + a_n b_n$$

is a real Hilbert space. The space $\mathbb{C}^n$ with

$$(\mathbf{a}, \mathbf{b}) = a_1 \bar{b}_1 + \cdots + a_n \bar{b}_n$$

is a complex Hilbert space. In each case, the norm is the Euclidean norm.

(b) Let H be the set $C(J)$ of the continuous real-valued functions $f(x)$ on $J : a \le x \le b$ with the inner product

$$(f,g) = \int_a^b f(x) g(x)\, dx. \tag{1}$$

In this space, the "distance" between two functions f, g is

$$\|f-g\| = \sqrt{\int_a^b (f-g)^2\, dx}. \tag{2}$$

It is easy to see that the above inner product has the required properties. This space is not complete and hence is not a Hilbert space. This is because there exist sequences of functions $f_n \in C(J)$ that are Cauchy sequences in the sense of the norm (2) but do not have a continuous function as a limit, e.g., the sequence

$$f_n(x) = \{\max(x, 1/n)\}^{-1/3}$$

on the interval $0 \le x \le 1$. The limit in the sense of the norm is the function $x^{-1/3}$, which, however, does not belong to H.

In order to extend this space to make it complete, one must include functions that are not continuous. This leads to

(c) the real Hilbert space $L^2(J)$ of measurable functions on J that are square-integrable, which means that the integral

$$\int_a^b f^2(x)\,dx < \infty.$$

It follows that the integral of f also exists (as long as the interval J is bounded). The inner product is defined as in (1). We note that measurability and integral are understood in the sense of Lebesgue.

(d) Correspondingly, the complex-valued continuous or square-integrable functions in J form a complex inner product space or Hilbert space, respectively, if one defines as inner product

$$(f, g) = \int_a^b f(x)\overline{g(x)}\,dx.$$

We note that our development of the Sturm–Liouville eigenvalue problem is largely carried out without the notion of the Lebesgue integral, that is, in the inner product space (b) of continuous functions.

(e) *Exercise.* Show that the inner product is a continuous function on $H \times H$, i.e., that $f_n \to f$, $g_n \to g$ implies $(f_n, g_n) \to (f, g)$.

III. Orthonormal Systems and Fourier Series. Let H be a real or complex inner product space. A sequence $(u_n)_0^\infty$ of elements of H is called a *(countable) orthonormal system* or an *orthonormal sequence* if

$$(u_m, u_n) = \delta_{mn} = \begin{cases} 1 & \text{for} \quad m = n, \\ 0 & \text{for} \quad m \ne n. \end{cases}$$

If f is an element of H, then

$$\sum_{i=0}^{\infty} c_i u_i \quad \text{with} \quad c_i := (f, u_i) \tag{3}$$

is called the *Fourier series* generated by f, and the c_i are called the *Fourier coefficients* of f. The following results deal with questions of the convergence of this series and its sum.

If $\sum$ denotes a finite sum and d_i are arbitrary constants, then

$$(f - \sum d_i u_i, f - \sum d_i u_i) = (f, f) - \sum d_i \bar{c}_i - \sum c_i \bar{d}_i + \sum_{i,j} d_i \bar{d}_j (u_i, u_j)$$

or

$$\|f - \sum d_i u_i\|^2 = \|f\|^2 + \sum |d_i - c_i|^2 - \sum |c_i|^2.$$

Thus $\sum d_i u_i$ is a best approximation to f if and only if $d_i = c_i$. In particular, the nth partial sum s_n of the Fourier series (3) satisfies

$$\|f - s_n\|^2 = \|f\|^2 - \sum_{i=0}^{n} |c_i|^2. \tag{4}$$

(a) *Bessel's inequality* holds,

$$\sum_{i=0}^{\infty} |c_i|^2 = \sum_{i=0}^{\infty} |(f, u_i)|^2 \leq \|f\|^2 \quad \text{for} \quad f \in H. \tag{5}$$

(b) The partial sums of the Fourier series (3) form a Cauchy sequence. Thus if H is a Hilbert space, then the Fourier series (3) converges, i.e, its partial sums converge in the sense of the norm to an element of H.

(c) Equality (in the sense of norm convergence)

$$f = \sum_{i=0}^{\infty} c_i u_i$$

holds if and only if equality holds in Bessel's inequality (5). If this is true for every $f \in H$, then (u_n) is called a *complete orthonormal system* or an *orthonormal basis.*

Propositions (a) and (c) follow immediately from (4), since the left side of the inequality is ≥ 0. To prove (b), let s_n be the nth partial sum of the Fourier series (3). If $m < n$, we have

$$\|s_n - s_m\|^2 = \sum_{i,j=m+1}^{n} c_i \bar{c}_j (u_i, u_j) = \sum_{i=m+1}^{n} |c_i|^2.$$

Thus, because of the convergence of the series (5), (s_n) is a Cauchy sequence.

(d) *Example.* The functions

$$u_0 = \frac{1}{\sqrt{2\pi}}, \quad u_{2n-1} = \frac{1}{\sqrt{\pi}} \sin nx, \quad u_{2n} = \frac{1}{\sqrt{\pi}} \cos nx \quad (n \in \mathbb{N})$$

form an orthonormal system in the space of Example II.(b) or II.(c) with $J = [0, 2\pi]$.

The functions $(e^{inx}/\sqrt{2\pi})$, $n \in \mathbb{Z}$, form an orthonormal system in the space from example II.(d) with $J = [0, 2\pi]$.

The proof of the following facts is recommended as an exercise.

(e) The partial sums of a series $\sum_{i=0}^{\infty} \alpha_i u_i$ form a Cauchy sequence if and only if $\sum_{i=0}^{\infty} |\alpha_i|^2$ converges. Thus, in a Hilbert space, this condition is necessary and sufficient for the convergence of the series.

(f) If the series $\sum_{i=0}^{\infty} a_i u_i$ converges, say to $f \in H$, then $\alpha_i = (f, u_i)$; i.e., a convergent series is the Fourier series of the function represented by the series. In particular,

$$\sum_{i=0}^{\infty} \alpha_i u_i = \sum_{i=0}^{\infty} \beta_i u_i \Longrightarrow \alpha_i = \beta_i \quad \text{for all} \quad i.$$

IV. Bounded, Self-Adjoint, and Compact Operators. Let H be a (real or complex) pre-Hilbert space and $T : H \to H$ a linear operator. T is called *bounded* if the *norm of* T,

$$\|T\| := \sup\{\|Tf\| : f \in H, \ \|f\| = 1\},$$

is finite. In this case,

$$\|Tf\| \le \|T\| \cdot \|f\| \quad \text{for all} \quad f \in H. \tag{6}$$

If T is linear and bounded and

$$(Tf, g) = (f, Tg) \quad \text{for} \quad f, g \in H,$$

then T is called *self-adjoint* or *Hermitian.*

A linear operator T is called *compact* if for every bounded sequence (f_n) from H, the sequence (Tf_n) has a convergent subsequence (with limit in H). It is easy to see that a compact linear operator is bounded.

(a) *If T is a self-adjoint operator T, then (Tf, f) is real for every $f \in H$, and*

$$\|T\| = \sup\{|(Tf, f)| : f \in H, \ \|f\| = 1\}.$$

Proof. Denote the right side of this equation by β. Then clearly,

$$|(Tf, f)| \le \beta \|f\|^2 \quad \text{for} \quad f \in H. \tag{7}$$

By (6) and the Schwarz inequality we have

$$|(Tf, f)| \le \|Tf\| \le \|T\| \ \text{ for } \|f\| = 1, \text{ whence } \ \beta \le \|T\|.$$

The proof of the reverse inequality follows from the identity

$$(Tf + Tg, f + g) - (Tf - Tg, f - g) = 2(Tf, g) + 2(Tg, f).$$

The left side is

$$\leq \beta\|f+g\|^2 + \beta\|f-g\|^2 = 2\beta(\|f\|^2 + \|g\|^2)$$

by (7) and the parallelogram identity. Using the particular choice $f = \lambda h$, $g = Th$ with $\lambda = \|Th\|$, $\|h\| = 1$, one then obtains

$$2(Tf,g) + 2(Tg,f) = 2\lambda(Th,Th) + 2\lambda(T^2h,h) = 4\lambda^3,$$

and hence

$$4\lambda^3 \leq 2\beta(\lambda^2+\lambda^2) \Longrightarrow \lambda = \|Th\| \leq \beta.$$

Since h with $\|h\| = 1$ is arbitrary, it follows that $\|T\| \leq \beta$; and therefore, $\|T\| = \beta$. ∎

V. Eigenvalues of Compact Self-Adjoint Operators. If the equation

$$Tu = \mu u \quad \text{with} \quad 0 \neq u \in H \tag{8}$$

holds, then μ is called an *eigenvalue* of T and u a corresponding *eigenelement*.

Let T be a compact self-adjoint operator. To determine an eigenvalue, we consider, if $T \neq 0$, a sequence (ϕ_n) from H such that

$$\|\phi_n\| = 1, \quad |(T\phi_n,\phi_n)| \to \|T\| \quad \text{as} \quad n \to \infty,$$

cf. IV.(a). By passing to a subsequence if necessary, we assume further that the sequence of real numbers $(T\phi_n,\phi_n)$ and the sequence $(T\phi_n)$ both converge (T is compact!),

$$(T\phi_n,\phi_n) \to \mu, \quad T\phi_n \to \mu u.$$

Here μ is real and $|\mu| = \|T\| > 0$. Now we have

$$\begin{aligned} 0 &\leq \|T\phi_n - \mu\phi_n\|^2 = \|T\phi_n\|^2 - 2\mu(T\phi_n,\phi_n) + \mu^2 \\ &\leq 2\mu^2 - 2\mu(T\phi_n,\phi_n) \to 0; \end{aligned}$$

that is,

$$T\phi_n = \mu\phi_n + \varepsilon_n \quad \text{with} \quad \varepsilon_n \in H, \ \|\varepsilon_n\| \to 0.$$

Since $T\phi_n \to \mu u$, we have $\mu\phi_n \to \mu u$; thus $\phi_n \to u$, and therefore $T\phi_n \to Tu$. It follows that equation (8) holds and $\|u\| = 1$.

VI. Theorem. *If T is a compact self-adjoint operator in the inner product space H, then T has an eigenvalue $\mu_0 \in \mathbb{R}$ with $|\mu_0| = \|T\|$. The corresponding eigenelement $u_0 \in H$ with*

$$Tu_0 = \mu_0 u_0, \ \|u_0\| = 1$$

has the property that the expression $|(Tu, u)|$ assumes its maximum on the unit ball at the point u_0. Since (8) implies that $(Tu, u) = \mu\|u\|^2$, every eigenvalue μ of T is real, and it satisfies $|\mu| \le \|T\|$.

Proof. This theorem was proved for $T \neq 0$ in V; for $T = 0$ it is trivial. ∎

Consider now the subspace H_1 of all elements $f \in H$ that are orthogonal to u_0:

$$H_1 := \{f \in H : (f, u_0) = 0\}.$$

It is easily seen that H_1 is a closed subspace of H. Furthermore, T maps H_1 into itself because

$$(Tf, u_0) = (f, Tu_0) = \mu_0(f, u_0) = 0 \quad \text{for} \quad f \in H_1,$$

and T is self-adjoint and compact in H_1.

Now Theorem VI can be applied to H_1. Thus there is an eigenvalue μ_1 and an eigenelement u_1 with

$$|\mu_0| \ge |\mu_1|, \ (u_0, u_1) = 0, \ \|u_1\| = 1.$$

Now let H_2 be the subspace of all elements $f \in H$ that are orthogonal to u_0 and u_1, etc. This procedure terminates only if the subspace H_n of elements f with

$$H_n : (f, u_i) = 0 \quad \text{for} \quad i = 0, 1, \ldots, n-1$$

is $\{0\}$. That is impossible in an infinite-dimensional space.

VII. Theorem. *Let H be an infinite-dimensional inner product space and $T : H \to H$ be linear, self-adjoint, and compact. Then the eigenvalue problem (8) has countably many real eigenvalues $\mu_0, \mu_1, \ldots$ with*

$$|\mu_0| \ge |\mu_1| \ge |\mu_2| \ge \cdots \quad \text{and} \quad \mu_n \to 0 \quad \text{as} \quad n \to \infty. \tag{9}$$

The corresponding eigenelements u_n,

$$Tu_n = \mu_n u_n,$$

form (with a suitable normalization) an orthonormal system,

$$(u_m, u_n) = \begin{cases} 1 & \text{for} \quad n = m, \\ 0 & \text{for} \quad n \neq m. \end{cases}$$

If H_n is the space of all $f \in H$ such that

$$(f, u_i) = 0 \quad \text{for} \quad i = 0, \ldots, n-1,$$

then

$$|\mu_n| = \sup \|Tf\| = \sup |(Tf, f)| \qquad (f \in H_n,\ \|f\| = 1) \tag{10}$$

and $(Tu_n, u_n) = \mu_n$; i.e., the supremum is assumed for $f = u_n$.

Each element in the image space of T is represented by its Fourier series, i.e., if $f \in H$, then

$$Tf = \sum_{i=0}^{\infty} d_i u_i \quad \text{with} \quad d_i = (h, u_i) = \mu_i (f, u_i). \tag{11}$$

The *proof* of this theorem, up to (11) and the limit relation in (9), is contained in the previous remarks. The sequence (μ_n) converges to 0 because otherwise the sequence $\psi_n = \frac{1}{\mu_n} u_n$ would be bounded, and then the sequence $(T\psi_n) = (u_n)$ would possess a convergent subsequence, which is impossible, since $\|u_n - u_m\| = 2$ for $m \neq n$.

Finally, in order to prove (11), we consider the function

$$g_n = \sum_{i=0}^{n-1} c_i u_i, \quad c_i = (f, u_i).$$

Clearly, $g_n \in H_n$ holds; therefore, by (10), (4), and (9),

$$\|Tg_n\| \leq |\mu_n| \cdot \|g_n\| \leq |\mu_n| \cdot \|f\| \to 0.$$

The conclusion now follows from the equation

$$Tf - \sum_{i=0}^{n-1} d_i u_i = Tg_n.$$

∎

Addendum. *Every eigenvalue $\mu \neq 0$ is equal to some μ_n, and the corresponding eigenspace (that is the set of all $u \in H$ that satisfy* (8)*) has finite dimension and is spanned by the eigenelements u_k corresponding to $\mu_k = \mu$.*

Proof. If u is a solution of (8) with $\mu \neq 0$, then u lies in the image of T, i.e., we have

$$u = \sum c_i u_i \quad \text{with} \quad c_i = (u, u_i), \quad Tu = \sum c_i \mu_i u_i.$$

Because of (8) and III.(f), the relation $\mu c_i = \mu_i c_i$ holds for all i. If $\mu \neq \mu_i$ for all i, then $c_i = 0$, and hence $u = 0$. If $\mu = \mu_n$, then $c_i = 0$ for all i with $\mu_i \neq \mu$ and $u = \sum c_k u_k$, where the sum extends over all k with $\mu_k = \mu$. ∎

For $T = 0$, the theorem is true, but not interesting.

VIII. Theorem. *If H is a Hilbert space and $\mu = 0$ is not an eigenvalue of T, then (u_n) is an orthonormal basis; i.e., a representation*

$$f = \sum_{i=0}^{\infty} c_i u_i \quad \text{with} \quad c_i = (f, u_i)$$

holds for all $f \in H$.

The equation $f = \sum c_i u_i$ also holds in a pre-Hilbert space if $\mu = 0$ is not an eigenvalue and the series belongs to H.

Proof. By III.(b), the Fourier series of f is convergent, say to g. Thus, by III.(f), $c_i = (g, u_i)$. It follows that Tf and Tg have the same Fourier coefficients $\mu_i c_i$ and hence are equal by the conclusion (11) of the theorem. From $T(f-g) = 0$ it follows that $f = g$, since 0 is not an eigenvalue of T. ∎

We apply these results now to

IX. The Sturm–Liouville Eigenvalue Problem. We consider the problem

$$Lu + \lambda r u = 0 \quad \text{in} \quad J = [a, b], \; R_1 u = R_2 u = 0, \tag{12}$$

where $Lu = (pu')' + qu$ and

$$R_1 u = \alpha_1 u(a) + \alpha_2 p(a) u'(a),$$
$$R_2 u = \beta_1 u(b) + \beta_2 p(b) u'(b)$$

under assumption (SL) of 27.I. Suppose λ^* is not an eigenvalue and $q(x)$ is replaced by $q^*(x) = q(x) + \lambda^* r(x)$. If (λ_n, u_n) are the eigenvalues and eigenfunctions for the original problem, then those for the new problem are $(\lambda_n - \lambda^*, u_n)$. In particular, 0 is not an eigenvalue for the new problem. Therefore, we will assume, without loss of generality, that $\lambda = 0$ is not an eigenvalue.

A solution u of (12) can be interpreted as a solution of the semihomogeneous Sturmian boundary value problem

$$Lu = g(x) \quad \text{with} \quad g(x) = -\lambda r(x) u(x),$$

$R_1 u = R_2 u = 0$. Thus, by (26.12), u satisfies the integral equation

$$u(x) = -\lambda \int_a^b \Gamma(x, \xi) r(\xi) u(\xi) \, d\xi. \tag{13}$$

Here $\Gamma(x, \xi)$ is Green's function for the Sturmian boundary value problem (26.4), whose existence is guaranteed by Theorem 26.VII, since $\lambda = 0$ is not an eigenvalue ($Lu = 0$, $R_i u = 0$ has only the trivial solution).

The relationship between the original problem and the integral equation is clarified in the following

X. Theorem. *Let assumption* (SL) *from* 27.I *hold, and suppose that* 0 *is not an eigenvalue of* (12). *Then* λ *is an eigenvalue and the function* $u(x)$ *is a corresponding eigenfunction if and only if* u *is continuous in* J *and* $\not\equiv 0$ *and satisfies the integral equation* (13).

The *proof* of Theorem X is contained, for the most part, in the above discussion. Just one small hole needs to be closed. If one wants to show that a solution u of (13) also represents a solution of (12), then one must first check that $u \in C^2(J)$, since u is only assumed to be continuous. However, this follows from Theorem 26.VII, since the integral on the right-hand side has the form (26.12) with $g = -\lambda r u$, and as it was proved there, this integral is twice continuously differentiable for continuous g. ∎

We have thus transformed the original eigenvalue problem into an analogous problem for (13). Equation (13) is called a *Fredholm integral equation.* (Fredholm integral equations are those with fixed limits of integration; those with variable limits, such as arise with initial value problems, are called Volterra integral equations.)

Let the operator T be defined by the relation

$$(Tf)(x) = -\int_a^b \Gamma(x,\xi) r(\xi) f(\xi)\, d\xi. \tag{14}$$

Then from Theorem 26.VII we get the equivalence

$$v = Tf \iff Lv + rf = 0, \quad R_1 u = R_2 u = 0. \tag{15}$$

If both sides of (13) are multiplied by $1/\lambda$, then

$$Tu = \mu u \quad \text{with} \quad \mu = 1/\lambda. \tag{16}$$

We now consider this equation in the real inner product space $H = C(J)$ of Example II.(b) and apply the earlier results. The operator T maps $C(J)$ to itself. From $Tf = 0$ we conclude, using (15), that $f = 0$; i.e., $\mu = 0$ is not an eigenvalue of T. Since $\lambda = 0$ is not an eigenvalue of (12), there is a one-to-one correspondence between the eigenvalues λ of (12) and μ of (16) given by $\lambda = 1/\mu$.

The discussion can be simplified if one uses a weighted inner product

$$(f,g)_r = \int_a^b r(x) f(x) g(x)\, dx \tag{17}$$

in the space $C(J)$ instead of the inner product (f,g) of Example II.(b) (for a first reading, the reader can take $r = 1$ without missing any essentials). First of all, it follows from our general assumptions (SL) in 27.I that there exist positive constants α, β with

$$0 < \alpha \le r(x) \le \beta \quad \text{in} \quad J.$$

Thus the weighted norm

$$\|f\|_r = (f,f)_r^{1/2} = \left(\int_a^b r(x) f^2(x)\, dx \right)^{1/2} \tag{18}$$

and the usual one $\|\cdot\|$ generated by the inner product (f,g) satisfy the relation

$$\alpha\|f\| \le \|f\|_r \le \beta\|f\|;$$

i.e., the two norms are equivalent; cf. 5.V or 10.III. We denote the space $C(J)$, equipped with the inner product $(f,g)_r$, by H_r.

The operator T is linear, self-adjoint, and compact. The self-adjointness follows from the symmetry of Γ,

$$(Tf,g)_r = -\int_a^b r(x)g(x) \int_a^b \Gamma(x,\xi)r(\xi)f(\xi)\, d\xi dx = (f,Tg)_r.$$

The compactness of T is contained in the following lemma.

XI. Lemma. *If (f_n) is a sequence in $C(J)$ with $\|f\|_r \le C$, then the sequence*

$$g_n(x) = Tf_n = -\int_a^b \Gamma(x,\xi)r(\xi)f_n(\xi)\, d\xi$$

satisfies the hypotheses of the Ascoli–Arzela theorem 7.IV*; i.e., it is equicontinuous and uniformly bounded,*

$$|g_n(x)| \le C_1 \quad \textit{for all} \quad x \in J,\ n \in \mathbb{N}.$$

Hence the sequence (g_n) has a subsequence that converges uniformly in J and therefore also in H_r to a function $g \in C(J)$.

Proof. Because of the continuity of $\Gamma(x,\xi)$, for $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|\Gamma(x,\xi) - \Gamma(x',\xi)| < \varepsilon \quad \text{for} \quad |x - x'| < \delta.$$

Therefore, if $g = Tf$ and $\|f\|_r \le C$, then by the Schwarz inequality,

$$|g(x) - g(x')| \le \int_a^b |\Gamma(x,\xi) - \Gamma(x',\xi)| r(\xi) |f(\xi)|\, d\xi$$

$$\le \varepsilon \int_a^b r|f|\, d\xi = \varepsilon(1,|f|)_r \le \varepsilon\|1\|_r\|f\|_r \le C\gamma\varepsilon$$

with $\gamma = \|1\|_r \le \beta\sqrt{b-a}$. This proves equicontinuity. The proof of boundedness is even simpler. The kernel Γ is continuous, hence bounded, $|\Gamma(x,\xi)| \le A$. Thus it follows from the Schwarz inequality that

$$|g_n(x)| = |(\Gamma(x,\cdot), f_n)_r| \le \|A\|_r\|f_n\|_r \le \gamma AC.$$

∎

The equation $Tu_i = \mu_i u_i$ ($\mu_i \neq 0$) can be written using $\lambda_i = 1/\mu_i$ as $u_i = T(\lambda_i u_i)$. By (15), the last equation is equivalent to

$$Lu_i + \lambda_i r(x) u_i = 0, \ R_1 u_i = R_2 u_i = 0 \quad (i = 0, 1, 2, \ldots).$$

This is a new proof for part of the existence theorem 27.II.

If we set aside for the moment the question of convergence, then an expansion of a given function $\phi(x)$ in terms of the eigenfunctions u_i is given by

$$\phi(x) = \sum_{i=0}^{\infty} d_i u_i(x) \text{ with } d_i = (\phi, u_i)_r = \int_a^b r(x)\phi(x)u_i(x)\,dx. \tag{19}$$

This matches the expansion given in Theorem 27.III.

Now let $\phi \in C^2(J)$, $R_1\phi = R_2\phi = 0$. By (26.12),

$$\phi(x) = \int_a^b \Gamma(x,\xi)(L\phi)(\xi)\,d\xi = Tf \quad \text{with} \quad f = -L\phi/r.$$

Therefore, Theorem VII applies, and (11) holds for $\phi = Tf$, which is (19). This equation is to be understood in the sense of convergence in the norm (18). We will now show that in fact, the convergence is uniform in J.

By (11), $d_i = \mu_i c_i$ with $c_i = (f, u_i)$. Further, for fixed x_0, the number $\mu_i u_i(x_0)$ can be interpreted as the Fourier coefficient of the function $-\Gamma(x_0, \xi)$; that is,

$$\mu_i u_i(x_0) = -(\Gamma(x_0, \cdot), u_i)_r.$$

Consider now a partial sum from $i = m$ to $i = n$ of (19) and apply the Schwarz inequality:

$$\left(\sum_{i=m}^{n} c_i \mu_i u_i(x_0)\right)^2 \leq \sum_{i=m}^{n} c_i^2 \sum_{i=m}^{n} (\mu_i u_i(x_0))^2.$$

If the second sum on the right-hand side is extended to 0 and to ∞, then by Bessel's inequality it is $\leq \|\Gamma(x_0, \cdot)\|_r^2$. The first sum on the right-hand side is likewise the partial sum of a convergent series. Thus for any $\varepsilon > 0$, there exists an n_0 such that

$$\left(\sum_{i=m}^{n} c_i \mu_i u_i(x_0)\right)^2 \leq \varepsilon \|\Gamma(x_0, \cdot)\|_r^2 \leq A\varepsilon \text{ for } n > m \geq n_0, \ x_0 \in J.$$

This establishes the uniform convergence of (19). Since convergence holds in H_r, Theorem 27.III follows for the case where $\phi \in C^2(J)$. We note that the theorem is also true for $\phi \in C^1(J)$. However, the proof is more difficult; cf., for instance, Kamke (1945) or Titchmarsh (1962). ∎

Our final result is a general theorem about the Fourier expansion of functions from $L^2(J)$. The proof requires results from the theory of the Lebesgue integral.

XII. Expansion Theorem. *Every function $\phi \in L^2(J)$ can be expanded in a series of eigenfunctions of the Sturm–Liouville problem* (12). *Equation* (19) *holds in the L_2-norm of Example* II.(b),

$$\int_a^b \left(\phi(x) - \sum_{i=0}^{n} d_i u_i(x)\right)^2 dx \longrightarrow 0 \quad \text{as} \quad n \to \infty.$$

The *proof* will be briefly indicated. One considers T as an operator in the real Hilbert space $L^2(J)$ with the inner product given by (17). It is easy to show that T is self-adjoint and compact in $L^2(J)$. Consequently, Theorem VII holds in $L^2(J)$. Now let $f \in L^2(J)$ and suppose $Tf = 0$. If $g \in C^2(J)$ and satisfies $R_1 g = R_2 g = 0$, then there is an $h \in C(J)$ such that $g = Th$, hence $(f, g)_r = (f, Th)_r = (Tf, h)_r = 0$. It follows that $f = 0$, i.e., 0 is not an eigenvalue of T. Then by Theorem VII, (u_i) is an orthonormal basis and the proposed relationship holds, first for the norm $\|\cdot\|_r$ and then also for the L_2-norm, because of the equivalence of the two norms. ∎

Our last theorem deals with the Fourier expansion under conditions that are weaker than (SL) and apply to the radial Δ operator, among others.

XIII. Coefficients with a Zero on the Boundary. We consider the eigenvalue problem (27.22) for the operator $Lu = (pu')'$ using the assumptions and the notation given in 27.XX. The radial elliptic eigenvalue problem for $\Delta u + \lambda u$ is covered as a special case.

In what follows, H_r is the Hilbert space of measurable functions in $J = [0, b]$ with a finite norm (18); the inner product is given in (17). Since $r(0) = 0$ is permitted, we have only $H_r \supset L_2(J)$.

Theorem. *The sequence (u_n) of eigenfunctions of the eigenvalue problem* (27.22) *is an orthonormal basis in H_r; that is, each element of H_r is represented by its Fourier series in the sense of convergence in H_r.*

In the following sketch of the proof, T is the operator defined in (14) $(a = 0)$, where Green's function is taken from 27.XX.(d) and $B = \int_0^b p^{-1}(x)R(x)\,dx < \infty$.

Let $f \in H_r$ and $v = Tf$; that is, $(pv')' + rf = 0$, $(pv')(0) = v(b) = 0$. Using the Cauchy–Schwarz inequality, it follows that

$$|pv'(x)| = \left|\int_0^x rf\,dt\right| = \left|\int_0^x \sqrt{r}\cdot\sqrt{r}\,f\,dt\right| \le \left(\int_0^x r\,dt \cdot \int_0^b rf^2\,dt\right)^{1/2},$$

and therefore

$$|v'(x)| \le \|f\|_r \sqrt{R(x)}/p(x). \tag{$\star$}$$

Writing $\sqrt{R}/p$ in the form $\sqrt{R/p} \cdot \sqrt{1/p}$, one obtains from ($\star$)

$$|v(x)| = \left| \int_x^b v'\,dt \right| \le \|f\|_r \int_x^b \frac{\sqrt{R}}{p}\,dt \le \|f\|_r \left(\int_x^b \frac{R}{p}\,dt \cdot \int_x^b \frac{1}{p}\,dt \right)^{1/2},$$

which implies

$$|v(x)| \le \|f\|_r \sqrt{B} \sqrt{|h(x)|}, \quad \text{where} \quad h(x) = -\int_x^b \frac{1}{p}\,dt,$$

and

$$\|v\|_r^2 \le \|f\|_r^2 B \int_0^b r|h|\,dx = B^2 \|f\|_r^2$$

(reverse the order of integration in the last integral).

Now all necessary estimates are assembled. Let (f_n) be a bounded sequence in H_r, $C = \sup \|f\|_r$, and $v_n = Tf_n$. The bounds on $v(x)$ and $v'(x)$ show that the Ascoli–Arzelà theorem can be applied in intervals $[\varepsilon, b]$ with $\varepsilon > 0$. It follows in a familiar way that a subsequence of (v_n) converges in $J_0 = [(0, b]$ to a function $v \in C^0(J_0)$ (one chooses a subsequence (v_n^1) converging in $\left[\frac{1}{2}, b\right]$, from that subseqence chooses a subsequence (v_n^2) converging in $\left[\frac{1}{3}, b\right], \ldots$, and considers the sequence (v_n^n)). Since $r(x)v_n^2(x) \le g(x) = BC^2|h(x)|r(x) \in L(J)$, the limit v satisfies the same inequality. Hence $v \in H_r$ and $\|v_n - v\|_r \to 0$ as $n \to \infty$ due to the theorem on majorized convergence. This shows that the operator T is *compact*; it is also self-adjoint, and 0 is not an eigenvalue of T because of 27.XX.(c). Now the theorem follows from Theorem VIII. ■

Exercise. Replace the boundary condition $u(b) = 0$ in (27.22) by $R_2 u = \beta_1 u(b) + \beta_2 p(b) u'(b) = 0$ and prove the corresponding theorem on the completeness of (u_n).

The eigenvalue problem takes on particular significance in connection with certain partial differential equations that play an important role in physics.

XIV. Partial Differential Equations. (a) We begin with the *parabolic differential equation* for the function $\phi = \phi(t, x)$,

$$\phi_t = \frac{1}{r(x)} \left[(p(x)\phi_x)_x + q(x)\phi \right] \quad \text{for} \quad a < x < b,\ t > 0, \tag{20}$$

with the boundary conditions

$$\begin{aligned} R_1\phi &:= \alpha_1 \phi(t, a) + \alpha_2 p(a) \phi_x(t, a) = 0, \\ R_2\phi &:= \beta_1 \phi(t, b) + \beta_2 p(b) \phi_x(t, b) = 0 \end{aligned} \tag{21}$$

and the initial condition

$$\phi(0, x) = f(x) \quad \text{for} \quad a \le x \le b. \tag{22}$$

If $p = \text{const}$, $r = \text{const}$, $q = 0$, and $\alpha_2 = \beta_2 = 0$, these equations describe the temperature distribution in a homogeneous rod of length $b - a$ whose initial temperature is equal to $f(x)$ and whose ends are held at temperature zero.

A product ansatz (or, as it is also sometimes called, a separation of variables ansatz) $\phi(t,x) = h(t)u(x)$ for a solution of (20) leads to

$$h'u = \frac{h}{r}[(pu')' + qu].$$

If one divides here by the product hu, then the functions to the left of the equal sign depend only on t and those on the right only on x. This equation can be valid (after dividing by hu) only if the left- and right-hand sides are constant. We call this constant $-\lambda$ and obtain the equations

$$\begin{aligned} h' + \lambda h = 0 \quad &\text{for} \quad h = h(t), \\ (pu')' + qu + \lambda r u = 0 \quad &\text{for} \quad u = u(x). \end{aligned}$$

If in addition, we require that $\phi(t,x) = h(t)u(x)$ satisfy the boundary conditions (21), then we must have $R_1 u = R_2 u = 0$. Thus we obtain the eigenvalue problem (12) for u. If λ_n is an eigenvalue and u_n the corresponding eigenfunction, then the product

$$\phi_n(t,x) = \mathrm{e}^{-\lambda_n t} u_n(x)$$

is a solution of (20) that satisfies the boundary conditions (21). The same also holds for a linear combination of the ϕ_n and—assuming appropriate convergence behavior—for the infinite series

$$\phi(t,x) = \sum_{n=0}^{\infty} c_n \phi_n(t,x) = \sum_{n=0}^{\infty} c_n \mathrm{e}^{-\lambda_n t} u_n(x). \tag{23}$$

The initial condition (22) then leads to the equation

$$\phi(0,x) = f(x) = \sum_{n=0}^{\infty} c_n u_n(x),$$

which is just the Fourier series for f with respect to the orthogonal system (u_n). We summarize:

The solution to the initial-boundary value problem (20)–(22) *is obtained (at first formally) as an infinite series of the form* (23), *where the coefficients c_n are the Fourier coefficients of the function f with respect to the orthonormal system of eigenfunctions (u_n) to the eigenvalue problem* (12).

(b) *An Example.* In the case of the heat equation

$$\phi_t = \phi_{xx} \quad \text{for} \quad 0 < x < \pi,\ t > 0$$

with the boundary conditions $\phi(t,0) = \phi(t,\pi) = 0$ and the initial condition (22), the procedure described above leads to $\lambda_n = n^2$, $u_n = \sqrt{2/\pi}\sin nt$ ($n \in \mathbb{N}$) and hence to the solution

$$\phi(t,x) = \alpha \sum_{n=1}^{\infty} c_n e^{-n^2 t} \sin nx \quad \text{with} \quad c_n = \alpha \int_0^{\pi} f(x) \sin nx \, dx,$$

where $\alpha = \sqrt{2/\pi}$.

A number of conclusions about the behavior of the solution can be obtained from this representation, for example, the estimate

$$\phi^2(t,x) \le \alpha^2 \sum_n c_n^2 \cdot \sum_n e^{-2n^2 t} \le \alpha^2 \int_0^{\pi} f^2(x)\, dx \cdot \frac{e^{-2t}}{1 - e^{-2t}}.$$

Here we have made use of Cauchy's inequality and Bessel's inequality. Thus $|\phi(t,x)| \le Ce^{-t}$. However, if $\int_0^{\pi} f(x) \sin x \, dx = 0$, then it follows that $|\phi(t,x)| \le C \cdot e^{-4t}$ (Proof?).

This is not the place to deal with the questions of the convergence of the series (23) and the existence of $\phi_t, \ldots$. Instead, we will consider another example,

(c) *The hyperbolic differential equation*

$$\phi_{tt} = \frac{1}{r(x)}\left[(p(x)\phi_x)_x + q(x)\phi\right] \quad \text{for} \quad a < x < b,\ t > 0$$

with the boundary condition (21) and initial condition

$$\phi(0,x) = f(x) \quad \text{and} \quad \phi_t(0,x) = g(x) \quad \text{for} \quad a \le x \le b. \tag{24}$$

The product ansatz $\phi(t,x) = h(t)u(x)$ now leads to the equation

$$h'' + \lambda h = 0 \quad \text{for} \quad h = h(t)$$

and the eigenvalue problem (12) for $u(x)$. Corresponding to each eigenvalue λ_n there are two solutions

$$\phi_n = u_n(x) \cos\sqrt{\lambda_n}\, t,\ \psi_n = u_n(x) \sin\sqrt{\lambda_n}\, t, \quad \text{provided that} \quad \lambda_n \ge 0.$$

A corresponding series ansatz has the form

$$\phi(t,x) = \sum_{n=0}^{\infty} c_n \phi_n + \sum_{n=0}^{\infty} d_n \psi_n,$$

which, together with the initial conditions (24), leads to the relations

$$f(x) = \sum_{n=0}^{\infty} c_n u_n(x), \quad g(x) = \sum_{n=0}^{\infty} \sqrt{\lambda_n} d_n u_n(x).$$

Exercise. Determine the formulas for c_n, d_n, and ϕ for the equation of the vibrating string

$$\phi_{tt} = \phi_{xx} \quad \text{for} \quad 0 < x < \pi,\ t > 0$$

with boundary conditions of the first kind: $\phi(t,0) = \phi(t,\pi) = 0$.

(d) As an example of a partial differential equation in several "space variables," we consider the heat equation

$$\phi_t = \Delta\phi \quad \text{for} \quad \phi = \phi(t, \boldsymbol{\xi}) = \phi(t, \xi_1, \ldots, \xi_n)$$

with $\Delta\phi = \phi_{\xi_1\xi_1} + \phi_{\xi_2\xi_2} + \cdots + \phi_{\xi_n\xi_n}$. Suppose that rotationally symmetric initial values are prescribed in the unit ball,

$$\phi(0, \boldsymbol{\xi}) = f(|\boldsymbol{\xi}|_e) \quad \text{for} \quad 0 \le |\boldsymbol{\xi}|_e \le 1,$$

and that the boundary values are given by

$$\phi(t, \boldsymbol{\xi}) = 0 \quad \text{for} \quad |\boldsymbol{\xi}|_e = 1, \quad t > 0.$$

According to 6.XIV,

$$\Delta u(|\boldsymbol{\xi}|_e) = u'' + \frac{n-1}{x} u' = x^{1-n}(x^{n-1}u')', \quad x = |\boldsymbol{\xi}|_e.$$

The ansatz $\phi(t, \boldsymbol{\xi}) = h(t)u(x)$ leads to the equation $h' + \lambda h = 0$ for $h(t)$ and the equation

$$(x^{n-1}u')' + \lambda x^{n-1} u = 0 \ \text{ for } \ u(x).$$

As boundary conditions one has $u'(0) = 0$ and $u(1) = 0$; cf. Lemma 6.XIV. Thus we are led to the eigenvalue problem (27.18) with $\alpha = n - 1$, for which the existence of eigenvalues was treated in 27.XVIII and the expansion theorem was proved in XIII. One obtains the solution in the form

$$\phi(t, \xi) = \sum_{n=0}^{\infty} c_n e^{-\lambda_n t} u_n(x) \ \text{ with } \ f(x) = \sum_{n=0}^{\infty} c_n u_n(x) \ \text{ and } \ x = |\xi|_e.$$

In the case $n = 2$, which corresponds to $\alpha = 1$, the differential equation in (27.17) is Bessel's equation of order 0. The solution y is the Bessel Function $J_0(x)$. Its zeros $0 < \xi_0 < \xi_1 < \cdots$ lead to the positive eigenvalues $\lambda_n = \xi_n^2$ and the corresponding eigenfunctions

$$u_n(x) = \alpha_n J_0(\xi_n x), \qquad n = 0, 1, 2, \ldots.$$

The normalization factor α_n is obtained from the relation

$$\int_0^1 x J_0^2(\xi_n x)\, dx = \frac{1}{2}|J_0'(\xi_n)|^2 \Longrightarrow \alpha_n = \sqrt{2}/|J_0'(\xi_n)|,$$

given without proof. From the expansion

$$f(x) = \sum_{n=0}^{\infty} c_n \alpha_n J_0(\xi_n x) \quad \text{with} \quad c_n = \alpha_n \int_0^1 x f(x) J_0(\xi_n x)\, dx,$$

one obtains the solution of the boundary value problem

$$\phi(t, \boldsymbol{\xi}) = \sum_{n=0}^{\infty} c_n \alpha_n \mathrm{e}^{-\xi_n^2 t} J_0(\xi_n x) \quad \text{with} \quad x = |\boldsymbol{\xi}|_e.$$

(e) *Exercise.* In example (b) the boundary conditions are changed to

$$\phi_x(t, 0) = \phi_x(t, \pi) = 0.$$

Physically, this means that the ends of the rod are thermally insulated; i.e., there is no heat flux from the ends. Solve the boundary value problem.

Chapter VII
Stability and Asymptotic Behavior

§ 29. Stability

I. Stability Theory. We resume the problems treated in §12. In contrast to the case investigated there, we now consider solutions defined on infinite intervals. In this setting, continuous dependence on initial conditions and on the right side of the differential equation is a significantly more complicated matter than in §12, where general results were obtained under restricted assumptions. Even in the simplest examples, new phenomena emerge when the interval is infinite.

Two Examples. Let $y(t)$ be the solution of

$$y' = y, \quad y(0) = \eta$$

and $z(t)$ be a solution with the initial value $z(0) = \eta + \varepsilon$. Then

$$z(t) - y(t) = \varepsilon \mathrm{e}^t,$$

i.e., the difference between two solutions to the same differential equation with different initial conditions tends to ∞ like e^t.

On the other hand, if y and z are two solutions of the differential equation

$$y' = -y$$

with initial values η and $\eta + \varepsilon$, then the difference is given by

$$z(t) - y(t) = \varepsilon \mathrm{e}^{-t},$$

and hence converges to 0 as $t \to \infty$.

Our goal in the present section is to give criteria that guarantee that solutions depend continuously on initial conditions in the sense that if the difference $z(0) - y(0)$ is small, then the function $z(t) - y(t)$ also remains small in the *whole* interval $t \geq 0$. Statements of this kind fall under the heading of "stability theory" for ordinary differential equations.

II. Stability, Asymptotic Stability. In the following, t is a real variable; the functions $\mathbf{f}$, $\mathbf{x}$, $\mathbf{y}$ can have values in either $\mathbb{R}^n$ or $\mathbb{C}^n$.

Let $\mathbf{x}(t)$ be a solution of the system

$$\mathbf{y}' = \mathbf{f}(t, \mathbf{y}) \tag{1}$$

for $0 \leq t < \infty$. We assume that $\mathbf{f}(t, \mathbf{y})$ is defined and continuous at least in $S_\alpha : 0 \leq t < \infty$, $|\mathbf{y} - \mathbf{x}(t)| < \alpha$ $(\alpha > 0)$. The solution $\mathbf{x}(t)$ is said to be *stable (in the sense of Lyapunov)* if the following statement is true:

For every $\varepsilon > 0$, there exists $\delta > 0$ such that every solution $\mathbf{y}(t)$ with

$$|\mathbf{y}(0) - \mathbf{x}(0)| < \delta$$

exists for all $t \geq 0$ and satisfies the inequality

$$|\mathbf{y}(t) - \mathbf{x}(t)| < \varepsilon \quad \text{for} \quad 0 \leq t < \infty.$$

A solution $\mathbf{x}(t)$ is called *asymptotically stable* if it is stable and if there exists $\beta > 0$ such that every solution $\mathbf{y}(t)$ with $|\mathbf{y}(0) - \mathbf{x}(0)| < \beta$ satisfies

$$\lim_{t \to \infty} |\mathbf{y}(t) - \mathbf{x}(t)| = 0.$$

A solution $\mathbf{x}(t)$ is called *unstable* if it is not stable.

More generally, one can consider a fundamental interval $[a, \infty)$ and replace $\mathbf{y}(0) - \mathbf{x}(0)$ by $\mathbf{y}(a) - \mathbf{x}(a)$ and $0 \leq t < \infty$ by $a \leq t < \infty$. This raises now a question: Are the stability definitions for the interval $[0, \infty)$ equivalent to the corresponding definitions for $[a, \infty)$? If one assumes additionally that $\mathbf{f}$ is locally Lipschitz continuous in $\mathbf{y}$, then the answer to this question is positive. The proof will be given in the next section; it can be omitted in a first reading.

Remarks. The norm $|\cdot|$ in these definitions is an arbitrary norm in $\mathbb{R}^n$ or $\mathbb{C}^n$. Using Theorem 10.III, it is easy to show that the definitions are independent of the choice of norm.

It is customary in stability theory to formulate the theorems for the case $t \to +\infty$; results for $t \to -\infty$ can be reduced to this case.

III. The Poincaré Map. The solution to equation (1) with the initial value $\mathbf{y}(t) = \boldsymbol{\eta}$ will be denoted by $\mathbf{y}(t; \tau, \boldsymbol{\eta})$ (uniqueness of solutions to initial value problems is assumed). Let $t = a$ and $t = b$ be two fixed points. The Poincaré map P associates an initial value at the point a with the value of the corresponding solution at the point b; in terms of formulas,

$$\boldsymbol{\eta} \to P\boldsymbol{\eta} = \mathbf{y}(b; a, \boldsymbol{\eta}) \qquad \textit{Poincaré map.} \tag{2}$$

From the uniqueness assumption, it follows that P is injective. We will assume in the following that $\mathbf{f}$ has values in $\mathbb{R}^n$; the extension to $\mathbb{C}^n$ is straightforward. Let $\mathbf{f}$ be defined on a set D that is open in the strip $S = [a, b] \times \mathbb{R}^n$, i.e., is the intersection of an open set with S.

Theorem. *Let* $\mathbf{f} : D \to \mathbb{R}^n$ *be continuous and locally Lipschitz continuous with respect to* $\mathbf{y}$. *Let the set* M *of all solutions of* (1) *that exist in all of the interval* $[a, b]$ *be nonempty. Then the sets* $M_a = \{\mathbf{y}(a) : \mathbf{y} \in M\}$ *and* $M_b = \{\mathbf{y}(b) : \mathbf{y} \in M\}$ *are open, and the Poincaré map* $P : M_a \to M_b$ *is a homeomorphism (i.e.,* P *is bijective,* P *and* P^{-1} *are continuous).*

Proof. Let $J = [a, b]$ and $z(t) \in M$. As in 13.X, we first determine an $\alpha > 0$ with $S_\alpha = \{(t, \mathbf{y}) : t \in J, |\mathbf{y} - \mathbf{z}(t)| \leq \alpha\} \subset D$ and extend $\mathbf{f}$ to the set $J \times \mathbb{R}^n$ while preserving the values in S_a, say, by setting

$$\mathbf{f}^*(t, \mathbf{y}) = \mathbf{f}(t, \mathbf{z}(t) + (\mathbf{y} - \mathbf{z}(t))h(|\mathbf{y} - \mathbf{z}(t)|))$$

with $h(s) = 1$ for $0 < s \leq \alpha$ and $h(s) = \alpha/s$ for $s > \alpha$. For any $(t, \mathbf{y}) \in J \times \mathbb{R}^n$, the argument of $\mathbf{f}$ appearing in the above formula belongs to S_α, i.e., $\mathbf{f}^*$ is defined in all of $J \times \mathbb{R}^n$. Further, $\mathbf{f} = \mathbf{f}^*$ in S_α, and $\mathbf{f}^*$ is Lipschitz continuous with respect to $\mathbf{y}$ in $J \times \mathbb{R}^n$.

Applying Theorem 13.II with $\mathbf{k} = \mathbf{f}^*$, $\lambda = \boldsymbol{\eta}$, $\mathbf{g}(x, \lambda) = \boldsymbol{\eta}$, $\alpha(\lambda) = a$ leads to the conclusion that the solution $\mathbf{y}^*(t; a, \boldsymbol{\eta})$ of (1) depends continuously on $(t; a, \boldsymbol{\eta})$. In particular, there exists $\delta > 0$ such that if $|\boldsymbol{\eta} - \mathbf{z}(a)| < \delta$, then $|\mathbf{y}^*(t; a, \boldsymbol{\eta}) - \mathbf{z}(t)| < \alpha$ in J. Therefore, if $\boldsymbol{\eta}$ is in this range, then $\mathbf{y}^*(t; a, \boldsymbol{\eta}) = \mathbf{y}(t; a, \boldsymbol{\eta}) \in M$ and $P\boldsymbol{\eta} = \mathbf{y}(b; a, \boldsymbol{\eta})$ is continuous. Since $\mathbf{z}(t) \in M$ is arbitrary, it follows that M_a is open and P is continuous in M_a. The continuity of P^{-1} and the openness of M_b follow in a corresponding manner. ∎

The question raised in II is now easily answered

Corollary. *Let the function* $\mathbf{f}$ *satisfy the assumptions for* (1) *and be locally Lipschitz continuous with respect to* $\mathbf{y}$. *A solution* $\mathbf{x}(t)$ *defined in* $[0, \infty)$ *is stable relative to* $t = 0$ *if and only if it is stable relative to an arbitrary point* $t = b$, $b > 0$.

Corresponding statements hold with respect to asymptotic stability and instability.

Proof. Suppose $\mathbf{x}(t)$ is stable relative to $t = b$. Then for any $\varepsilon > 0$, there exists a neighborhood U of $\mathbf{x}(b)$ such that if $\mathbf{y}(b) \in U$, then $|\mathbf{y}(t) - \mathbf{x}(t)| < \varepsilon$ for $t \geq b$. We apply the theorem with $a = 0$. Since the set M is not empty ($\mathbf{x}(t)$ belongs to M), it follows that $V = U \cap M_b$ is a neighborhood of $\mathbf{x}(b)$ and $W := P^{-1}(V)$ is a neighborhood of $\mathbf{x}(0)$. Making W smaller as needed, one obtains that the inequality $|\mathbf{y}(t) - \mathbf{x}(t)| < \varepsilon$ holds in $[0, b]$ (Theorem 13.II). However, because $P(W) \subset U$, the inequality holds in the whole interval $[0, \infty)$; i.e., $\mathbf{x}(t)$ is stable relative to $t = 0$. The proof of the converse is simpler and will be left to the reader. ∎

IV. Linear Systems. In the (real or complex) linear system

$$\mathbf{y}' = A(t)\mathbf{y} + \mathbf{b}(t) \tag{3}$$

let $A(t)$ and $\mathbf{b}(t)$ be continuous in the interval $J = [0, \infty)$. Then every solution exists in J (Theorem 14.VI). Let $X(t)$ be the fundamental system for the homogeneous equation with the initial value $X(0) = I$. We begin by clarifying the relation between the current definition of stability and the one given earlier.

(a) *The definition of stability for the zero solution of the homogeneous equation given in* 17.XI *agrees with the definition given in* II.

Proof. If the zero solution is stable in the sence of 17.XI, then there exists a $K > 0$ with $|X(t)| \le K$ in J. If $\mathbf{y}$ is an arbitrary solution, then the estimate $|\mathbf{y}(t)| \le K|\mathbf{y}(0)|$ follows from the representation $\mathbf{y}(t) = X(t)\mathbf{y}(0)$; hence $|\mathbf{y}(t)| \le \varepsilon$ in J if $|\mathbf{y}(0)| \le \delta := \varepsilon/K$. Therefore, the zero solution is stable according to Definition II. The converse is proved in a similar manner. The proofs for the two remaining cases are left to the reader. ∎

Theorem. *If the zero solution of the homogeneous equation* $\mathbf{y}' = A(t)\mathbf{y}$ *is stable, then every solution of the nonhomogeneous equation* (3) *is also stable. A corresponding theorem holds for asymptotic stability and instability.*

Since the definition of stability deals with the difference $\mathbf{z}(t) = \mathbf{y}(t) - \mathbf{x}(t)$ of two solutions to (3) and since this difference is a solution to the corresponding homogeneous equation, the conclusion follows at once. ∎

Thus when dealing with stability questions for linear systems, it is sufficient to study the stability of the zero solution of the homogeneous equation.

Because of its importance, we reformulate the result obtained in 17.XI for the equation with constant coefficients:

$$\mathbf{y}' = A\mathbf{y} \quad (A \text{ a real or complex } n \times n \text{ matrix}). \tag{4}$$

Stability Theorem. *Let* $\gamma = \max\{\operatorname{Re}\lambda : \lambda \in \sigma(A)\}$. *The stability of the trivial solution* $\mathbf{x}(t) \equiv \mathbf{0}$ *of* (4) *is determined as follows:*

$\gamma < 0 \Rightarrow \mathbf{x}(t)$ *is asymptotically stable;*

$\gamma > 0 \Rightarrow \mathbf{x}(t)$ *is unstable; and*

$\gamma = 0 \Rightarrow \mathbf{x}(t)$ *is not asymptotically stable, and is stable if and only if* $m'(\lambda) = m(\lambda)$ *for all eigenvalues* λ *with* $\operatorname{Re}\lambda = 0$ *(see* 17.VIII*).*

This determines the stability behavior for linear systems with constant coefficients completely. The following result gives a bound on $X(t) = e^{At}$.

V. Theorem. *If the eigenvalues* λ_i *of the constant (real or complex) matrix* A *satisfy the inequality*

$$\operatorname{Re}\lambda_i < \alpha, \tag{5}$$

then

$$\left|e^{At}\right| \le c e^{\alpha t} \quad \text{for} \quad t \ge 0 \tag{6}$$

for some positive constant c.

The *proof* follows from the fact that by 17.VIII, the differential equation (4) has n linearly independent solutions of the form

$$\mathbf{y}(t) = e^{\lambda t}\mathbf{p}(t), \tag{7}$$

where λ is an eigenvalue of A and $\mathbf{p}(t) = (p_1(t), \ldots, p_n(t))^\top$ is a polynomial of degree $\le n$.

If $\alpha - \operatorname{Re}\lambda = \varepsilon > 0$ here, then certainly $|p_i(t)| \le c_i e^{\varepsilon t}$ holds, and hence

$$|e^{\lambda t} p_i(t)| \le e^{(\varepsilon + \operatorname{Re}\lambda)t} c_i = c_i e^{\alpha t}.$$

If $Y(t)$ denotes the fundamental system consisting of n solutions of the form (7), then each of the n^2 components of Y can be estimated by an expression of the form const $\cdot\, e^{\alpha t}$. The same estimate also holds then for $Y(t)$, and since e^{At} is likewise a principal system and can be represented in the form $e^{At} = Y(t)C$, it holds for e^{At} as well. ∎

A norm generated by a scalar product (cf. 28.I) will be called a *Hilbert norm.* In the following we show that there exists a Hilbert norm in $\mathbb{C}^n$ such that (6) holds with $c = 1$ and, in addition, that a corresponding lower estimate holds; such a result has useful applications. The expression $(\cdot,\cdot)$ denotes the classical scalar product, $|\cdot|$ denotes the Euclidean norm; cf. 28.II(a). The conclusions hold for real A in $\mathbb{R}^n$.

Corollary. *Let the eigenvalues of the matrix A satisfy*

$$\beta < \operatorname{Re}\lambda_i < \alpha. \tag{5'}$$

Then there exists a Hilbert norm $\|\cdot\|$ *in* $\mathbb{C}^n$ *such that the estimates*

$$e^{\beta t}\|\mathbf{c}\| \le \|e^{At}\mathbf{c}\| \le e^{\alpha t}\|\mathbf{c}\| \quad \textit{for} \quad t \ge 0,\ \mathbf{c} \in \mathbb{C}^n \tag{6'}$$

hold. From here one obtains that

$$e^{\beta t} \le \|e^{At}\| \le e^{\alpha t} \quad \textit{for} \quad t \ge 0,$$

where $\|e^{At}\|$ *is the operator norm of* e^{At} *corresponding to* $\|\cdot\|$.

Proof. We assume first that $|\operatorname{Re}\lambda_i| < \delta$ and consider the scalar product

$$\langle \mathbf{c}, \mathbf{d}\rangle := \int_{-\infty}^{\infty} e^{-2\delta|t|}(e^{At}\mathbf{c}, e^{At}\mathbf{d})\, dt \qquad (\mathbf{c}, \mathbf{d} \in \mathbb{C}^n).$$

Let $\varepsilon > 0$ be chosen such that $|\operatorname{Re}\lambda_i| < \delta - \varepsilon$. Then it follows from (6) that $|(e^{At}\mathbf{c}, e^{At}\mathbf{d})| \le |e^{At}|^2|\mathbf{c}||\mathbf{d}| \le \text{const} \cdot e^{2(\delta-\varepsilon)t}$ for $t \ge 0$; therefore, the integral over $[0, \infty)$ is convergent. The integral over $(-\infty, 0]$ can be transformed into an integral over $[0, \infty)$ using the change of variables $t' = -t$, where A is replaced by $-A$. This integral is likewise convergent, since the eigenvalues of $-A$ satisfy the same estimates. Therefore, the scalar product is well-defined, and the properties 28.I are satisfied. Let $\|\mathbf{c}\| := \sqrt{\langle \mathbf{c}, \mathbf{c}\rangle}$. If one chooses $\mathbf{c} = \mathbf{d} = e^{As}\mathbf{a}$, then

$$\|e^{As}\mathbf{a}\|^2 = \int_{-\infty}^{\infty} e^{-2\delta|t|}|e^{A(s+t)}\mathbf{a}|^2 dt = \int_{-\infty}^{\infty} e^{-2\delta|t-s|}|e^{At}\mathbf{a}|^2 dt.$$

For $s \geq 0$, we have $|t| - s \leq |t - s| \leq |t| + s$, and hence

$$e^{-2\delta s}e^{-2\delta|t|} \leq e^{-2\delta|t-s|} \leq e^{2\delta s}e^{-2\delta|t|}.$$

It follows, taking into account the definition of $\|\mathbf{a}\|$, that

$$e^{-2\delta s}\|\mathbf{a}\|^2 \leq \|e^{As}\mathbf{a}\|^2 \leq e^{2\delta s}\|\mathbf{a}\|^2 \quad \text{for} \quad s \geq 0. \tag{$*$}$$

We now consider the general case and set $\gamma = (\alpha + \beta)/2$, $\delta = \alpha - \gamma = \gamma - \beta$. The eigenvalues μ_i of the matrix $A' = A - \gamma I$ can be obtained from the eigenvalues λ_i of the matrix A by setting $\mu_i = \lambda_i - \gamma$; in particular, they are $< \delta$ in magnitude. We define the above scalar product with A' in place of A. Taking square roots and using the relation $e^{A's} = e^{-\gamma s}e^{As}$, we obtain an estimate corresponding to $(*)$:

$$e^{-\delta s}\|\mathbf{a}\| \leq \|e^{A's}\mathbf{a}\| = e^{-\gamma s}\|e^{As}\mathbf{a}\| \leq e^{\delta s}\|\mathbf{a}\| \quad \text{for} \quad s \geq 0.$$

Since $\gamma - \delta = \beta$ and $\gamma + \delta = \alpha$, the inequalities $(6')$ are proved. ∎

We now turn to nonlinear problems. An important tool for this study is

VI. Gronwall's Lemma (1918). *Let the real function $\phi(t)$ be continuous in $J : 0 \leq t \leq a$, and let*

$$\phi(t) \leq \alpha + \beta \int_0^t \phi(\tau)\, d\tau \quad \text{in } J \text{ with} \quad \beta > 0.$$

Then

$$\phi(t) \leq \alpha e^{\beta t} \quad \text{in } J.$$

Proof. Denote the right side of the inequality in the assumption by $\psi(t)$. Then $\psi' = \beta\phi$, and since $\phi \leq \psi$, we have

$$\psi' \leq \beta\psi, \quad \text{or equivalently,} \quad (e^{-\beta t}\psi(t))' \leq 0.$$

Hence $e^{-\beta t}\psi(t)$ is decreasing, which implies

$$e^{-\beta t}\psi(t) \leq \psi(0) = \alpha,$$

and we obtain $\phi(t) \leq \psi(t) \leq \alpha e^{\beta t}$, as was claimed. ∎

The following two theorems constitute classical results of stability theory. They have to do with a differential equation with "linear principal part,"

$$\mathbf{y}' = A\mathbf{y} + \mathbf{g}(t, \mathbf{y}), \tag{8}$$

which means that $\mathbf{g}(t, \mathbf{y})$ is small relative to $\mathbf{y}$ for small $\mathbf{y}$. Under this assumption, which will be made precise in (9), stability properties of the linear equation (4) carry over to the nonlinear equation (8).

VII. Stability Theorem. *Let the function $\mathbf{g}(t, \mathbf{z})$ be defined and continuous for $t \geq 0$, $|z| \leq \alpha$ $(\alpha > 0)$, and let*

$$\lim_{|\mathbf{z}| \to 0} \frac{|\mathbf{g}(t, \mathbf{z})|}{|z|} = 0 \quad \textit{uniformly for} \quad 0 \leq t < \infty; \tag{9}$$

thus, in particular, $\mathbf{g}(t, \mathbf{0}) = \mathbf{0}$. Let A be a constant matrix, and suppose

$$\operatorname{Re} \lambda_i < 0$$

for all eigenvalues λ_i of A.

Then the zero solution $\mathbf{x}(t) \equiv \mathbf{0}$ of the nonlinear equation (8) is asymptotically stable.

Proof. The assumptions together with Theorem III imply that there exist two constants $c > 0$ and $\beta > 0$ such that $\operatorname{Re} \lambda_i < -\beta$ and

$$|\mathrm{e}^{At}| \leq c \cdot \mathrm{e}^{-\beta t} \quad \text{for} \quad t \geq 0.$$

Moreover, by (9), there exists a δ, $0 < \delta < \alpha$, such that

$$|\mathbf{g}(t, \mathbf{z})| \leq \frac{\beta}{2c} |\mathbf{z}| \quad \text{for} \quad |\mathbf{z}| \leq \delta,\ t \geq 0. \tag{10}$$

The theorem is proved if we can show that

$$|\mathbf{y}(0)| \leq \varepsilon < \frac{\delta}{2c} \ \text{ implies } \ |\mathbf{y}(t)| \leq c\varepsilon \mathrm{e}^{-\beta t/2} \ \text{ for } \ t \geq 0. \tag{$*$}$$

We know from 18.VI that every solution of the nonhomogeneous equation

$$\mathbf{y}' = A\mathbf{y} + \mathbf{b}(t)$$

can be represented in the form

$$\mathbf{y}(t) = \mathrm{e}^{At}\mathbf{y}_0 + \int_0^t \mathrm{e}^{A(t-s)}\mathbf{b}(s)\, ds, \quad \text{where} \quad \mathbf{y}_0 := \mathbf{y}(0).$$

Now, if $\mathbf{y}(t)$ is a solution of (8), then accordingly, it satisfies the integral equation

$$\mathbf{y}(t) = \mathrm{e}^{At}\mathbf{y}_0 + \int_0^t \mathrm{e}^{A(t-s)}\mathbf{g}(s, \mathbf{y}(s))\, ds.$$

Using (10), the inequality

$$|\mathbf{y}(t)| \le |\mathbf{y}_0| c e^{-\beta t} + \int_0^t c e^{-\beta(t-s)} \frac{\beta}{2c} |\mathbf{y}(s)|\, ds \tag{11}$$

follows, at least as long as (10) can be applied, i.e., as long as $|\mathbf{y}| \le \delta$. Now let $\mathbf{y}(t)$ be a solution of (8) with $|\mathbf{y}_0| < \varepsilon$ and $\phi(t) = |\mathbf{y}(t)| e^{\beta t}$. From (11) it follows (as long as $|\mathbf{y}| \le \delta$) that

$$\phi(t) \le c\varepsilon + \frac{\beta}{2} \int_0^t \phi(s)\, ds,$$

and therefore by Gronwall's lemma,

$$\phi(t) \le c\varepsilon e^{\beta t/2}, \quad \text{or} \quad |\mathbf{y}(t)| \le c\varepsilon e^{-\beta t/2} < \frac{1}{2}\delta. \tag{12}$$

This inequality implies that $|\mathbf{y}(t)|$ cannot take on the value δ for any positive t and hence that the inequality (12), and consequently $(*)$ holds for all $t \ge 0$. Note that $\mathbf{y}(t)$ can be extended to the boundary of the domain of $\mathbf{g}$, hence because of (12) to the whole interval $0 \le t < \infty$. ∎

VIII. Instability Theorem. *Assume that* $\mathbf{g}(t, \mathbf{z})$ *satisfies the assumptions of Theorem* VII. *Further, let* A *be a constant matrix and suppose*

$$\operatorname{Re} \lambda > 0$$

for at least one eigenvalue λ *of* A. *Then the solution* $\mathbf{x}(t) \equiv 0$ *of the nonlinear differential equation* (8) *is unstable.*

Proof. We first transform the differential equation (8), using a linear transformation, into a form that is better suited for our purposes. Let $\lambda_1, \ldots, \lambda_n$ be the zeros of the characteristic polynomial of A (counting multiplicities). Let A be transformed into the Jordan normal form B by the matrix C:

$$B = C^{-1} A C = (b_{ij})$$

with

$$b_{ii} = \lambda_i, \; b_{i,i+1} = 0 \text{ or } 1, \quad b_{ij} = 0 \quad \text{otherwise}.$$

Further, let H be the diagonal matrix

$$H = \operatorname{diag}(\eta, \eta^2, \ldots, \eta^n) \quad (\eta > 0).$$

It is easy to check that $H^{-1} = \operatorname{diag}(\eta^{-1}, \eta^{-2}, \ldots, \eta^{-n})$ and that

$$D = H^{-1} B H \Leftrightarrow d_{ij} = b_{ij} \eta^{j-i},$$

i.e.,

$$d_{ii} = \lambda_i, \; d_{i,i+1} = 0 \text{ or } \eta, \quad d_{ij} = 0 \quad \text{otherwise}. \tag{13}$$

If we now set $\mathbf{y}(t) = CH\mathbf{z}(t)$, then the differential equation transforms into

$$\mathbf{z}' = H^{-1}C^{-1}\mathbf{y}' = H^{-1}C^{-1}\left\{ACH\mathbf{z} + \mathbf{g}(t, CH\mathbf{z})\right\},$$

or

$$\mathbf{z}' = D\mathbf{z} + \mathbf{f}(t, \mathbf{z}) \tag{14}$$

with

$$\mathbf{f}(t, \mathbf{z}) = H^{-1}C^{-1}\mathbf{g}(t, CH\mathbf{z}) \quad \text{and} \quad D = H^{-1}C^{-1}ACH.$$

If $\mathbf{g}$ satisfies assumption (9), then so does $\mathbf{f}$, since from $|\mathbf{g}| \le \varepsilon|\mathbf{z}|$ for $|\mathbf{z}| \le \delta$ it follows that

$$|\mathbf{f}(t, \mathbf{z})| \le |H^{-1}C^{-1}| \cdot |CH|\varepsilon|\mathbf{z}| \quad \text{for} \quad |\mathbf{z}| \le \delta/|CH|.$$

Instead of (14) one can also write

$$z_i' = \lambda_i z_i\{+\eta z_{i+1}\} + f_i(t, \mathbf{z}) \qquad (i = 1, \ldots, n). \tag{14'}$$

The term in braces appears only if the index i corresponds to a Jordan block with more than one row and does not correspond to the last row in this Jordan block.

We denote by j or k those indices for which

$$\operatorname{Re}\lambda_j > 0 \quad \text{or} \quad \operatorname{Re}\lambda_k \le 0, \quad \text{respectively},$$

and by ϕ, ψ the real-valued functions

$$\phi(t) = \sum_j |z_j(t)|^2, \qquad \psi(t) = \sum_k |z_k(t)|^2,$$

where $\mathbf{z}(t)$ is a solution of (14′). Now let $\eta > 0$ be chosen so small that

$$0 < 6\eta < \operatorname{Re}\lambda_j \quad \text{for all } j,$$

and choose $\delta > 0$ so small that

$$|\mathbf{f}(t, \mathbf{z})|_e < \eta|\mathbf{z}|_e \quad \text{for} \quad |\mathbf{z}|_e \le \delta.$$

If $\mathbf{z}(t)$ is a solution of (14′) with

$$|\mathbf{z}(0|_e < \delta, \quad \psi(0) < \phi(0), \tag{15}$$

then as long as $|\mathbf{z}(t)|_e \le \delta$ and $\psi(t) \le \phi(t)$, we have

$$\begin{aligned} \phi' &= 2\textstyle\sum_j \operatorname{Re} z_j'\bar{z}_j \\ &= 2\textstyle\sum_j(\operatorname{Re}\lambda_j z_j\bar{z}_j\{+\eta\operatorname{Re} z_{j+1}\bar{z}_j\} + \operatorname{Re}\bar{z}_j f_j(t, \mathbf{z})), \end{aligned} \tag{16}$$

and further, by the Schwarz inequality ($j+1$ is an index of type j),

$$\left|\textstyle\sum \operatorname{Re} z_{j+1}\bar{z}_j\right| \le \textstyle\sum|z_j z_{j+1}| \le \sqrt{\textstyle\sum|z_j|^2\textstyle\sum|z_j|^2} = \phi$$

and

$$\left|\sum \operatorname{Re} \bar{z}_j f_j\right| \leq \sqrt{\sum |z_j|^2 \sum |f_j|^2} \leq \sqrt{\phi}|\mathbf{f}|_e,$$

as well as $\operatorname{Re} \sum \lambda_j z_j \bar{z}_j > 6\eta\phi$ and

$$|\mathbf{f}|_e \leq \eta |\mathbf{z}|_e = \eta\sqrt{\phi + \psi} \leq 2\eta\sqrt{\phi}.$$

Hence

$$\frac{1}{2}\phi' > 6\eta\phi - \eta\phi - 2\eta\phi = 3\eta\phi.$$

An equation analogous to (16) holds for $\psi(t)$ (one simply replaces j by k). Thus, because $\operatorname{Re} \lambda_k \leq 0$, it follows, using the same estimates, that

$$\frac{1}{2}\psi' \leq \eta\psi + 2\eta\phi.$$

Therefore, as long as $\psi(t) \leq \phi(t)$, we have

$$\frac{1}{2}(\phi' - \psi') > 3\eta\phi - (\eta\psi + 2\eta\phi) = \eta(\phi - \psi) \geq 0,$$

i.e., the difference $\phi - \psi$ is increasing as long as it is positive. This shows that as long as $|\mathbf{z}(t)| \leq \delta$, the inequalities $\psi(t) < \phi(t)$ and $\phi' > 6\eta\phi$ are satisfied, and hence $\phi(t) \geq \phi(0)e^{6\eta t}$ (Lemma 9.I); i.e., for every solution $\mathbf{z}(t)$ satisfying (15) there exists a t_0 with $|\mathbf{z}(t_0)| = \delta$. But this signifies that the solution $\mathbf{x}(t) \equiv \mathbf{0}$ is not stable. ∎

IX. Autonomous Systems. Linearization. Autonomous systems are systems of the form

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}). \tag{17}$$

The right-hand side of such equations does not depend explicitly on t; consequently, if $\mathbf{y}(t)$ is a solution, then so is $\mathbf{y}(t+t_0)$. Other properties of autonomous systems were already discussed in 10.XI.

Let us assume that $\mathbf{f} \in C^1(D)$, where $D \subset \mathbb{R}^n$ is neighborhood of $\mathbf{0}$, and that $\mathbf{0}$ is a critical point of $\mathbf{f}$, i.e., $\mathbf{f}(\mathbf{0}) = \mathbf{0}$. The equation $\mathbf{y}' = A\mathbf{y}$, where A is now the Jacobian $\mathbf{f}'(\mathbf{0})$, is called the *linearized equation* at the point $\mathbf{0}$, and the transition from the nonlinear equation (17) to the linear equation $\mathbf{y}' = \mathbf{f}'(\mathbf{0})\mathbf{y}$ is denoted as *linearization*. If equation (17) is written in the form

$$\mathbf{y}' = A\mathbf{y} + \mathbf{g}(\mathbf{y}), \tag{18}$$

then

$$\mathbf{g}(\mathbf{y}) = \mathbf{f}(\mathbf{y}) - \mathbf{f}'(\mathbf{0})\mathbf{y},$$

and hence

$$\lim_{\mathbf{y}\to\mathbf{0}} \frac{\mathbf{g}(\mathbf{y})}{|\mathbf{y}|} = 0$$

by the definition of differentiability. This shows that the function $\mathbf{g}(\mathbf{y})$ satisfies the main assumption (9) of the two preceding theorems.

By the stability theorem VII, *the "equilibrium state"* $\mathbf{x} \equiv \mathbf{0}$ *of the nonlinear equation* (17) *is asymptotically stable if the same is true of the linearized equation* (4). *By Theorem* VIII, *it is certainly unstable if* $\operatorname{Re}\lambda > 0$ *holds for some eigenvalue of* A.

We have already seen several types of *unstable* linear systems in the case $n = 2$ with completely different phase portraits, for instance, the saddle point and unstable nodes and vortex points. This raises the question whether the structural similarity between the linear system (4) and the "perturbed" equation (18) (with $\mathbf{g}(\mathbf{y}) = o(|\mathbf{y}|)$) reaches still deeper and also includes the phase portraits. The answer is sometimes, but not always, positive; it requires a new notion.

The origin is called a *hyperbolic critical point* of $\mathbf{f}$ if $\mathbf{f}(\mathbf{0}) = \mathbf{0}$ and the Jacobian matrix $A = f'(0)$ has only eigenvalues with $\operatorname{Re}\lambda \neq 0$. For such points D. Grobman (1959) and Ph. Hartman (1963) proved the following

Linearization Theorem. (Grobman–Hartman). *Let D be a neighborhood of the origin and $\mathbf{f} \in C^1(D)$. If the point $\mathbf{0}$ is a hyperbolic critical point of $\mathbf{f}$, then there exist neighborhoods U, V of the origin and a homeomorphism $\mathbf{h} : U \to V$ (a bijection that is continuous in both directions) that transforms the trajectories of the linear equation* (4) *(as long as they belong to U) into trajectories of the nonlinear equation* (17), *preserving the sense of direction.*

(a) The above conclusions carry immediately over to the case where another point $\mathbf{a}$, instead of $\mathbf{0}$, is a critical point of $\mathbf{f}$. This is because the difference $\mathbf{z}(t) = \mathbf{y}(t) - \mathbf{a}$, which is the concern of stability questions, satisfies the equation $\mathbf{z}' = \mathbf{h}(\mathbf{z})$ with $\mathbf{h}(\mathbf{z}) = \mathbf{f}(\mathbf{a} + \mathbf{z})$ whenever $\mathbf{y}$ is a solution of (17). Here we have $\mathbf{h}(\mathbf{0}) = \mathbf{f}(\mathbf{a}) = \mathbf{0}$ and $A = \mathbf{h}'(\mathbf{0}) = \mathbf{f}(\mathbf{a})$. The critical point $\mathbf{a}$ is called *hyperbolic* whenever $\operatorname{Re}\lambda \neq 0$ for $\lambda \in \sigma(A)$. In the linearization theorem, which remains valid, V is now a neighborhood of $\mathbf{a}$.

Examples. 1. The real linear systems for $n = 2$ with $\det A \neq 0$ discussed in 17.X have the origin as the only critical point. It is hyperbolic in all cases with one exception, the center (in 17.X.(d), it is denoted by $K(0, \omega)$).

2. The equation of the mathematical pendulum

$$u'' + \sin u = 0 \Longleftrightarrow \begin{pmatrix} x \\ y \end{pmatrix}' = \begin{pmatrix} y \\ -\sin x \end{pmatrix}$$

has critical points $(0, 0)$ and $(\pi, 0)$. Corresponding linearizations are

$$A = \mathbf{f}'(0,0) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad A = \mathbf{f}'(\pi, 0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

For the first of these, the linear part is the equation for the harmonic oscillator $u'' + u = 0$. The trajectories are circles around the origin, and the phase portrait of the mathematical pendulum shows closed, approximately circular Jordan curves near $(0,0)$. The linearization theorem, however, does not lead to any conclusion in this case ($\operatorname{Re}\lambda = 0$), and for good reason. The differential equation

$$u'' + u^2 u' + \sin u = 0$$

has the same linearization at the origin. Nevertheless, this equation is asymptotically stable at the origin (this follows from 30.X.(e)), while the harmonic oscillator has a center there.

By contrast, in the second case $\det(A - \lambda I) = \lambda^2 - 1$, whence $\lambda = \pm 1$, and one has a saddle point. By the linearization theorem the phase portrait of the mathematical pendulum likewise has a saddle point structure in a neighborhood of the point $(\pi, 0)$. Compare the corresponding pictures in 11.X.(d) and 17.X.(c).

3. Let $n = 1$ and consider the equation

$$y' = \alpha y + \beta y^3 \qquad (\alpha, \beta \in \mathbb{R}).$$

The linearized equation is $y' = \alpha y$. The following table describes the stability behavior of the solution $y \equiv 0$:

linearized equation		nonlinear equation
$\alpha < 0$	asymptotically stable	asymptotically stable
$\alpha > 0$	unstable	unstable
$\alpha = 0$	stable	$\begin{cases} \text{asymptotically stable} & \text{if } \beta < 0 \\ \text{stable} & \text{if } \beta = 0 \\ \text{unstable} & \text{if } \beta > 0 \end{cases}$

The conclusions for $\alpha \neq 0$ follow from VII and VIII; the proof for the case $\alpha = 0$ is suggested as an exercise. The point $y = 0$ is hyperbolic for $\alpha \neq 0$ only; for $\alpha = 0$ the nonlinear equation changes its behavior with β.

Linearization is an excellent tool for the study of nonlinear autonomous systems in a neighborhood of its critical points. This viewpoint gives the classification of plane linear systems a deeper significance. However, the connection disappears if the linear system has eigenvalues with vanishing real part.

Proofs of the linearization theorem are found in the books by Amann (1983) and Hartman (1964); they are not simple. The books by Jordan and Smith (1988) and Drazin (1992) are well suited for obtaining a deeper understanding of the global behavior of nonlinear systems. More advanced are the treatises of Hale–Koçak (1991) and Wiggins (1988).

X. The Generalized Lemma of Gronwall. *Let the real-valued function $\phi(t)$ be continuous in $J = [0, a]$, and let*

$$\phi(t) \le \alpha + \int_0^t h(s)\phi(s)\,ds \quad \text{in} \quad J,$$

where $\alpha \in \mathbb{R}$ and $h(t)$ is nonnegative and continuous (sufficient: Lebesgue integrable) in J. Then

$$\phi(t) \le \alpha e^{H(t)} \quad \text{with} \quad H(t) = \int_0^t h(s)\,ds.$$

The *proof* from VI carries over (Exercise!).

XI. Exercise. (a) In the (real or complex) system of differential equations

$$\mathbf{y}' = A\mathbf{y} + \mathbf{g}(t, \mathbf{y})$$

let A be a constant matrix and $\operatorname{Re}\lambda < \alpha$ for every eigenvalue λ of A. Further, let $\mathbf{g}(t, \mathbf{y})$ be continuous for $t \ge 0$, $\mathbf{y} \in \mathbb{R}^n$ or $\mathbb{C}^n$, and

$$|\mathbf{g}(t, \mathbf{y})| \le h(t)|\mathbf{y}|,$$

where $h(t)$ is a continuous (sufficient: locally integrable) function for $t \ge 0$. Show that every solution $\mathbf{y}(t)$ satisfies an estimate

$$|\mathbf{y}(t)| \le K|\mathbf{y}(0)|e^{\alpha t + KH(t)} \quad \text{with} \quad H(t) = \int_0^t h(s)\,ds$$

for some constant $K > 0$ that is independent of $\mathbf{y}$.

Hint. Derive an integral equation for $\phi(t) = e^{-\alpha t}|\mathbf{y}(t)|$ and use the generalized lemma of Gronwall.

From (a) conclude the following:

(b) If $h(t)$ is integrable over $0 \le t < \infty$ and if all eigenvalues of A have negative real part, then the solution $\mathbf{y} \equiv \mathbf{0}$ is asymptotically stable and all solutions tend to zero as $t \to \infty$.

(c) In the linear system

$$\mathbf{y}' = (A + B(t))\mathbf{y},$$

let $B(t)$ be a continuous matrix for $t \ge 0$, and let

$$\int_0^\infty |B(t)|\,dt < \infty.$$

If all eigenvalues of A have negative real part, then the solution $\mathbf{y} \equiv \mathbf{0}$ is asymptotically stable.

XII. Exercise. Let $n = 3$, $\mathbf{y} = (x, y, z)$, and

$$\begin{aligned} x' &= -x - y + z + r_1(x,y,z)x, \\ y' &= x - 2y + 2z + r_2(x,y,z)y, \\ z' &= x + 2y + z + r_3(x,y,z)z, \end{aligned}$$

where $r_i(x,y,z)$ is continuous and $r_i(0,0,0) = 0$ $(i = 1,2,3)$. Show that the zero solution is unstable.

§ 30. The Method of Lyapunov

The Russian mathematician and engineer A.M. Lyapunov (1857–1918) introduced in his dissertation of 1892 two methods for dealing with stability questions. While the first method is of a special nature, his *second*, or *direct, method* has developed into an extraordinarily useful tool. The method is based on a real-valued *Lyapunov function* V, which can be viewed as a generalized distance from the origin.

We consider *real autonomous systems*

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}), \tag{1}$$

where $\mathbf{f}$ is continuous in the open set $D \subset \mathbb{R}^n$, $\mathbf{0} \in D$, and $\mathbf{f}(\mathbf{0}) = \mathbf{0}$. The zero solution $\mathbf{x}(t) \equiv \mathbf{0}$ of (1) is also called the *rest state* or *equilibrium state.* Our theorems deal with this case. The extension to an equilibrium state $\mathbf{x}(t) \equiv \mathbf{a}$ in the case where $\mathbf{f}(\mathbf{a}) = \mathbf{0}$ is elementary; cf. 29.IX.(a).

Notation. The functions, vectors, and matrices that appear have (unless otherwise noted) real-valued components. The expressions $(\mathbf{x}, \mathbf{y})$, $|\mathbf{x}|$, and B_r denote respectively the scalar product, the Euclidean norm and the open ball $|\mathbf{x}| < r$ in $\mathbb{R}^n$. In many results in this section it is assumed that $\mathbf{f}$ is locally Lipschitz continuous in D. When this is the case, the solution $\mathbf{y}(t)$ of (1) with initial condition $\mathbf{y}(0) = \boldsymbol{\eta}$ is uniquely determined and will be denoted by $\mathbf{y}(t; \boldsymbol{\eta})$.

Expositions of the Lyapunov method are found in the books cited at the end of 29.IX and in the monographs by Cesàri (1971) and Hahn (1967).

I. Lyapunov Functions. We first introduce an additional concept of stability. The equilibrium state is said to be *exponentially stable* if there exist positive constants β, γ, c such that for every solution $\mathbf{y}(t)$ of (1)

$$|\mathbf{y}(0)| < \beta \quad \text{implies} \quad |\mathbf{y}(t)| < ce^{-\gamma t} \quad \text{for} \quad t > 0;$$

in particular, it is required that these solutions exist in $[0, \infty)$.

(a) *If* $\mathbf{f}$ *is locally Lipschitz continuous, then exponential stability implies asymptotic stability.*

Proof. For every $\varepsilon > 0$ there exists an $a \geq 0$ such that $ce^{-a\gamma} < \varepsilon$. Thus if $|\boldsymbol{\eta}| < \beta$, then $|\mathbf{y}(t;\boldsymbol{\eta})| < \varepsilon e^{-\gamma(t-a)}$ in $[a,\infty)$. By Theorem 13.II, there exists a positive $\delta < \beta$ such that $|\boldsymbol{\eta}| < \delta$ implies that $|\mathbf{y}(t;\boldsymbol{\eta})| < \varepsilon$ in $[0,a]$. Therefore, the last inequality holds in $[0,\infty)$ if $|\boldsymbol{\eta}| < \delta$, i.e., the equilibrium state is stable and, in fact, is asymptotically stable. ∎

Given a real-valued function $V \in C^1(D)$, we define

$$\dot{V} := (\operatorname{grad} V(\mathbf{x}), \mathbf{f}(\mathbf{x})) = f_1(\mathbf{x}) \cdot V_{x_1}(\mathbf{x}) + \cdots + f_n(\mathbf{x}) \cdot V_{x_n}(\mathbf{x}). \tag{2}$$

It is easy to recognize that $\dot{V}$ is the directional derivative of V in the (not normalized) direction of $\mathbf{f}$:

$$\dot{V}(\mathbf{x}) = \lim_{t\to 0} \frac{1}{t}\left[V(\mathbf{x}+t\mathbf{f}(\mathbf{x})) - V(\mathbf{x})\right]. \tag{2'}$$

Because of the following property, $\dot{V}$ is also called the *derivative of V along trajectories.*

(b) *If $\mathbf{y}(t)$ is a solution to* (1), *then, by the chain rule and* (1),

$$\frac{d}{dt}V(\mathbf{y}(t)) = \dot{V}(\mathbf{y}(t)).$$

This formula can be used to obtain information about the behavior of V along a trajectory *without prior knowledge of the solution.* In the "direct method", this idea is exploited. A *Lyapunov function* for (1) is a function $V \in C^1(D)$ that satisfies the relations

$$V(\mathbf{0}) = 0, \quad V(\mathbf{x}) > 0 \quad \text{for} \quad \mathbf{x} \neq \mathbf{0} \quad \text{and} \quad \dot{V}(\mathbf{x}) \leq 0 \quad \text{in} \quad D.$$

II. Stability Theorem (Lyapunov). *Let $\mathbf{f} \in C(D)$ with $\mathbf{f}(\mathbf{0}) = \mathbf{0}$ and let there exist a Lyapunov function V for $\mathbf{f}$. Then*

(a) *$\dot{V} \leq 0$ in $D \Longrightarrow$ the zero solution of* (1) *is stable.*

(b) *$\dot{V} < 0$ in $D \setminus \{\mathbf{0}\} \Longrightarrow$ the zero solution of* (1) *is asymptotically stable.*

(c) *$\dot{V} \leq -\alpha V$ and $V(\mathbf{x}) \geq b|\mathbf{x}|^\beta$ in D $(\alpha,\beta,b > 0)$ $\Longrightarrow$ the zero solution is exponentially stable.*

Proof. (a) Let $\varepsilon > 0$ be chosen so small that the closed ball $\bar{B}_\varepsilon$ lies in D. We choose a positive γ such that $V(\mathbf{x}) > \gamma$ holds for $|\mathbf{x}| = \varepsilon$, and then choose a δ with $0 < \delta < \varepsilon$ such that $V(\mathbf{x}) < \gamma$ for $|\mathbf{x}| < \delta$. If $\mathbf{y}$ is a solution of (1) with $|\mathbf{y}(0)| < \delta$, then by I.(b), the derivative of the function $\phi(t) = V(\mathbf{y}(t))$ satisfies $\phi'(t) \leq 0$, and hence we have $\phi(t) \leq \phi(0) < \gamma$. Since $V(\mathbf{x})$ only takes on values $\geq \gamma$ on the sphere $|\mathbf{x}| = \varepsilon$, it follows that $|\mathbf{y}(t)|$ remains $< \varepsilon$ for $t > 0$. Both the existence of the solution in the whole interval $J = [0,\infty)$ and the estimate $|\mathbf{y}(t)| < \varepsilon$ in J follow from here.

(b) If $\mathbf{y}(t)$ is a solution as defined in (a) and $\phi(t) = V(\mathbf{y}(t))$, then $\lim_{t\to\infty} \phi(t) = \beta < \gamma$. We first show that $\beta = 0$. Let us assume that this is not the case. Then the set $M = \{\mathbf{x} \in \bar{B}_\varepsilon : \beta \leq V(\mathbf{x}) \leq \gamma\}$ is a compact subset of $\bar{B}_\varepsilon \setminus \{\mathbf{0}\}$ and

$\max\{\dot{V}(\mathbf{x}) : \mathbf{x} \in M\} = -\alpha < 0$. Since the solution $\mathbf{y}$ stays in M, we would have $\phi'(t) \leq -\alpha$, which leads to a contradiction. Thus $\lim \phi(t) = 0$.

The limit relation $\mathbf{y}(t) \to \mathbf{0}$ $(t \to \infty)$ follows from here. For a positive $\varepsilon' < \varepsilon$, the function V has a positive minimum δ on the set $\varepsilon' \leq |\mathbf{x}| \leq \epsilon$. Therefore, $|\mathbf{y}(t)| < \varepsilon'$ as soon as $\phi(t) < \delta$, i.e., for all large t.

(c) The hypotheses imply that $b|\mathbf{y}(t)|^\beta \leq V(\mathbf{y}(t)) = \phi(t)$ and $\phi' \leq -\alpha\phi$, and hence $\phi(t) \leq \phi(0)e^{-\alpha t}$. It follows that $|\mathbf{y}(t)| \leq ce^{-\gamma t}$ with $\gamma = \alpha/\beta > 0$. ∎

III. Instability Theorem (Lyapunov). *Suppose that $V \in C^1(D)$, $V(\mathbf{0}) = 0$, and $V(\mathbf{x}_k) > 0$ for some sequence $(\mathbf{x}_k)$ in $D \setminus \{\mathbf{0}\}$ with $\mathbf{x}_k \to \mathbf{0}$. If $\dot{V} > 0$ for $\mathbf{x} \neq \mathbf{0}$ or $\dot{V} \geq \lambda V$ in D with $\lambda > 0$, then the zero solution is unstable. In particular, it is unstable if $V(\mathbf{x}) > 0$ and $\dot{V}(\mathbf{x}) > 0$ for $\mathbf{x} \neq \mathbf{0}$.*

Proof. Let $\mathbf{y}$ be a solution of (1) with $\mathbf{y}(0) = \mathbf{x}_k \neq \mathbf{0}$; it follows that $\phi(0) = \alpha > 0$, where once again $\phi(t) = V(\mathbf{y}(t))$. We consider the first case and choose $\varepsilon > 0$ such that $V < \alpha$ in $\bar{B}_\varepsilon$. Since $\phi' \geq 0$, and hence $\alpha = \phi(0) \leq \phi(t)$, we have $|\mathbf{y}(t)| > \varepsilon$. Now let $\bar{B}_r$ be a closed ball contained in D $(r > \varepsilon)$. If $\varepsilon \leq |\mathbf{x}| \leq r$, then $\dot{V}(\mathbf{x}) \geq \beta > 0$, and hence $\phi' \geq \beta$ and $\phi(t) \geq \alpha + \beta t$, as long as $\mathbf{y}(t) \in B_r$. Since V is bounded in B_r, the solution $\mathbf{y}(t)$ must leave the ball B_r in finite time.

In the second case we have $\phi'(t) \geq \lambda\phi(t)$, whence it follows that $\phi(t) \geq \alpha e^{\lambda t}$. Here, too, $\mathbf{y}(t) > r$ for large t. Therefore, because $\mathbf{x}_k \to \mathbf{0}$, there exist solutions with arbitrarily small initial values that leave the ball B_r. ∎

IV. Examples. There is no general recipe for constructing Lyapunov functions. In specific cases one may rely on experience and examples; some imagination is also helpful. For many problems, the scalar product $V(\mathbf{x}) = (\mathbf{x}, \mathbf{x}) = |\mathbf{x}|^2$ works. More generally, we consider an arbitrary scalar product $\langle \mathbf{x}, \mathbf{y} \rangle$ in $\mathbb{R}^n$ and compute the derivative $\dot{V}$ associated with the function $V(\mathbf{x}) = \langle \mathbf{x}, \mathbf{x} \rangle$ using (2′):

$$V(\mathbf{x} + t\mathbf{f}(\mathbf{x})) - V(\mathbf{x}) = 2t\langle \mathbf{x}, \mathbf{f}(\mathbf{x}) \rangle + t^2\langle \mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}) \rangle.$$

After dividing by t, one obtains, as $t \to 0$,

(a) $V(\mathbf{x}) = \langle \mathbf{x}, \mathbf{x} \rangle$ satisfies $\dot{V}(\mathbf{x}) = 2\langle \mathbf{x}, \mathbf{f}(\mathbf{x}) \rangle$ and $V(\mathbf{x}) > 0$ for $\mathbf{x} \neq 0$.

(b) In the differential equation

$$\mathbf{y}' = A\mathbf{y} + \psi(\mathbf{y})B\mathbf{y} + \mathbf{g}(\mathbf{y}),$$

assume that $\psi : D \to \mathbb{R}$ and $\mathbf{g} : D \to \mathbb{R}^n$ with $\mathbf{g}(\mathbf{0}) = \mathbf{0}$ are continuous and $B = -B^T$ is a skew-symmetric matrix. Then it follows from the result in (a) that the derivative of $V(\mathbf{x}) = (\mathbf{x}, \mathbf{x}) = |\mathbf{x}|^2$ is given by

$$\dot{V}(\mathbf{x}) = 2(\mathbf{x}, A\mathbf{x}) + 2(\mathbf{x}, \mathbf{g}(\mathbf{x})).$$

By the skew-symmetry of B, we have $(\mathbf{x}, B\mathbf{x}) = 0$; therefore, the term $\psi(\mathbf{y})B\mathbf{y}$ in the differential equation has absolutely no effect on $\dot{V}$. We consider three cases.

(i) If $(\mathbf{x}, A\mathbf{x}) \leq 0$ and $(\mathbf{x}, \mathbf{g}(\mathbf{x})) \leq 0$ in D, then the rest state is stable.

Now suppose $\mathbf{g}(\mathbf{x}) = o(|\mathbf{x}|)$ as $\mathbf{x} \to \mathbf{0}$. Then the rest state is

(ii) exponentially stable if $(\mathbf{x}, A\mathbf{x}) \le -\alpha|\mathbf{x}|^2$ with $\alpha > 0$,

(iii) unstable if $(\mathbf{x}, A\mathbf{x}) \ge \alpha|\mathbf{x}|^2$ with $\alpha > 0$ holds.

Proposition (i) follows from Theorem II.(a). To prove (ii) and (iii), let $r > 0$ be determined such that $B_r \subset D$ and $|\mathbf{g}(\mathbf{x})| < \frac{1}{2}\alpha|\mathbf{x}|$ in B_r. It follows that $|(\mathbf{x}, \mathbf{g}(\mathbf{x}))| < \frac{1}{2}\alpha|\mathbf{x}|^2$, and hence $\dot{V} \le -\alpha$ or $\dot{V} \ge \alpha V$ in B_r. Then (ii) follows from Theorem II.(c) and (iii) from Theorem III, applied on B_r. ∎

(c) *Linear Systems.* Consider the linear system $\mathbf{y}' = A\mathbf{y}$ and suppose $\operatorname{Re}\lambda < 0$ for all $\lambda \in \sigma(A)$. We use the scalar product

$$\langle \mathbf{x}, \mathbf{y} \rangle = \int_0^\infty (\mathrm{e}^{At}\mathbf{x}, \mathrm{e}^{At}\mathbf{y})\, dt$$

(the convergence of the integral is proved as in 29.V). If $V(\mathbf{x}) = \langle \mathbf{x}, \mathbf{x} \rangle$ and $\mathbf{y}(t) = \mathrm{e}^{At}\mathbf{x}$, $\mathbf{y}'(t) = A\mathrm{e}^{At}\mathbf{x}$, then

$$\dot{V}(\mathbf{x}) = 2\langle \mathbf{x}, A\mathbf{x} \rangle = \int_0^\infty 2(\mathbf{y}(t), \mathbf{y}'(dt))\, dt = |\mathbf{y}(t)|^2\Big|_0^\infty = -|\mathbf{x}|^2.$$

By Theorem II.(b), the zero solution is asymptotically stable, something we have known all along (17.XI). However, the approach used here gives additional information.

(d) *"Lightning proof" of the Stability Theorem* 29.VII *in the Autonomous Case.* If $\operatorname{Re}\lambda < 0$ holds for $\lambda \in \sigma(A)$ and if $\mathbf{g}(\mathbf{x}) = o(|\mathbf{x}|)$ as $\mathbf{x} \to \mathbf{0}$, then the zero solution of the equation

$$\mathbf{y}' = A\mathbf{y} + \mathbf{g}(\mathbf{y})$$

is exponentially stable.

Proof. The function V introduced in (c) satisfies (with $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$)

$$\dot{V}(\mathbf{x}) \le -|\mathbf{x}|^2 + 2\langle \mathbf{x}, \mathbf{g}(\mathbf{x}) \rangle \le -|\mathbf{x}|^2 + 2\|\mathbf{x}\|\,\|\mathbf{g}(\mathbf{x})\|.$$

There exists $c > 0$ such that $\|\mathbf{x}\| \le c|\mathbf{x}|$; cf. Lemma 10.III. Let $r > 0$ be such that $B_r \subset D$ and $|\mathbf{g}(\mathbf{x})| \le (1/(4c^2))|\mathbf{x}|$ in B_r. Then

$$\dot{V}(\mathbf{x}) \le -|\mathbf{x}|^2 + 2c^2|\mathbf{x}||\mathbf{g}(\mathbf{x})| \le -\frac{1}{2}|\mathbf{x}|^2 \le -\frac{1}{2c^2}V(\mathbf{x}) \text{ in } B_r.$$

The conclusion now follows from Theorem II.(c). ∎

(e) *Nonlinear Oscillations without Friction.* For the equation

$$u'' + h(u) = 0 \Leftrightarrow x' = y,\ y' = -h(x)$$

with $xh(x) > 0$ for $x \ne 0$, studied in 11.X, an obvious choice for a Lyapunov function is the energy function

$$E(x, y) = \frac{1}{2}y^2 + H(x) \quad \text{with} \quad H(x) = \int_0^x h(s)\, ds.$$

Here $E(x, y) > 0$ for $(x, y) \neq (0, 0)$ and $\dot{E}(x, y) \equiv 0$. Therefore, the zero solution is stable.

(f) *Nonlinear Oscillations with Friction.* We consider an equation with a linear friction term $\varepsilon u'$ $(\varepsilon > 0)$

$$u'' + \varepsilon u' + h(u) = 0 \Leftrightarrow x' = y, \ y' = -h(x) - \varepsilon y.$$

As a Lyapunov function we take again the energy function $E(x, y)$ from (e); it has now the derivative

$$\dot{E} = -\varepsilon y^2.$$

Thus the energy decreases, as might be expected. By Theorem II.(a), the rest state is stable. On physical grounds, one would guess that it is, in fact, asymptotically stable. This, however, does not follow from Theorem II.(b), since the inequality $\dot{V} < 0$ is violated when $y = 0$. In the next section we will derive a more general stability theorem that implies, among other things, the asymptotic stability in this example.

V. Limit Points and Limit Sets. Invariant Sets. In the autonomous differential equation

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}) \tag{1}$$

let $\mathbf{f}$ be locally Lipschitz continuous on the open set $D \subset \mathbb{R}^n$. The solution $\mathbf{y}(t)$ with $\mathbf{y}(0) = \boldsymbol{\eta} \in D$ will be denoted by $\mathbf{y}(t; \boldsymbol{\eta})$. This solution exists in a maximal interval $J = (t^-, t^+)$ with $-\infty \leq t^- < 0 < t^+ \leq \infty$ and generates an orbit $\gamma = \mathbf{y}(J)$. The sets $\gamma^+ = \mathbf{y}([0, t^+))$ and $\gamma^- = \mathbf{y}((t^-, 0])$ are called the *positive semiorbit* and *negative semiorbit*, respectively. A point $\mathbf{a} \in \mathbb{R}^n$ is called a *positive limit point* or *ω-limit point* if $t^+ = \infty$ and if there exists a sequence (t_k) tending to ∞ such that $\lim \mathbf{y}(t_k) = \mathbf{a}$. The set L^+ of all ω-limit points is called the *ω-limit set.* Correspondingly, an α-limit point $\mathbf{a}$ is defined by the conditions $t^- = -\infty$, $\lim t_k = -\infty$, and $\lim \mathbf{y}(t_k) = \mathbf{a}$, and the α-limit set L^- as the set of all α-limit points. In order to emphasize the dependence on the initial value $\mathbf{y}(0) = \boldsymbol{\eta}$, one writes $t^+(\boldsymbol{\eta})$, $\gamma^+(\boldsymbol{\eta})$, $L^+(\boldsymbol{\eta})$, ... For a set $A \subset D$, $L^+(A)$ denotes the union of the sets $L^+(\mathbf{a})$ for $\mathbf{a} \in A$. Since $\mathbf{z}(t) = \mathbf{y}(t + t_0)$ is a solution whenever $\mathbf{y}(t)$ is, and since both solutions clearly have the same limit sets, it follows that $L^+(\boldsymbol{\eta}) = L^+(\gamma(\boldsymbol{\eta}))$ and $L^-(\boldsymbol{\eta}) = L^-(\gamma(\boldsymbol{\eta}))$.

A set $M \subset D$ is called *positively invariant* or *negatively invariant* or *invariant* with respect to the differential equation (1) if $\boldsymbol{\eta} \in M$ implies that $\gamma^+(\boldsymbol{\eta}) \subset M$ or $\gamma^-(\boldsymbol{\eta}) \subset M$ or $\gamma(\boldsymbol{\eta}) \subset M$, respectively. The following simple propositions will be stated for positive invariance; corresponding statements also hold for negative invariance and invariance.

(a) If $\mathbf{y}(t)$ is a periodic solution, then $\gamma = \gamma^+ = \gamma^- = L^+(\gamma) = L^-(\gamma)$.

(b) Any union of positively invariant sets is positively invariant. Thus every subset of D contains a largest positively invariant subset (it could be empty).

(c) If $\mathbf{y}(t)$ is a solution with maximal interval of existence J and if $0, s, s+t \in J$, then

$$\mathbf{y}(s+t) = \mathbf{y}(t; \mathbf{y}(s)).$$

(d) Every positive semiorbit is positively invariant, every orbit is invariant.

(e) Let $\mathbf{y}$ be a solution in (t^-, ∞) with $t^- < 0$. Then $\gamma^+ \cap L^+ \neq \emptyset$ implies $\gamma^+ \subset L^+$.

(f) If $M \subset D$ is positively invariant, then so is $\bigcup\{\gamma^+(\boldsymbol{\eta}) : \boldsymbol{\eta} \in M\}$.

The reader is invited to provide the details of the proofs as initiation into the new concepts. The statement (c) says simply that $\mathbf{z}(t) := \mathbf{y}(t+s)$ is the unique solution with $\mathbf{z}(0) = \mathbf{y}(s)$. In (e), let $\mathbf{a}$ be a point in the intersection; then on the one hand, $\mathbf{a} = \mathbf{y}(\tau)$ with $\tau \geq 0$; on the other hand, $\mathbf{a} = \lim \mathbf{y}(t_k)$, where $\lim t_k = \infty$. Thus for arbitrary $s \geq -\tau$, the limit $\mathbf{y}(s+t_k) = \mathbf{y}(s; \mathbf{y}(t_k)) \to \mathbf{y}(s; \mathbf{a}) = \mathbf{y}(s+\tau)$ holds because of the continuous dependence of the solution on initial values. ∎

In the following, $\operatorname{dist}(\mathbf{x}, A) = \inf\{|\mathbf{x} - \mathbf{a}| : \mathbf{a} \in A\}$ is the distance between a point and a set, and $\operatorname{dist}(A, B) = \inf\{|\mathbf{a} - \mathbf{b}| : \mathbf{a} \in A, \mathbf{b} \in B\}$ is the distance between two sets. The next theorem is crucial for later considerations.

Theorem. *Let $\mathbf{y}(t)$ be a solution of (1) in a maximal interval J with $0 \in J$. If $\gamma^+ \subset K$, where K is a compact subset of D, then $t^+ = \infty$ and the limit set $L^+ \subset K$ is nonempty, compact, connected, and (two-sided) invariant, and*

$$\lim_{t \to \infty} \operatorname{dist}(\mathbf{y}(t), L^+) = 0.$$

In particular, all solutions $\mathbf{y}(t; \boldsymbol{\eta})$ with $\boldsymbol{\eta} \in L^+$ exist in $\mathbb{R}$.

Proof. Since $\mathbf{y}(t)$ lies in K on its maximal interval of existence to the right, it follows that the solution exists for all $t > 0$. Therefore, by the Bolzano–Weierstrass theorem, every sequence of the form $(\mathbf{y}(t_k))$ has a convergent subsequence, so L^+ is a nonempty subset of K.

L^+ is closed. To show: If $\mathbf{b}$ is an accumulation point of L^+, then for arbitrary $\varepsilon > 0$ and $T > 0$ there exists a $t > T$ such that $|\mathbf{y}(t) - \mathbf{b}| < \varepsilon$. To prove this, one takes a point $\mathbf{a} \in L^+$ with $|\mathbf{a} - \mathbf{b}| < \varepsilon/2$ and a $t = t_k > T$ such that $|\mathbf{y}(t) - \mathbf{a}| < \varepsilon/2$; then it follows that $|\mathbf{y}(t) - \mathbf{b}| < \varepsilon$.

L^+ is connected. Suppose that L^+ is not connected, i.e., that there exist nonempty, disjoint, compact sets K_1 and K_2 with $L^+ = K_1 \cup K_2$ and $\operatorname{dist}(K_1, K_2) = 2\rho > 0$. Let $d_i(t) := \operatorname{dist}(\mathbf{y}(t), K_i)$, $i = 1, 2$. For each k $(= 1, 2, \ldots)$ there exist points $t_k^1, t_k^2 > k$ such that

$$d_1(t_k^1) < \rho \quad \text{and} \quad d_2(t_k^2) < \rho.$$

Since $d_1(t) + d_2(t) \geq 2\rho$ and since these two functions are continuous, there is a point t_k, lying between t_k^1 and t_k^2, such that

$$d_1(t_k) = \rho \quad \text{and} \quad d_2(t_k) \geq \rho.$$

The sequence $(\mathbf{y}(t_k))$ lies in the compact set K, and it has an accumulation point $\mathbf{a}$ in K. On the other hand, $\operatorname{dist}(\mathbf{a}, K_i) \geq \rho$, $i = 1, 2$, which is a contradiction.

Invariance. Let $\mathbf{a} \in L^+$ and $\lim \mathbf{y}(t_k) = \mathbf{a}$, where $t_k \to \infty$. The solution $\mathbf{y}(t; \mathbf{a})$ exists in a maximal interval J_1. We fix $t \in J_1$ and choose a compact interval $I \subset J_1$ containing 0 and t. According to Theorem 13.X, the solution with initial value $\mathbf{y}(t_k)$ exists at least in I for k large. Now (c) implies

$$\mathbf{y}(t + t_k) = \mathbf{y}(t; \mathbf{y}(t_k)) \to \mathbf{y}(t; \mathbf{a}) \in L^+ \quad \text{as} \quad k \to \infty.$$

Therefore, since $t \in J_1$ is arbitrary, $\gamma(\mathbf{a}) \subset L^+ \subset K$, and consequently, L^+ is invariant. Furthermore, since K is a compact subset of D, it follows that $J_1 = \mathbb{R}$.

The limit relation. Let $\varepsilon > 0$ be chosen so small that the ε-neighborhood L^+_ε of L^+ is contained in D. If a sequence (t_k) exists such that $t_k \to \infty$ and $\mathbf{y}(t_k) \notin L^+_\varepsilon$, then the sequence $(\mathbf{y}(t_k))$ has an accumulation point outside of L^+. With this contradiction, the final assertion $\operatorname{dist}(\mathbf{y}(t); L^+) \to 0$ is also proved. ∎

VI. Attractor and Domain of Attraction. Again $\mathbf{f}$ is assumed to be locally Lipschitz continuous in D. If $\mathbf{f}(\mathbf{0}) = \mathbf{0}$ and if the equilibrium solution $\mathbf{x}(t) \equiv \mathbf{0}$ is asymptotically stable, then the set of all $\boldsymbol{\eta} \in D$ with the property that $\mathbf{y}(t; \boldsymbol{\eta}) \to \mathbf{0}$ as $t \to \infty$ is a neighborhood of the origin. This set is called the *domain of attraction* of $\mathbf{0}$ and is denoted by $\mathcal{A}(\mathbf{0})$. More generally, if $M \subset D$ is a positively invariant set, we define the domain of attraction $\mathcal{A}(M)$ of M to be the set of all points $\boldsymbol{\eta} \in D$ such that $\operatorname{dist}(\mathbf{y}(t; \boldsymbol{\eta}), M) \to 0$ as $t \to \infty$. If $\mathcal{A}(M)$ is a neighborhood of M (superset of an ε-neighborhood), then M is called an *attractor.* If $D = \mathbb{R}^n$ and $\mathcal{A}(M) = \mathbb{R}^n$, then M is called a *global attractor.* In particular, a singleton $M = \{\mathbf{a}\}$ with $\mathbf{f}(\mathbf{a}) = \mathbf{0}$ is an attractor if the solution $\mathbf{x}(t) \equiv \mathbf{a}$ is asymptotically stable.

Lemma. *Let $G \subset D$ be open, $V \in C^1(G)$, and $\dot{V} \leq 0$ in G. Suppose that the set $G_\alpha = \{\mathbf{x} \in G : V(\mathbf{x}) \leq \alpha\}$ is compact for some $\alpha \in V(G)$. Then the following hold:*

(a) *Every solution $\mathbf{y}(t; \boldsymbol{\eta})$ with $\boldsymbol{\eta} \in G_\alpha$ exists for all $t > 0$.*

(b) *G_α is positively invariant.*

(c) *If $\boldsymbol{\eta} \in G_\alpha$, then $L^+(\boldsymbol{\eta}) \subset G_\alpha$ is nonempty and $\dot{V} = 0$ on $L^+(\boldsymbol{\eta})$.*

Proof. We write $\mathbf{y}(t)$ for $\mathbf{y}(t; \boldsymbol{\eta})$ and $\phi(t) = V(\mathbf{y}(t))$. If $\boldsymbol{\eta} \in G_\alpha$, then $\phi(0) \leq \alpha$. As long as $\mathbf{y}(t)$ remains in G, the inequality $\phi'(t) \leq 0$ holds; hence $\phi(t) \leq \alpha$, or, what amounts to the same thing, $\mathbf{y}(t) \in G_\alpha$. Since the distance from G_α to the boundary of G is positive, one arrives at a familiar conclusion, namely, that the solution exists for all $t > 0$ and remains in G_α. This proves (a) and (b).

By Theorem V, $L^+(\boldsymbol{\eta}) =: L^+$ is nonempty and contained in G_α. Assume that $\dot{V}(\mathbf{a}) < 0$ for some $\mathbf{a} \in L^+$. Then $\dot{V}(\mathbf{x}) \leq -\gamma < 0$ holds in a ball $B : |\mathbf{x} - \mathbf{a}| \leq 2\varepsilon$. There exists a sequence (t_k) tending to ∞ such that $|\mathbf{y}(t_k) - \mathbf{a}| < \varepsilon$ and a number $c > 0$, independent of k, such that $|\mathbf{y}(t) - \mathbf{a}| < 2\varepsilon$ for $t \in J_k = (t_k - c, t_k + c)$

and $k = 1, 2, 3, \ldots$ (this follows from the boundedness of $|\mathbf{y}'(t)|$ in $[0, \infty)$). Thus in every interval J_k, $\phi' \leq -\gamma$, and therefore $\phi(t) \to -\infty$ as $t \to \infty$. This contradiction shows that $\dot{V}(\mathbf{a}) = 0$. ∎

Determining, or at least estimating, the domain of attraction is a problem of great practical importance. The following theorem shows how Lyapunov functions can be used in this context.

VII. Theorem. *Let $G \subset D$ be open. Let the function $V \in C^1(G)$ satisfy $\dot{V} \leq 0$ in G and have the property that for every $\alpha \in V(G)$, the set $G_\alpha = \{\mathbf{x} \in G : V(\mathbf{x}) \leq \alpha\}$ is compact. Let M be the largest invariant subset of the set $N := \{\mathbf{x} \in G : \dot{V}(\mathbf{x}) = 0\}$. Then G is contained in the domain of attraction of M; i.e., for $\boldsymbol{\eta} \in G$,* $\operatorname{dist}(\mathbf{y}(t; \boldsymbol{\eta}), M)$ *tends to* 0 *as* $t \to \infty$.

The essential steps in the proof have already been dealt with in the lemma. A point $\boldsymbol{\eta} \in G$ belongs to G_α for $\alpha = V(\boldsymbol{\eta})$. By VI.(c), $L^+ = L^+(\boldsymbol{\eta}) \subset N$, and by Theorem V, L^+ is invariant. Hence $L^+ \subset M$ and, again by Theorem V, $0 \leq \operatorname{dist}(\mathbf{y}(t; \boldsymbol{\eta}), M) \leq \operatorname{dist}(\mathbf{y}(t; \boldsymbol{\eta}), L^+) \to 0$ as $t \to \infty$. ∎

With this we have tools needed to derive sharper theorems on asymptotic stability and instability. The main idea of the stability theorem goes back to LaSalle (1968). The instability theorem was proved by Četaev with the stronger assumption $\dot{V} > 0$ in (b) as early as 1934 and later generalized by Krasovsky.

VIII. Stability Theorem (LaSalle). *Let the function* $\mathbf{f}$ *with* $\mathbf{f}(\mathbf{0}) = 0$ *be locally Lipschitz continuous in D, and let $V \in C^1(D)$ be a Lyapunov function for* $\mathbf{f}$. *If $M = \{\mathbf{0}\}$ is the largest invariant subset of $N = \{\mathbf{x} \in D : \dot{V}(\mathbf{x}) = 0\}$, then the rest state is asymptotically stable.*

Proof. Let $\bar{B}_r \subset D$ and $V(\mathbf{x}) > \gamma > 0$ for $|\mathbf{x}| = r$ $(r > 0)$. Then the set $G = \{\mathbf{x} \in B_r : V(\mathbf{x}) < \gamma\}$ satisfies the hypotheses of the previous theorem, and this theorem gives the conclusion. ∎

IX. Instability Theorem (Četaev–Krasovsky). *Let* $\mathbf{f}$ *satisfy the hypotheses of Theorem* VIII *and let G be an open subset of D with $\mathbf{0} \in \partial G$. Let the function $V \in C^1(G) \cap C(\bar{G})$ satisfy the conditions*

(a) *$V > 0$ in G, $V = 0$ on $\partial G \cap D$;*

(b) *$\dot{V} \geq 0$ in G.*

If the empty set is the only invariant subset of the set $N = \{\mathbf{x} \in G : \dot{V}(\mathbf{x}) = 0\}$, then the equilibrium state is unstable.

Proof. We choose $r > 0$ such that $\bar{B}_r$ lies in D. Let $\boldsymbol{\eta}$ be an arbitrary point from $G \cap B_r$ and $\mathbf{y}(t) := \mathbf{y}(t; \boldsymbol{\eta})$. We will derive a contradiction from the assumption that $\mathbf{y}(t)$ remains in B_r for all $t > 0$. To do this, let $\phi(t) = V(\mathbf{y}(t))$, $\phi(0) = V(\boldsymbol{\eta}) = \alpha > 0$, and $G_\alpha = \{\mathbf{x} \in \bar{G} \cap \bar{B}_r : V(\mathbf{x}) \geq \alpha\}$. Since V vanishes on $\partial G \cap \bar{B}_r$, G_α is a compact subset of G. As long as $\mathbf{y}(t)$ remains in G, $\phi'(t) \geq 0$, which implies that $\phi(t) \geq \alpha$ and hence $\mathbf{y}(t) \in G_\alpha$. From here it follows in the usual way that $\mathbf{y}(t) \in G_\alpha$ for all $t > 0$. Thus the trajectory $\gamma^+(\boldsymbol{\eta})$ is contained in the compact set G_α, and $L^+ = L^+(\boldsymbol{\eta})$ is nonempty and invariant by Theorem V. It follows that $\phi(t)$ is also bounded. One shows exactly as in Lemma VI that the assumptions $\boldsymbol{\eta} \in L^+$, $\dot{V}(\boldsymbol{\eta}) > 0$ imply the relation $\lim \phi(t) = \infty$. Thus we have $\dot{V}(\boldsymbol{\eta}) = 0$ and hence $L^+ \subset N$. Since L^+ is invariant, this gives a contradiction. It shows that each solution that begins in G leaves the ball B_r; since $\mathbf{0} \in \partial G$, the conclusion follows. ∎

We now apply these two theorems to a class of second order equations, which describe nonlinear oscillations with friction. The case of frictionless oscillations was discussed in detail in 11.X.

X. Nonlinear Oscillations with Friction. We consider a differential equation

$$x'' + r(x, x') = 0 \tag{3}$$

for $x = x(t)$. The corresponding autonomous system for $(x, x') = (x, y)$ is given by

$$x' = y, \qquad y' = -r(x, y). \tag{3'}$$

We assume that $r \in C^1(D)$, where D is open and $(0,0) \in D$, $r(0,0) = 0$, and $r(x,0) \neq 0$ for $x \neq 0$. This condition means that $\mathbf{f}(x,y) = (y, -r(x,y))$ has no critical points besides the origin. As a Lyapunov function we choose

$$V(x,y) = \frac{1}{2}y^2 + R(x) \quad \text{with} \quad R(x) = \int_0^x r(s,0)\,ds. \tag{4}$$

In the frictionless case, where $r = r(x)$, this is precisely the energy function from 11.X. One obtains, using the mean value theorem,

$$\dot{V}(x,y) = -y[r(x,y) - r(x,0)] = -y^2 r_y(x, \theta y) \ \text{ with } \ 0 < \theta < 1. \tag{5}$$

The following propositions deal with the stability of the equilibrium state $x(t) \equiv 0$.

(a) $xr(x,0) > 0$ for $x \neq 0$, $r_y(x,y) \geq 0 \Longrightarrow$ the equilibrium state is stable.

(b) $xr(x,0) > 0$ for $x \neq 0$, $r_y(x,y) > 0$ for $xy \neq 0 \Longrightarrow$ the equilibrium state is asymptotically stable.

(c) $r_y(x,y) < 0$ for $xy \neq 0 \Longrightarrow$ the equilibrium state is unstable.

Proof. If $xr(x,0) > 0$, then it follows that $R(x) > 0$ $(x \neq 0)$; hence $0 = V(0,0) < V(x,y)$ for $(x,y) \neq \mathbf{0}$. Therefore, in case (a), $\dot{V} \leq 0$, and the stability theorem II.(a) applies.

In case (b), $\dot{V} < 0$ in $D \setminus N$, $N = \{(x,y) \in D : xy = 0\}$. Let $\xi \neq 0$ and $\eta \neq 0$. The solution with initial value $(\xi, 0)$ satisfies $y'(0) = -r(\xi, 0) \neq 0$, and the solution with initial value $(0, \eta)$ satisfies $x'(0) \neq 0$. Therefore, these solutions do not remain in the set N; i.e., $M = \{\mathbf{0}\}$ is the largest invariant subset of N. Case (b) now follows from Theorem VIII.

In case (c), there are four cases to be distinguished depending on whether $r(x,0)$ is positive or negative for $x > 0$ or $x < 0$. Suppose, for instance, that $r(x,0) < 0$ for $x > 0$ and $x < 0$, which implies that $R(x) < 0$ for $x > 0$ and $R(x) > 0$ for $x < 0$. In this case, the set G consists of all points $(x,y) \in D$ with $x < 0$ and those points with $x \geq 0$ for which $|y| > \sqrt{2|R(x)|}$. On the set G we have $V > 0$; on the two curves $y = \pm\sqrt{2|R(x)|}$, $x \geq 0$, which belong to the boundary, $V = 0$; finally, $\dot{V} > 0$ in G, except for the set $N = \{(x,y) \in G : xy = 0\}$, where $\dot{V}$ vanishes. Every solution that starts on N leaves N, as we have seen in case (b) above, i.e., N does not contain an invariant subset. The conclusion now follows from Theorem IX.

In the case where $xr(x,0) > 0$ for $x \neq 0$, one can choose $G = G \setminus \{(0,0)\}$. The two remaining cases are left to the reader as an exercise. ∎

(d) Propositions (a) through (c) remain valid if the continuous differentiability condition for r is replaced with local Lipschitz continuity. The hypotheses on r_y must then be replaced by corresponding monotonicity conditions, for instance, $r_y < 0$ by "r is strongly monotone decreasing in y." It is easy to see that even less is sufficient, namely a corresponding sign condition for the difference $r(x,y) - r(x,0)$. Incidentally, in the rubber-band example 11.X.(c) the force term r does not belong to C^1.

(e) *The Liénard Equation*

$$x'' + g(x)x' + h(x) = 0$$

describes an oscillation, where $g(x)x'$ represents a friction term that is linear in the velocity and $h(x)$ describes a restoring force. We will assume that g and h are locally Lipschitz continuous. In addition, it will be assumed that $g(x) > 0$ (the friction force acts opposite to the velocity vector).

Here $r(x,y) = g(x)y + h(x)$, and since $r(x,0) = h(x)$, $R(x)$ in (4) is the function $H(x) = \int_0^x h(s)\,ds$ introduced in 11.X and V is the corresponding energy function E. The derivative of this function along trajectories is given by $\dot{V}(x,y) = -g(x)y^2$. Since $r_y = g(x)$, (a), (b), (c) lead to the following

Theorem. *Suppose $xh(x) > 0$ for $x \neq 0$. Then the equilibrium state of the Liénard equation is stable if $g(x) \geq 0$, asymptotically stable if $g(x) > 0$ for $x \neq 0$, and unstable if $g(x) < 0$ for $x \neq 0$.*

To study the behavior of solutions as $t \to -\infty$, one introduces the function $z(t) = x(-t)$. It satisfies the differential equation $z'' - g(z)z' + h(z) = 0$, that is, the original differential equation with $-g(x)$ in place of $g(x)$. The stability properties can then be read off from the theorem.

(f) The

Van der Pol Equation $\qquad x'' = \varepsilon(1 - x^2)x' - x$

is a special case of the Liénard equation. If $\varepsilon > 0$, the zero solution is unstable, for $\varepsilon < 0$ (this corresponds to $\varepsilon > 0$ in the direction $t \to -\infty$), it is asymptotically stable.

Exercises. (g) *Global Attractor.* In the Liénard equation, let g and h be locally Lipschitz continuous. Let $xh(x) > 0$ and $g(x) > 0$ for $x \neq 0$ and suppose $H(x) = \displaystyle\int_0^x h(s)\,ds$ tends to ∞ as $x \to \pm\infty$. Show: The zero solution is a global attractor.

(h) *Nonlinear Friction Force.* Extend the theorem in (e) and the conclusion (g) concerning the Liénard equation to the case of the differential equation

$$x'' + g(x)\psi(x') + h(x) = 0,$$

where $\psi(y)$ with $\psi(0) = 0$ is locally Lipschitz continuous and strongly monotone increasing. An important example is the quadratic resistance law $\psi(y) = y^2 \operatorname{sgn} y$ that is used to describe air resistance in high-velocity motion.

(i) *Domain of Attraction.* Let g and h be locally Lipschitz continuous in $J = (a, b)$ with $a < 0 < b$, and let $xh(x) > 0$ and $g(x) > 0$ in $J \setminus \{0\}$. Denote the limiting values of $H(x) = \displaystyle\int_0^x h(s)\,ds$ as $x \to a$ and $x \to b$ by $H(a)$ and $H(b)$, respectively. Show that in the case of the Liénard equation and, more generally, the equation considered in (h), the set

$$G = \{(x, y) \in J \times \mathbb{R} : \tfrac{1}{2}y^2 + H(x) < \min(H(a), H(b))\}$$

is contained in the domain of attraction of the attractor $M = \{(0, 0)\}$.

(j) Sharpen the result of the Theorem in (e) by showing: If $xh(x) > 0$ for $x \neq 0$ and $g(x) \geq 0$, then the zero solution of the Liénard equation is asymptotically stable if and only if there exists a null sequence (x_k) with $g(x_k) > 0$.

Remarks. The Dutch physicist and radio engineer Balthasar van der Pol (1889–1959) came upon equation (f) in 1926 in describing an electrical circuit with a triode valve. Soon thereafter, A. Liénard investigated the general equation of type (e). The results about the more general equation (3) go back to W. Leighton. Numerous additional results about individual differential equations of second order, in particular, results dealing with the occurrence of periodic solutions, are described in the book by Reissig-Sansone-Conti (1963).

XI. Additional Examples and Remarks. (a) *Gradient Systems.* This is the name given to systems in which $\mathbf{f}$ has a potential function $g \in C^1(D)$ such that $\mathbf{f}(\mathbf{y}) = -\operatorname{grad} g(\mathbf{y})$ (the minus sign is used in physics). For the equation

$$\mathbf{y}' = -\operatorname{grad} g(\mathbf{y}) \tag{6}$$

a natural choice of a Lyapunov function is the function $V(\mathbf{x}) = g(\mathbf{x})$. We have then

$$\dot{V} = -|\operatorname{grad} g(\mathbf{x})|^2.$$

Thus, if g has a local minimum at the point $\mathbf{a} \in D$ and if there exists a neighborhood N of $\mathbf{a}$ such that $g(\mathbf{x}) > g(\mathbf{a})$ and $\operatorname{grad} g(\mathbf{x}) \neq \mathbf{0}$ in $N \setminus \{\mathbf{a}\}$, then the equilibrium state $\mathbf{x}(t) \equiv \mathbf{a}$ is asymptotically stable. This follows from Theorem II.(b).

The following example is important from both a theoretical and a historical standpoint.

(b) *Motion in a Conservative Force Field.* Let a conservative force field $\mathbf{k}$ be defined on an open set $D_1 \subset \mathbb{R}^n$, i.e., let there exist a potential $U \subset C^1(D_1)$ with $\mathbf{k}(\mathbf{x}) = -\operatorname{grad} U(\mathbf{x})$. The equation of motion $\mathbf{x}'' = \mathbf{k}(\mathbf{x})$ then reads

$$\mathbf{x}'' = -\operatorname{grad} U(\mathbf{x}) \Leftrightarrow \mathbf{x}' = \mathbf{y},\ \mathbf{y}' = -\operatorname{grad} U(\mathbf{x}). \tag{7}$$

Thus we are dealing with a system of $2n$ equations in the set $D = D_1 \times \mathbb{R}^n \subset \mathbb{R}^{2n}$. As a Lyapunov function, we take the energy function

$$V(\mathbf{x}, \mathbf{y}) = U(\mathbf{x}) + \frac{1}{2}|\mathbf{y}|^2$$

(the sum of the potential and kinetic energy). A simple calculation shows that $\dot{V}(\mathbf{x}, \mathbf{y}) \equiv 0$. Therefore, V is constant along trajectories of solutions; this is the theorem of conservation of energy.

The equation $\operatorname{grad} V(\mathbf{x}, \mathbf{y}) = (\operatorname{grad} U(\mathbf{x}), \mathbf{y}) = (\mathbf{0}, \mathbf{0})$ is satisfied if and only if $\operatorname{grad} U(\mathbf{x}) = \mathbf{0}$ and $\mathbf{y} = \mathbf{0}$. From this observation we obtain the following:

(b_1) Let $\operatorname{grad} U(\mathbf{a}) = \mathbf{0}$, where $\mathbf{a} \in D_1$. If the potential U has a strong minimum at $\mathbf{a}$, then the constant solution $\mathbf{x}(t) \equiv \mathbf{a}$, that is, the solution $(\mathbf{x}(t), \mathbf{y}(t)) \equiv (\mathbf{a}, \mathbf{0})$, is stable. This follows again from Theorem II. Incidentally, the scalar equation $x'' + h(x) = 0$ with $U(x) = H(x)$ is a special case.

(c) *Motion in a Force Field with Friction.* As a rule, the frictional force has the direction of $-\mathbf{x}'$. We allow a general term of the form $-\psi(\mathbf{x}, \mathbf{x}')A\,\mathbf{x}'$ with $(A\mathbf{y}, \mathbf{y}) \geq \alpha|\mathbf{y}|^2$ ($\alpha > 0$) and nonnegative ψ; this implies that the angle between $-\mathbf{x}'$ and the frictional force is smaller that $\pi/2$. The resulting equation has the form

$$\mathbf{x}'' + \psi(\mathbf{x}, \mathbf{x}')A\mathbf{x}' + \operatorname{grad} U(\mathbf{x}) = \mathbf{0}.$$

We use the energy function V from (b) and obtain

$$\dot{V}(\mathbf{x}, \mathbf{y}) = -\psi(\mathbf{x}, \mathbf{y})(A\mathbf{y}, \mathbf{y}) \leq -\alpha\psi(\mathbf{x}, \mathbf{y})|\mathbf{y}|^2.$$

(c_1) Let ψ and α, with $(A\mathbf{y}, \mathbf{y}) \geq \alpha|\mathbf{y}|^2$, be positive and let U have a strong minimum at $\mathbf{0}$. Then the equilibrium state $\mathbf{x}(t) \equiv \mathbf{0}$ is asymptotically stable.

This result follows from Theorem VIII using a setup that is similar to the one-dimensional case in X.(b).

(d) *Hamiltonian Systems.* Let the real-valued function $H(\mathbf{x}, \mathbf{y})$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, belong to $C^2(D)$, where $D \subset \mathbb{R}^{2n}$ is open. An autonomous system of $2n$ differential equations of the form

$$\mathbf{x}' = H_{\mathbf{y}}(\mathbf{x}, \mathbf{y}), \qquad \mathbf{y}' = -H_{\mathbf{x}}(\mathbf{x}, \mathbf{y}) \tag{8}$$

is called a *Hamiltonian system*, and the function H is called a *Hamiltonian function.* The Hamiltonian function can be used as a Lyapunov function; indeed,

$V = H$ satisfies $\dot{V} \equiv 0$ in D, as one easily sees. It follows then from Theorem II that

(d$_1$) A strong minimum of the Hamiltonian function is a stable equilibrium state for equation (8).

A Hamiltonian system for $n = 1$ was already treated in 3.V. The potential function F is a Hamiltonian function for the differential equation (3.13). Also, the equation of motion treated in (b) is of type (8), where the total energy function is the Hamiltonian function.

As an outlook on more recent developments, we consider a three-dimensional autonomous system that despite its simplicity, exhibits exceptionally rich and complicated dynamics. The equations were proposed by the meteorologist and mathematician E. N. Lorenz as a very crude model of a convective (predominantly vertical) flow realized by a fluid that is warmed from below and cooled from above. The example has attracted great attention, and its stimulating effect persists in the recent research on chaotic motion.

XII. The Lorenz Equations. These equations read

$$\begin{aligned} x' &= \sigma(y - x), \\ y' &= rx - y - xz, \\ z' &= xy - bz, \end{aligned} \tag{9}$$

where σ, r, and b are positive constants.

We formulate some properties of the solutions to this system as exercises, with hints for the proof.

(a) *Symmetry.* If $(x(t), y(t), z(t))$ is a solution, then so is $(-x(t), -y(t), z(t))$.

(b) The positive and negative z-axes are invariant sets.

(c) The origin is a critical point for all parameter values. If $0 < r < 1$, the origin is a global attractor and the zero solution is asymptotically stable.

(d) If $r > 1$, then the zero solution is unstable.

(e) Every solution has a maximal interval of existence of the form $J = (t^-, \infty)$. There exists a compact, positively invariant set $E \subset \mathbb{R}^3$ (depending on σ, r, b) that every solution enters at some time and thereafter never again leaves.

Hints for the proofs. (c) Use the Lyapunov function $V(x, y, z) = x^2 + \sigma y^2 + \sigma z^2$ and show that the hypotheses of Theorem II.(b) and Theorem VII with $G = \mathbb{R}^3$ are satisfied.

(d) Calculate the matrix A of the linearized system and show that A has three real eigenvalues, two negative and one positive.

(e) Consider the Lyapunov function $V = rx^2 + \sigma y^2 + \sigma(z - 2r)^2$, calculate the derivative $\dot{V}$, and show that the set $A = \{(x, y, z) \in \mathbb{R}^3 : \dot{V}(x, y, z) \geq -\delta\}$ is compact ($\delta > 0$). Let M be the maximum of V on A and let E be the set of all points with $V(x, y, z) \leq K$ (E is an ellipsoid with center $(0, 0, 2r)$). If $\mathbf{v}(t) = (x(t), y(t), z(t))$ is a solution and $\phi(t) = V(\mathbf{v}(t))$, then show that $(*)$ $\phi(t) \geq K$ implies $\phi'(t) \leq -\delta$, and derive the conclusion from $(*)$.

These properties lie on the surface. Anyone interested in digging deeper can consult the book *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors* by C. Sparrow (Springer Verlag 1982).

Appendix

In this appendix, concepts and theorems from topology, real, and complex analysis and functional analysis that are used in the text are formulated. In most cases the theorems are given with proofs or at least with sketches of the proofs. At some points the theory is deepened.

A. Topology

In this section we present some basic facts about paths and curves for use in the investigation of differential equations. We begin with some definitions and elementary results, followed by a discussion of the polar-coordinate representation of curves that is needed for the Prüfer transformation in § 27.IV. We then introduce the winding number and give a statement of the Jordan curve theorem without proof. Next, theorems on level curves that are used to check for the existence of periodic solutions are presented. The section concludes with some theorems on autonomous systems of differential equations for $n = 2$ that in essence spell out that a solution that starts on a level curve traces it out entirely. The proofs for these theorems are independent of the previous results.

There is no consensus in the textbook literature about the concept of a curve. In some branches of mathematics the emphasis is on the curve as a set—a one-dimensional manifold; in others (particularly in mechanics) it is important to know how the curve is traced out, which is accomplished by introducing a time dependent path function. Both aspects come forward here.

I. Paths and Curves. A continuous function $\phi : I = [a, b] \to \mathbb{R}^n$ is called a *path* in $\mathbb{R}^n$, and the image set $C = \phi(I)$ is called a *curve* with the parametric representation ϕ; we will also use the notation $\phi|_I$ when we wish to call attention to the interval I. The point $\phi(a)$ is called the *initial point* of the path, the point $\phi(b)$ the *terminal point*. The path ϕ is called a *Jordan path* if the mapping ϕ is injective, and a *closed Jordan path* if $\phi(a) = \phi(b)$ and ϕ is injective on $[a, b)$. If $\phi \in C^1(I)$ and $\phi'(t) \neq 0$, then ϕ is called a *smooth path*.

The same terminology is used for the curve generated by ϕ. Thus, for instance, the set $C \subset \mathbb{R}^n$ is called a *closed Jordan curve* if there exists a closed Jordan path $\phi|_I$ with $\phi(I) = C$.

(a) *Smooth Closed Paths and Curves.* Of a smooth closed path we demand that besides being smooth and closed, it satisfy (i) $\phi'(a) = \lambda\phi'(b)$ with $\lambda > 0$. The relation (i) is equivalent to the condition that the curve $\phi(I)$ have a tangent at the point $\phi(a) = \phi(b)$ (there is a peak at $\phi(a)$ if (i) holds with $\lambda < 0$). If (i) holds, then there is a change of parameter of the form $\phi_1(t) = \phi(h(t))$ with $h(I) = I$ such that ϕ_1 satisfies (i) with $\lambda = 1$ and generates the same curve C (for example, one can take $h(t) = a + \alpha(t-a) + \beta(t-a)^2$ with $\alpha = 2/(1+\lambda)$, $\beta = (1-\alpha)/(b-a)$). This ϕ_1 can be extended as a periodic C^1-function of period $p = b - a$ to all of $\mathbb{R}$, and each of the closed paths $\phi_1|_{[c,c+p]}$ (c arbitrary) generate the same curve C. In particular, any point of C can be taken as the initial point.

(b) *Splicing Paths.* Paths can be joined together: If $\phi|_I$ and $\psi|_J$ are paths with $I = [a,b]$ and $J = [b,c]$ and if $\phi(b) = \psi(b)$, then we denote by $\omega = \phi \oplus \psi$ the path in the interval $I \cup J$ defined by $\omega|_I = \phi$, $\omega|_J = \psi$. This construction can also be carried out when J does not connect to I. For example, if $J = [\alpha, \beta]$ and, of course, the relation $\phi(b) = \psi(\alpha)$ holds, then one introduces a change of parameter in ψ ($t' = t + b - \alpha$) and then proceeds as above.

(c) *Reversing Orientation.* From the path $\phi|_I$ one obtains, by reversing the orientation, the path $\phi^-|_I$, defined by $\phi^-(t) = \phi(a+b-t)$. The path is traced out in the reverse direction with initial and terminal points exchanged. However, ϕ and ϕ^- generate the same curve.

More results concerning paths and curves, including the definition of the path length L, the formula

$$L = \int_a^b |\phi'(t)|\, dt \quad \text{for} \quad \phi \in C^1(I),$$

and the corresponding formula for the arclength, can be found in standard analysis textbooks.

II. Connectedness. An open set $G \subset \mathbb{R}^n$ is called *connected* (or, more exactly, *path-connected*) if every pair of points from G can be connected by a path in G, that is, given $x, y \in G$, there exists $\phi|_I$ with $\phi(a) = x$, $\phi(b) = y$, and $\phi(I) \subset G$. A nonempty open connected set will be called a *domain.* We note that in topological spaces a different definition of connectedness is used; the two notions are equivalent for open sets in $\mathbb{R}^n$.

Now let G be an arbitrary nonempty open set and $x, y \in G$. If x and y can be connected in G, then we write $x \sim y$. This relation is an equivalence relation, and the corresponding equivalence classes are pairwise disjoint open connected subsets of G whose union is G. These subsets are called *components* of G. The set G is a domain if it has only one component, namely G. It is easy to prove

(a) *If G, H are domains with $G \cap H \neq \emptyset$, then $G \cup H$ is also a domain.*

III. Plane Curves. Polar Coordinate Representation. A point (x,y) in the plane can be represented as a complex number $z = (x,y)$, also written as $z = x + \mathrm{i}y$. In the second notation, x stands for $(x, 0)$ and i for $(0, 1)$, and the definition of multiplication in $\mathbb{C}$ leads to $\mathrm{i}y = (0, y)$. The function

$$\mathrm{e}^{\mathrm{i}t} = (\cos t, \sin t) = \cos t + \mathrm{i} \sin t$$

is 2π-periodic, and $|\mathrm{e}^{\mathrm{i}t}| = 1$.

Every point $z \neq 0$ has a polar coordinate representation

$$z = r\mathrm{e}^{\mathrm{i}\phi};$$

here $r = |z| = \sqrt{x^2 + y^2}$ is the *modulus* of z. The *argument* $\phi = \arg z$ is uniquely determined up to a multiple of 2π. For $z, z' \neq 0$, we have the relations

$$\arg \frac{1}{z} = -\arg z, \quad \arg zz' = \arg z + \arg z' \pmod{2\pi}. \tag{1}$$

Using the principal value of the arc functions, the relations

$$\arg z = \begin{cases} \arctan(y/x) & \text{for} \quad x > 0 \quad [+\pi \text{ for } x < 0], \\ \operatorname{arccot}(x/y) & \text{for} \quad y > 0 \quad [+\pi \text{ for } y < 0] \end{cases} \tag{2}$$

determine one value of the argument. All other values are obtained from here adding $2k\pi$ (k an integer). The principal value of the argument is denoted by $\operatorname{Arg} z$; it is defined by the inequalities $-\pi < \operatorname{Arg} z \leq \pi$.

The following results apply to paths in the plane $\zeta(t) = (\xi(t), \eta(t)) : I \to \mathbb{R}^2$. We first show that such a path has a polar coordinate representation

$$\zeta(t) = r(t)\mathrm{e}^{\mathrm{i}\phi(t)} \quad (\textit{with continuous argument function} \quad \phi(t)). \tag{3}$$

This result requires a simple fact that follows from the intermediate value theorem.

(a) *Suppose that for a function $f \in C(I)$ there exists a $\delta > 0$ such that at each point $t \in I$, either $f(t) = 0$ or $|f(t)| \geq \delta$. Then if f vanishes at a point of I, $f(t) \equiv 0$ in I.*

Theorem. *A path $\zeta|_I$ with $\zeta(t) \neq 0$ has a representation* (3) *with an argument function ϕ that is continuous in I. This representation is unique modulo 2π; i.e., every other continuous argument function is of the form $\phi(t) + 2k\pi$, where k is an integer.*

If $\zeta \in C^k(I)$, then the functions $r(t) = |\zeta(t)|$ and $\phi(t)$ are also in $C^k(I)$.

Proof. Let $\zeta(t) = (\xi(t), \eta(t)) \in C^k(I)$ $(k \geq 0)$. Suppose, to consider a specific case, that at $t = \tau$, $\xi(\tau) > 0$. Then in a neighborhood J_τ of τ one obtains an argument function in $C^k(J_\tau)$ by setting $\phi(t) := \arctan \eta(t)/\xi(t)$. Proceeding in a similar manner at each point $t \in I$ and using the appropriate choice of the two possibilities given in (2), one obtains a corresponding interval neighborhood J_t and a function $\phi = \arg \zeta \in C^k(J_t)$. By the Borel covering theorem, finitely many of these interval neighborhoods, say $J_1, \ldots, J_p$, are a cover for I. Let these neighborhoods be numbered such that $J_1 = [a, t_1)$, $t_1 \in J_2 = (s_2, t_2)$, $t_2 \in J_3 = (s_3, t_3)$, $\dots$, $t_{p-1} \in J_p = (s_p, b]$, and let ϕ_j be the argument function that corresponds to J_j. We begin the construction of an argument function $\phi \in C^k(I)$ by setting $\phi(t) := \phi_1(t)$ in J_1. Then a point t in $J_1 \cap J_2$ is chosen, and one determines an m such that $\phi_1(t) = \phi_2(t) + 2m\pi$. By (a), $\phi_1 \equiv \phi_2 + 2m\pi$ holds in $J_1 \cap J_2$, and hence the definition $\phi(t) := \phi_2(t) + 2m\pi$ in J_2 gives a function in $C^k(J_1 \cup J_2)$. Proceeding in this manner one eventually obtains $\phi = \arg\zeta \in C^k(I)$.

The uniqueness of ϕ modulo 2π follows immediately from (a). ∎

Corollary. *Let $\zeta|_I$ and $\zeta^*|_I$ be two paths that do not contain the origin and suppose* $\operatorname{Re}(\zeta^*(t)/\zeta(t)) > 0$. *If the argument function $\phi(t) = \arg\zeta(t)$ is continuous, then*

$$\phi^*(t) := \phi(t) + \operatorname{Arg}\frac{\zeta^*(t)}{\zeta(t)}$$

defines a continuous argument function $\phi^*(t) = \arg\zeta^*(t)$.

By (1) $\phi^* = \arg\zeta^*$, and since $\operatorname{Re}(\zeta^*/\zeta) > 0$, the arctan formula applies in all of I. Hence ϕ^* is continuous. ∎

IV. The Winding Number. Let $I = [a, b]$ and $\zeta|_I$ be a closed path that does not pass through the origin. Making use of the (continuous) polar coordinate representation $\zeta(t) = r(t)\mathrm{e}^{\mathrm{i}\phi(t)}$, we define the winding number of ζ (with respect to 0) by

$$U(z) := \frac{1}{2\pi}\{\phi(b) - \phi(a)\} \qquad \textit{winding number.}$$

Since $\zeta(a) = \zeta(b)$, the winding number U is an integer, and by Theorem III, it is independent of the choice of the argument function ϕ. The name winding number points to the fact that $U(\zeta)$ is the number of times the path ζ winds around the origin in the positive (counterclockwise) sense. The winding number is also called *index* of ζ.

Example. For $\zeta(t) = \mathrm{e}^{\mathrm{i}kt}$ ($k \neq 0$ an integer) on $I = [0, 2\pi]$ we have $U(\mathrm{e}^{\mathrm{i}kt}) = k$.

The winding number $U(z;\zeta)$ of a closed path $\zeta|_I$ about a point $z \notin \zeta(I)$ is defined in an entirely analogous manner. First a representation $\zeta(t) = z + r(t)\mathrm{e}^{\mathrm{i}\phi(t)}$, i.e., a representation (3) of $\zeta(t) - z$, is constructed, and then one sets

$$U(z;\zeta) = \frac{1}{2\pi}\{\phi(b) - \phi(a)\}.$$

The earlier statements about U also hold in this case.

Theorem. *Let $C = \zeta(I)$. The winding number $U(z;\zeta)$ is constant on each component of the open set $G = \mathbb{R}^2 \setminus C$.*

Proof. It is sufficient to show that the function $U(z) := U(z;\zeta)$ is continuous in G. To see that this is the case, let $z_1, z_2 \in G$ and $\psi|_{[0,1]}$ be a path connecting z_1 and z_2 in G. If U is continuous, then so is the function $h(t) := U(\psi(t))$. Since $h(t)$ is integer-valued, it follows from III(a) that h is constant on $[0,1]$ and hence that $U(z_1) = U(z_2)$.

To prove the continuity of U at a point $z \in G$, we consider points z^* such that $|z - z^*| < \rho := \operatorname{dist}(z, C)$. Then $|z - z^*| < |\zeta(t) - z|$, and hence

$$\operatorname{Re} \frac{\zeta(t) - z^*}{\zeta(t) - z} = \operatorname{Re} \left(1 + \frac{z - z^*}{\zeta(t) - z}\right) > 0.$$

By Corollary III the function

$$\phi^*(t) = \phi(t) + \operatorname{Arg} \frac{\zeta(t) - z^*}{\zeta(t) - z},$$

where $\phi(t) = \arg(\zeta(t) - z)$, is a continuous argument function. Clearly, $\phi^*(t) \to \phi(t)$ and with it $U(z^*) \to U(z)$ as $z^* \to z$. ∎

We come now to the Jordan curve theorem. This is one of those theorems that appear obvious on the surface but are difficult to prove. A proof can be found in textbooks on topology or complex analysis, e.g., R.B. Burckel (1979).

V. Jordan Curve Theorem. *A closed Jordan curve C separates the plane into two connected parts. More precisely: The open set $\mathbb{R}^2 \setminus C$ consists of two components, a bounded component* $\operatorname{Int}(C)$, *the inside, and an unbounded component* $\operatorname{Ext}(C)$, *the outside of C, and C is the boundary of each component.*

If C is generated by the closed Jordan path ζ, then the winding number $U(z;\zeta)$ is $+1$ or -1 in $\operatorname{Int}(C)$ *and zero in* $\operatorname{Ext}(C)$.

Positive and Negative Orientation. One says that the path ζ is *positively oriented* if the winding number is $+1$ in $\operatorname{Int}(C)$ and is *negatively oriented* if it is -1 there. Intuitively, positive orientation means that the interior lies to the left if one proceeds on C in the direction of the path. The unit circle is positively oriented in the conventional representation $z = e^{it}$, $0 \le t \le 2\pi$.

VI. Simply Connected Domains. A domain $G \subset \mathbb{R}^2$ is called *simply connected* if the inside of every closed Jordan curve lying in G also belongs to G. This property means that G does not have any holes. The notion of a simply connected domain arises in 3.III and in the existence theorem 21.II, among others.

Exercise. A plane domain G said to be *convex* if the line segment $\overline{z_1 z_2}$ connecting two arbitrary points z_1, z_2 from G also lies in G; it is called *starlike* with respect to a point $a \in G$ if for every $z \in G$, $\overline{az} \subset G$. Prove: Every convex domain is starlike, and every starlike domain is simply connected. Give an example of a nonconvex starlike domain.

VII. Level Curves. For a continuously differentiable function $F : G \subset \mathbb{R}^2 \to \mathbb{R}$ (G open), we consider the level sets

$$M_\alpha = \{z \in G : F(z) = \alpha\} = F^{-1}(\alpha).$$

A point z where $\operatorname{grad} F(z) = 0$ is called a *critical* (or *stationary*) *point* of F. The following theorem gives a criterion for level sets to be closed Jordan curves. It plays an important role in the investigation of periodic solutions to differential equations.

Theorem. *Let $G \subset \mathbb{R}^2$ be open and $F \in C^1(G, \mathbb{R})$. If $M_\alpha = F^{-1}(\alpha)$ is a nonempty compact subset of G that does not contain any critical points, then M_α consists of finitely many smooth closed Jordan curves.*

Proof. The implicit function theorem implies that in a neighborhood of a point $z_0 \in M_\alpha$ the equation $F(x, y) = \alpha$ can be resolved either in the form (i) $y = f(x)$ or (ii) $x = g(y)$. More specifically, there exists an open neighborhood $R = I \times J$ of z_0 with $I = (a, b)$, $J = (c, d)$ and in the case (i) a function $f \in C^1(\bar{I})$ such that $F(x, f(x)) = \alpha$ in $\bar{I}$ and $F(x, y) \neq \alpha$ for all (x, y) from $\bar{R}$ with $y \neq f(x)$. This means that

$$\bar{R} \cap M_\alpha = \operatorname{graph} f|_{\bar{I}} =: C, \quad R \cap M_\alpha = \operatorname{graph} f|_I =: C^0.$$

Here C is a Jordan curve with the parametric representation $z = \zeta(t) = (t, f(t))$, $t \in \bar{I}$, and $C = C^0 \cup \{z', z''\}$, $z' = \zeta(a)$, $z'' = \zeta(b)$. In the case (ii) the statement is similar with $\zeta(t) = (g(t), t)$, $C = \zeta(\bar{J})$, $z' = \zeta(c)$, ...

Every point of M_α is associated with such a rectangular neighborhood. By the Borel covering theorem, a finite number of these neighborhoods, say R_1, ..., R_p, that already cover M_α can be chosen from this collection. We modify the above notation,

$$\begin{aligned} z_k \in R_k &= I_k \times J_k, \quad I_k = (a_k, b_k), \quad J_k = (c_k, d_k), \\ C_k^0 &= R_k \cap M_\alpha, \quad C_k = \bar{R}_k \cap M_\alpha = C_k^0 \cup \{z_k', z_k''\}, \end{aligned}$$

and associate the corresponding parametric representation $\zeta_k|_{L_k}$ with $L_k = [a_k', b_k']$, where $L_k^0 = I_k$ in case (i) and $L_k^0 = J_k$ in case (ii).

We assume further that no unnecessary rectangles appear, i.e., that from $C_k \subset C_l$ it follows that $k = l$. We start with R_1 and assume that case (i) is present. We have $\bar{R}_1 \cap M_\alpha = C_1 = \zeta_1(L_1)$. The path $\zeta_1|_{L_1}$ has an orientation (from left to right), the endpoint $z_1'' = \zeta_1(b_1')$ does not lie in R_1. Thus there exists a rectangle, say R_2, with $z_1'' \in R_2$. An end piece $C_1^* = C_1 \cap R_2$ of C_1 is simultaneously a starting piece of C_2. We carry the orientation of C_1 over to C_2 in that we reorient ζ_2 if necessary (however, we retain the notation ζ_2 as well as z_2', z_2'' for the initial and terminal points); cf I.(c). Again $z_2'' = \zeta_2(b_2') \notin R_2$. There exists a rectangle R_3 that contains z_2'' and in which $C_3 = C_3^0 \cup \{z_3', z_3''\}$ lies. By assumption C_2 is not a subset of C_3; hence $z_3' \in C_2^0$. We proceed in this manner. Since M_α is covered by a finite number of rectangles R_k, the following case arises at, say, the mth step: We have $z_m'' := \zeta_m(b_m') \notin R_m$, but $z_m \in R_k$ for some $k < m$, and hence $z_m'' \in C_k^0$. Since $R := R_1 \cup \cdots \cup R_m$ is open and all points of $C_1 \cup \cdots \cup C_m$ with the exception of z_1' are interior points of R, the path ζ_m can enter R only at the point z_1'. With this we have essentially shown, that $C = C_1 \cup \cdots \cup C_m$ is a closed Jordan curve.

In order to obtain a unified representation $\zeta|_L$ for C from the separate parametric representations $\zeta_k|_{L_k}$ of the individual curve segments C_k, one must first decrease the size of the L_k such that the curve segments no longer overlap but rather connect to one another. On the new intervals $L_k = [\alpha_k, \beta_k]$ that arise in this process we have $\zeta_1(\beta_1) = \zeta_2(\alpha_2)$. However, only $\zeta_1'(\beta_1) = \lambda\zeta_2'(\alpha_2)$ with $\lambda > 0$. In order to get $\lambda = 1$, one replaces the parameter t in ζ_2 by γt and L_2 by $\frac{1}{\gamma}L_2$, using a suitable $\gamma > 0$. Proceeding in this way successively with $\zeta_3, \ldots, \zeta_{m-1}, \zeta_m$, one obtains new representations of the C_k, for which $\zeta_k(\beta_k) = \zeta_{k+1}(\alpha_{k+1})$ and $\zeta_k'(\beta_k) = \zeta_{k+1}'(\alpha_{k+1})$ (with the understanding that $\zeta_{n+1} = \zeta_1$). By coupling the parametric representations (cf., I.(b)) one eventually obtains a smooth closed Jordan path for C satisfying the conditions I.(a). ∎

As a simple consequence, one obtains the following useful theorem.

Theorem. *Let G be a simply connected domain and suppose $F \in C^1(G)$ has a global maximum at the point z_0 and no other critical points in G. Let $F(z_0) =: B$ and let there exist an $A < B$ ($A = -\infty$ is allowed) such that for every sequence (z_n) in G with $\lim z_n \in \partial G$ or $\lim |z_n| = \infty$ the relation $\limsup F(z_n) \le A$ holds.*

Then for every $\alpha \in (A, B)$, the level set $C_\alpha = F^{-1}(\alpha)$ is a closed smooth Jordan curve and $F(z) > \alpha$ in $\operatorname{Int}(C_\alpha)$ and $F(z) < \alpha$ in $\operatorname{Ext}(C_\alpha)$.

It follows immediately that $z_0 \in \operatorname{Int}(C_\alpha)$ and $C_\alpha \subset \operatorname{Int}(C_\beta)$ for $A < \beta < \alpha < B$.

Proof. Let $A < \alpha < B$. Since G is connected, $(A, B) \subset F(G)$, and hence $M_\alpha = F^{-1}(\alpha)$ is not empty. Assume that M_α is not a compact subset of G. Then there exists a sequence (z_n) in M_α with $\lim z_n \in \partial G$ or $\lim |z_n| = \infty$, which leads to a contradiction to the assumptions because $F(z_n) = \alpha > A$. Hence M_α is compact, and by the previous theorem there exists a closed Jordan curve $C_\alpha \subset M_\alpha$. We write I_α for $\operatorname{Int}(C_\alpha)$ and E_α for $\operatorname{Ext}(C_\alpha)$. Since z_0 is the only critical point, F has no local minimum in I_α and we have $F(z) > \alpha$ in I_α.

Let $A < \beta < \alpha$ and $M = \{z \in G : \beta \le F(z)\}$. Using a similar argument, one sees that M is a compact subset of G. The set $N = M \setminus I_\alpha$ is likewise compact. Let $\gamma = \max F(N) = F(z_1)$ with $z_1 \in N$. From $\gamma > \alpha$ it follows that $z_1 \in E_\alpha$, but since E_α does not contain any critical points, this case is ruled out. Therefore, we have $\gamma = \alpha$. A point $z_1 \in E_\alpha$ with $F(z_1) = \alpha$ would also be critical, and therefore $F(z) < \alpha$ in E_α. ∎

VIII. Autonomous Differential Equations in the Plane. We now give some important results on autonomous differential equations in the plane. The results in this subsection do not make use of the previous theorems.

First some preliminary results. Here $I = [\alpha, \beta]$ and $J = [a, b]$ are compact intervals.

(a) Let the function $h : J \to \mathbb{R}$ be continuous and locally injective (i.e., for every $t \in J$ there exists a neighborhood U such that the restriction $h|_{U \cap J}$ is injective). Then h is injective in J and hence strongly monotone increasing or decreasing.

(b) Let $\zeta|_I$ be a Jordan path, $C = \zeta(I)$, and $z|_J$ a smooth path with $z(J) \subset C$. Then z is a Jordan path, and there exists a uniquely determined continuous and strongly monotone function $h : J \to I$ with $z(t) = \zeta(h(t))$ for $t \in J$.

Proof. The proof of (a) is elementary. First, h is injective in an interval $[a, a + \varepsilon]$ and hence, for example, strongly increasing. If $c = \sup\{t \in J :$ h is strongly increasing in $[a, t]\}$, then the assumption $c < b$ leads to a contradiction.

(b) Because of the compactness of I and the bijectivity of $\zeta : I \to C$, the inverse function ζ^{-1} is continuous and injective, and hence $h = \zeta^{-1} \circ z : J \to I$ is continuous. From $z'(t) = (x'(t), y'(t)) \neq 0$ it follows, to take a specific case, that $x'(t) \neq 0$. Therefore, x is strongly monotone in a neighborhood U of t, and therefore z and also h are injective in U. By part (a), h is injective in J. Clearly, h is uniquely determined by $z = \zeta \circ h \Longleftrightarrow h = \zeta^{-1} \circ z$. ∎

In the example of the mathematical pendulum (11.X.(d)), three kinds of level curves arise: closed Jordan curves, separatices, and infinite curves. We can now give an answer in general to the question whether a solution of an autonomous system

$$\dot{x} = f(x, y), \quad \dot{y} = g(x, y) \tag{4}$$

that begins on such a curve traces out the entire curve. We assume that f and g are locally Lipschitz continuous in a plane domain G and refer to the results from 10.XI.

Theorem on Periodic Solutions. *Let $C \subset G$ be a closed Jordan curve and $(f, g) \neq 0$ on C. Let the solution $z(t) = (x(t), y(t))$ of (4) with the maximal interval of existence $J^\circ = (a, b)$ $(\infty \leq a < b \leq \infty)$ run along C; i.e., let $z(J^\circ) \subset C$ hold. Then $z(J^\circ) = C$, $J^\circ = \mathbb{R}$, and the solution $z(t)$ is periodic.*

Remark. The significance of this theorem comes from the fact that in important cases the trajectories are determined as level curves of a potential function $F(x, y)$ as described in 3.V. Examples of this type were presented in 3.VI–VII, 11.XI and elsewhere.

Proof. By 10.XI.(a) $J^\circ = \mathbb{R}$. We may assume that $C = \zeta(I)$ with $I = [\alpha, \beta]$, $\zeta(\alpha) = \zeta(\beta) = z(0)$; cf. I.(a). Let $c = \sup\{t \in J^\circ : z([0,t]) \neq C\}$. Clearly, $0 < c \leq \infty$. For arbitrary $c' < c$, $z([0,c'])$ is contained in a (not closed!) Jordan curve $C' \neq C$, $C' \subset C$. By (b) we have (i) $z(t) = \zeta(h(t))$ for $0 \leq t \leq c'$, where h is injective. We assume that h is strongly monotone increasing (if not, the path ζ is reoriented). Thus since c' is arbitrary, equation (i) holds in the half-open interval $[0,c)$. Here $h(0) = \alpha$, and we set $\gamma = \lim_{t \to c-} h(t)$. By (i), $z(t) \to \zeta(\gamma) \in C$ as $t \to c$. If $c = \infty$, then by 10.XI.(h), $\zeta(\gamma)$ would be a critical point of the system (4), in contradiction to the assumptions. Therefore, $c < \infty$. We set $h(c) = \gamma$. Then h is continuous in $[0,c]$, strongly monotone increasing, and (i) holds in $[0,c]$.

The assumption $\gamma < \beta$ contradicts the maximality of c, since then the curve $z([0,c])$ is disjoint from the curve segment $\zeta((\gamma,\beta))$ and thus $z([0,c+\varepsilon]) \neq C$ for small positive ε. Therefore, we have $\gamma = \beta$, i.e., $z(c) = z(0)$, and it follows from 10.XI.(b) that $z(t) = z(t+c)$ in $\mathbb{R}$. Finally, from (i) we have that $z([0,c]) = C$ and c is the smallest positive period of $z(t)$. ∎

(c) *Open Curves.* Let the function $\zeta : I^\circ = (\alpha, \beta) \to G$ $(-\infty \leq \alpha < \beta \leq \infty)$ be continuous and injective, let $C^\circ = \zeta(I^\circ)$, and assume that the inverse function $\zeta^{-1} : C^\circ \to I^\circ$ is continuous. Then we say that ζ is an open Jordan path and C° is an open Jordan curve.

Remark. If the domain of ζ is not compact, then the inverse function is not continuous in general. However, ζ^{-1} is continuous if one requires, in addition, that there not exist a sequence (t_k) in I° with the properties $\lim t_k = \alpha$ or β and $\lim \zeta(t_k) \in C^\circ$. This assumption is, by the way, necessary and sufficient for the continuity of ζ^{-1}.

Corollary. *Let $\zeta|_{I^\circ}$ be an open Jordan curve, $C^\circ = \zeta(I^\circ) \subset G$, and $(f,g) \neq 0$ in C°. Let the solution $z(t) = (x(t), y(t))$ of system (4), with the maximal interval of existence $J^\circ = (a,b)$, satisfy $z(J^\circ) \subset C^\circ$. Then $z(J^\circ) = C^\circ$. Further, $z = \zeta \circ h$, where $h : J^\circ \to I^\circ$ is continuous and bijective.*

Proof. If $J' \subset J^\circ$ is a compact interval, then $C' = z(J')$ is compact; it follows that $\zeta^{-1}(C')$ is compact and hence is contained in a compact interval $I' \subset I^\circ$. Because $(f,g) \neq 0$, $z|_{I'}$ is a smooth path. Applying (b) to I' and J', we have $z(t) = \zeta(h(t))$ in J', where $h : J' \to I'$ is continuous and (say) strongly monotone increasing. Since h is uniquely determined and J' is arbitrary, one obtains $z(t) = \zeta(h(t))$ in J° with $h : J^\circ \to I^\circ$ continuous, strongly increasing.

It remains to show that $h(J^\circ) = I^\circ = (\alpha, \beta)$; the remaining statements of the corollary follow then without difficulty. We confine ourselves to the proof of $\lim_{t \to b} h(t) = \beta$. Suppose $\lim_{t \to b} h(t) = \gamma < \beta$. From this it follows first that $\lim_{t \to b} z(t) = \zeta(\gamma) \in C^\circ$. The assumption $b < \infty$ contradicts the maximality of b. From $b = \infty$ it follows as above that $\zeta(\gamma)$ is a critical point of (4). This contradiction shows that $\gamma = b$. One shows correspondingly that $h(t) \to \alpha$ as $t \to a$. ∎

Remarks. *1.* This result can be applied to the case where $\zeta|_{[\alpha,\beta]}$ is a path connecting two stationary points of (f,g) that does not contain any other stationary points. As $t \to \alpha$ or β, the solution tends toward the corresponding stationary endpoint. The behavior is similar if $|\zeta(t)|$ tends to ∞ as $t \to \alpha$ or β. In this case the solution $z(t)$ also traces out the whole curve C°. Both cases occur in connection with the mathematical pendulum in 11.X.(d).

2. The above theorem and its corollary answer the question raised in 3.V.(d), while the question 3.V.(c) about the nature of level curves has already been answered by Theorems VII and VIII.

3. The theorem and corollary remain true for autonomous systems in $\mathbb{R}^n$; the proofs carry over as well.

B. Real Analysis

We first prove some theorems on Dini derivatives and convex functions as they relate to differential inequalities and as a particular example obtain some important norm estimates. Then we give a proof of the Brouwer fixed point theorem. This fundamental theorem is used in D.XII to derive the Schauder fixed point theorem.

In this section, $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^n$, and points in $\mathbb{R}^n$ will not be represented in boldface type.

I. Dini Derivatives. For a function $u : J \to \mathbb{R}$ (J an interval) the upper and lower right-sided *Dini derivatives* are defined by

$$D^+u(t) = \limsup_{s\to t+} Q(s,t), \quad D_+(t) = \liminf_{s\to t+} Q(s,t),$$

where $Q(s,t) = [u(s) - u(t)]/(s-t) = Q(t,s)$ is the difference quotient of u. For the corresponding left-sided Dini derivatives D^-, D_-, the limit $s \to t+$ is replaced with $s \to t-$. These definitions also appear in the text in 9.I.

In the following,$D \in \{D^+, D_+, D^-, D_-\}$ represents an arbitrarily chosen Dini derivative. However, within a single theorem or formula, D is fixed.

(a) *A right-sided Dini derivative satisfies $D(u+v)(t) = Du(t) + v'_+(t)$, provided that the right-sided derivative of v exists (with finite value). A corresponding statement holds for the left-sided derivatives.*

The proof of this result follows immediately from the formula

$$\limsup(a_n + b_n) = \limsup a_n + \lim b_n$$

and a corresponding formula for lim inf, where (b_n) is a convergent sequence.

A differentiable function with a nonnegative derivative is increasing; this is an immediate consequence of the mean value theorem. The following important theorem gives a far-reaching generalization of this statment.

Theorem. *If $u \in C(J)$ and $Du \geq 0$ in $J \setminus N$, where N is at most countable, then u is monotone increasing in J.*

The *proof* prodeeds in two steps. In the first step we assume strict inequalities, $Du > 0$ in $J \setminus N$, and show that u is increasing. If not, then there are points $a < b$ in J with $u(a) > u(b)$. We choose a number α with the properties $u(a) > \alpha > u(b)$ and $\alpha \notin u(N)$ (this is possible because the set $u(N)$ is countable). Let $c \in (a, b)$ be the largest point with $u(t) = \alpha$, i.e., $u(c) = \alpha$ and $u(t) < \alpha$ in $(c, b]$. Since the difference quotients $Q(c, t)$ are negative for $t > c$, we obtain $D^+u(c) \leq 0$. On the other hand, $c \notin N$, and therefore $Du(c) > 0$. This is a contradiction if D is D^+ or D_+. If D is a left-sided derivative, one takes c as the smallest point with $u(t) = \alpha$ and considers quotients $Q(c, t)$ with $t < c$. Hence u is increasing in all cases.

In the second step, we have the assumption of the theorem $Du \geq 0$ in $J \setminus N$ and consider the functions $u_\varepsilon(t) = u(t) + \varepsilon t$ $(\varepsilon > 0)$. It follows from (a) that $Du_\varepsilon \geq \varepsilon > 0$ in $J \setminus N$ and hence from the first part that u_ε is increasing. The theorem is now obtained by taking the limit as $\varepsilon \to 0$. ∎

As a simple consequence, one obtains the following

Generalized Mean Value Theorem. *Let I, J be intervals, $N \subset J$ countable, and $u \in C(J)$. Then*

$$Du(t) \in I \quad \text{for} \quad t \in J \setminus N \Longrightarrow Q(s, t) \in I \quad \text{for} \quad s, t \in J, \ s \neq t.$$

Hint for the proof: Apply the theorem to $\pm u(t) + \lambda$.

We draw another consequence, which allows a surprising application.

II. Theorem. *Let the functions $u, h \in C(J)$ satisfy $D^*u(t) \geq h(t)$ in $J \setminus N$, where D^* denotes a Dini derivative and N an at most countable set. Then $Du(t) \geq h(t)$ for every $t \in J$ and every Dini derivative.*

Proof. Let H be an antiderivative of h, that is, $H' = h$ in J. The function $v(t) = u(t) - H(t)$ satisfies $D^*v(t) = D^*u(t) - h(t) \geq 0$ in $J \setminus N$ by I.(a). By Theorem I, v is monotone increasing; hence $Dv(t) \geq 0$ for all $t \in J$ and every D. This inequality is, again by I.(a), equivalent to the conclusion. ∎

Application. If u has one-sided derivatives and if $u'_+ \leq f(t, u)$ in J, where f is continuous, then by Theorem II, $u'_- \leq f(t, u)$ in J. The formulation of many theorems found in the literature indicates that this result is not generally known.

III. Convex Functions. The function $u : J \to \mathbb{R}$ is called *convex* if

$$u(\lambda a + (1 - \lambda)b) \leq \lambda u(a) + (1 - \lambda)u(b) \quad \text{for} \quad 0 < \lambda < 1 \tag{1}$$

and $a, b \in J$. Elementary properties of convex functions, in particular, the existence of the one-sided derivatives and the inequalities $u'_-(s) \leq u'_+(s) \leq$

$u'_-(t) \le u'_+(t)$ for $s < t$, from which the differentiability of u in $J \setminus N$ (N countable) follows, can be found in standard reference works in analysis.

Lemma. *Let X be a Banach space with the norm $\|\cdot\|$ and $x, y \in X$. Then the function $p(t) = \|x + ty\|$ is convex and*

$$-\|y\| \le p'_-(t) \le p'_+(t) \le \|y\| \quad \text{for} \quad t \in \mathbb{R}.$$

To prove the convexity, we apply the triangle inequality and obtain, with the notation $\mu = 1 - \lambda$,

$$\begin{aligned} p(\lambda a + \mu b) &= \|x + (\lambda a + \mu b)y\| \\ &= \|\lambda(x + ay) + \mu(x + by)\| \le \lambda p(a) + \mu p(b), \end{aligned}$$

hence (1). In a similar manner the triangle inequality yields the estimate

$$|p(t+h) - p(t)| \le \|hy\|,$$

from which the second assertion follows after dividing by h and passing to the limit (the existence of one-sided derivatives is guaranteed by convexity). ∎

In going from a smooth vector function $u(t)$ to its norm $|u(t)|$, one generally loses differentiability. However the following theorem shows that one-sided differentiability is retained and that this is true even in the general case of functions with values in a Banach space. Here the derivative of a function $u : J \to X$ (X a Banach space) is defined as in the classical case: $u'(t) = \lim\limits_{h\to 0}[u(t+h) - u(t)]/h \in X$ (the limit is taken with respect to the norm of X).

IV. Theorem. *Let X be a Banach space with the norm $\|\cdot\|$. If $u : J \to X$ is a continuous function, then the function $\phi(t) := \|u(t)\|$ is also continuous. Moreover, if u has a right-sided derivative at the point t, then the same holds for ϕ, and the inequality*

$$-\|u'_+(t)\| \le \phi'_+(t) \le \|u'_+(t)\|$$

holds; a corresponding result is true for left-sided derivatives. In particular, if u is differentiable at the point t, then

$$-\|u'(t)\| \le \phi'_-(t) \le \phi_+(t) \le \|u'(t)\|.$$

Proof. From the triangle inequality we get $|\phi(s) - \phi(t)| \le \|u(s) - u(t)\|$ and from here the continuity of ϕ. Now let u be differentiable to the right at t, that is,

$$u(t+h) = u(t) + hu'_+(t) + he(h) \quad \text{with} \quad \lim_{h\to 0+} \|e(h)\| = 0.$$

It follows from the triangle inequality in the form (5.2) that

$$|\phi(t+h) - \|u(t) + hu'_+(t)\|| \le h\|\mathrm{e}(h)\|$$

and hence

$$\phi(t+h) = \|u(t) + hu'_+(t)\| + h\delta(h) \quad \text{with} \quad \lim_{h\to 0+} \delta(h) = 0.$$

Using the notation $x = u(t)$, $y = u'_+(t)$, $p(h) = \|x + hy\|$, we have

$$\phi(t+h) - \phi(t) = p(h) - p(0) + h\delta(h).$$

The conclusion now follows from Lemma III after dividing by h. ∎

We come now to the fixed point theorem discovered in 1912 by the Dutch mathematician Luitzen Egbertus Jan Brouwer (1881–1966).

V. Brouwer's Fixed Point Theorem. *Let B be the closed unit ball in $\mathbb{R}^n$ and $f : B \to B$ continuous. Then f has at least one fixed point.*

Note that since $\mathbb{C}^n$ is norm-isomorphic to $\mathbb{R}^{2n}$, the theorem also holds for the closed unit ball in $\mathbb{C}^n$.

Proof. First a preliminary remark. By applying the Weierstrass approximation theorem to each of the n components of f, one can show that for every $\varepsilon > 0$ there exists a (vector) polynomial P with $\|f - P\| < \varepsilon$.

Here and in the results that follow the maximum norm in B, $\|f\| = \max\{|f(x)| : x \in B\}$ is used. Then $\|P\| \le 1 + \varepsilon$; hence $Q(x) = P(x)/(1+\varepsilon)$ is a smooth mapping of B into B. It is easy to show that $\|f - Q\| < 2\varepsilon$. Now assume that x is a fixed point of Q. Then it follows that it is an "approximate fixed point" of f satisfying $|x - f(x)| = |Q(x) - f(x)| < 2\varepsilon$. Thus by Theorem 7.X, f has a fixed point if every smooth mapping of B into B has a fixed point.

We now prove the theorem in a sequence of steps under the assumption that $f \in C^1(B)$.

(a) Let

$$P(\lambda) = a\lambda^2 + 2b\lambda + c \quad \text{with} \quad a > 0$$

be a real quadratic polynomial with the property $P(0) \le 1$, $P(1) \le 1$. Since P is convex, there are exactly two values λ_1 and λ_2 such that $P(\lambda_1) = P(\lambda_2) = 1$; moreover,

$$\lambda_1 \le 0 < 1 \le \lambda_2 \quad \text{and} \quad P(\lambda) < 1 \quad \text{for} \quad \lambda_1 < \lambda < \lambda_2.$$

Thus $\lambda_{1,2} = A \pm \sqrt{C}$ with $A = -b/a$, $C = (b/a)^2 + (1-c)/a \ge 1/4$, the latter because $\lambda_2 - \lambda_1 \ge 1$.

(b) Suppose f does not have a fixed point. Then the continuous function $|f(x) - x|$ is positive. Since B is compact, there exists a $\gamma > 0$ with $|f(x) - x| \ge \gamma$ in B. For every $x \in B$, the quadratic polynomial

$$P(\lambda) = |x + \lambda(f(x) - x)|^2$$

satisfies the properties from (a): $P(0) = c = |x|^2 \leq 1$, $P(1) = |f(x)|^2 \leq 1$, $a = |f(x) - x|^2 \geq \gamma^2$, and $b = (x, f(x) - x)$. The function $\lambda_1 = \lambda_1(x)$ is ≤ 0 and belongs to $C^1(B)$, as one can read from the representation $\lambda_1 = A - \sqrt{C}$ in (a).

(c) We define C^1-functions $g(x) := \lambda_1(x)(f(x) - x)$ and

$$h(t, x) = x + tg(x) \quad \text{for} \quad 0 \leq t \leq 1$$

and consider the integral

$$V(t) = \int_B \det \frac{\partial h(t,x)}{\partial x}\, dx = \int_B \det \left(I + t\frac{\partial g(x)}{\partial x} \right) dx.$$

Here $\partial h/\partial x$ and $\partial g/\partial x$ are the $n \times n$ Jacobian matrices of h and g.

The proof of the theorem by contradiction proceeds as follows. We show, in order, that (i) $V(0) = |B| = \Omega_n$ (volume of the unit ball), (ii) $V(1) = 0$, and (iii) $V(t) = \text{const}$.

Assertion (i) follows immediately from the definition of V. To prove (ii) one first notes that $|h(1,x)|^2 = |x + \lambda_1(x)(f(x) - x)|^2 = P(\lambda_1) = 1$. Therefore $h(1, \cdot)$ maps B onto ∂B. Hence for $x \in B^\circ$ the matrix $\partial h(1,x)/\partial x$ is singular, since otherwise a neighborhood of x would be mapped bijectively by $h(1, \cdot)$ onto a neighborhood of $h(1,x)$ by the Inverse function theorem. Thus for $x \in B^\circ$, $\det \partial h(1,x)/\partial x = 0$, from which (ii) follows.

The proof of (iii) begins with the observation that the C^1-function g satisfies a Lipschitz condition

$$|g(x) - g(x')| \leq L|x - x'| \quad \text{in} \quad B.$$

Further, $g(x) = 0$ for $x \in \partial B$, since in this case $P(0) = |x|^2 = 1$ and hence $\lambda_1(x) = 0$. Let Q denote the projection onto the unit ball

$$Qx = x \quad \text{for} \quad |x| \leq 1 \quad \text{and} \quad Qx = \frac{x}{|x|} \quad \text{for} \quad |x| > 1.$$

It is easy to show $|Qx - Qx'| \leq |x - x'|$. Therefore, the function $\bar{g}(x) := g(Qx)$ satisfies a Lipschitz condition in $\mathbb{R}^n$ with the same constant L ($\bar{g}$ is simply the extension of g to $\mathbb{R}^n$ by 0 outside B). We show:

(d) For $0 \leq t < 1/L$, the mapping $h(t, \cdot)$ is a bijection of B onto B.

To prove this, let $\bar{h}(t,x) = x + t\bar{g}(x)$ and let $a \in \mathbb{R}^n$ be arbitrary. The equation $\bar{h}(t,x) = a$ is equivalent to

$$x = a - t\bar{g}(x).$$

Since the right side of this equation is a contraction with Lipschitz constant $tL < 1$, there exists exactly one $x = x_a$ with $\bar{h}(t, x_a) = a$. Thus the function $\bar{h}(t, \cdot)$ maps $\mathbb{R}^n$ bijectively onto itself. Now, however, $\bar{h}(t, \cdot)$ is the identity mapping on $\mathbb{R}^n \setminus B$ and equal to $h(t, \cdot)$ on B. Therefore, $h(t, \cdot)$ is a bijection of B onto B.

(e) From the substitution rule for n-dimensional integrals it follows that $V(t) = \text{const} = \Omega_n$, at least as long as $h(t, \cdot)$ is a bijection $B \to B$ and $\det \partial h(t,x)/\partial x > 0$. Hence there is an interval $0 \leq t \leq \varepsilon < 1/L$ where $V(t)$

is constant. However, since $V(t)$ is a polynomial in t of degree $\leq n$, it follows that $V(t) \equiv \Omega_n$ for $0 \leq t \leq 1$. This completes the proof by contradiction of the Brouwer fixed point theorem. ∎

In the following corollaries we make use of some abbreviated terminology. We say that a subset A of a Banach space has the *fixed point property* if every continuous mapping of A into A has a fixed point. The Brouwer fixed point theorem can then be rephrased: *The closed unit ball in* $\mathbb{R}^n$ *has the fixed point property.*

Let X, Y be Banach spaces or more general topological spaces. Two sets $A \subset X$ and $B \subset Y$ are said to be *homeomorphic* if there exists a homeomorphism $h : A \to B$ (a bijective mapping that, along with its inverse, is continuous).

Corollary 1. *If the sets A and B are homeomorphic and if A has the fixed point property, then B also has the fixed point property.*

The proof is very simple. Let $h : A \to B$ be a homeomorphism and $f : B \to B$ a continuous mapping. Then $F = h^{-1} \circ f \circ h$ is a continuous mapping of A to itself. If x is a fixed point of F, then the image point $\xi = h(x)$ is a fixed point of f, as one easily verifies. ∎

Corollary 2. *Let the set $A \subset \mathbb{R}^n$ be compact, and let there exist a continuous mapping $P : \mathbb{R}^n \to A$ with $P|_A = \mathrm{id}_A$, i.e., $P(x) = x$ for $x \in A$. Then A has the fixed point property.*

For the proof let $B \supset A$ be a closed ball and $f : A \to A$ continuous. Then $F = f \circ P$ is a continuous mapping of B into itself. By the Brouwer fixed point theorem, F has a fixed point ξ, and because $F(B) \subset A$, this fixed point belongs to A, whence $\xi = f(\xi)$, i.e., ξ is a fixed point of f. ∎

Corollary 3. *A nonempty, convex, and compact set $A \subset \mathbb{R}^n$ has the fixed point property.*

Proof. For every $x \in \mathbb{R}^n$ there exists, since A is convex and compact, exactly one "closest point" $y = Px \in A$ with $\operatorname{dist}(x, A) = |x - y|$. The mapping P, also called the (metric) projection on A, is continuous, and the assertion follows from Corollary 2. ∎

Remark. A property of sets that is carried over to the image set by a homeomorphism is also called a topological property. For example, openness and compactness are topological properties. Corollary 1 shows that the fixed point property is also a topological property.

C. Complex Analysis

Here we give some tools from complex analysis that are used when working with ordinary differential equations in the complex domain. We are dealing primarily with the Banach space of holomorphic functions introduced in Example 5.III.(d) and with the properties of holomorphic functions described in 8.I. In the following, G is a region in the complex plane.

I. Holomorphic Functions. A function $f : G \to \mathbb{C}$ is called *holomorphic* in G, written $f \in H(G)$ (21.I), if f is continuously differentiable (in the complex sense) in G. In most textbooks on complex analysis, only differentiability is required; the continuity of the derivative then follows as a theorem. The next two theorems form the foundation of the Cauchy function theory. We formulate them only in the generality that is necessary for our purposes.

II. Cauchy's Integral Theorem. *If G is simply connected, $f \in H(G)$, and $\zeta|_I$ is a piecewise continuously differentiable closed path in G with $I = [a, b]$, then*

$$\int_\zeta f(z)\,dz = \int_a^b f(\zeta(t))\zeta'(t)\,dt = 0.$$

The integral ($z_0 \in G$ is fixed)

$$F(z) = \int_{z_0}^z f(z')\,dz'$$

is independent of path (i.e., for each path $\zeta|_I$ in G with $\zeta(a) = z_0$, $\zeta(b) = z$, the integral has the same value), and F is an antiderivative of f. The latter means that $F \in H(G)$ and $F' = f$ in G.

III. Cauchy's Integral Formula for the Disk. *If the disk $B : |z - z_0| < r$, together with its boundary ∂B, lies in G, then for $f \in H(G)$,*

$$f(z) = \frac{1}{2\pi i}\int_{\partial B} \frac{f(\zeta)}{\zeta - z}\,d\zeta \quad \text{for} \quad z \in B. \tag{1}$$

Here the boundary of the disk ∂B is oriented in the positive direction.

Proofs for these two theorems can be found in any textbook on complex analysis.

IV. Applications of the Cauchy Integral Formula. We begin by expanding the factor $1/(\zeta - z)$ that appears in (1) in a geometric series,

$$\frac{1}{\zeta - z} = \sum_{n=0}^\infty \frac{(z - z_0)^n}{(\zeta - z_0)^{n+1}} \quad \text{for} \quad |z - z_0| < |\zeta - z_0| = r.$$

This series is uniformly convergent in every concentric disk $|z - z_0| \leq \rho \leq r$. Therefore, (1) can be integrated termwise and one obtains in this manner the following propositions.

(a) *Under the assumptions from* III, f *has a power series expansion*

$$f(z) = \sum_{n=0}^{\infty} c_n (z - z_0)^n \quad \text{for} \quad |z - z_0| < r. \tag{2}$$

(b) *Let* (f_n) *be a sequence from* $H(G)$ *that converges locally uniformly in* G. *Then the limit function* $f(z) = \lim f_n(z)$ *belongs to* $H(G)$.

The hypothesis means that the sequence is uniformly convergent in every closed disk $\bar{B} \subset G$. Thus the limit function is continuous in G. If f_n is written in the form (1), then the limit as $n \to \infty$ can be taken under the integral sign, i.e., the limit function f also satisfies equation (1). By (a), f can be expanded in a power series of the form (2). Therefore, f is holomorphic in B. Since B is arbitrary, f is holomorphic in all of G.

The next theorem is a simple consequence of (a).

V. Theorem. *If* $f \in H(G)$ *and if the disk* $B : |z - z_0| < r$ *lies in* G, *then* f *has a expansion* (2) *that is valid at least in* B, *which implies that the radius of convergence of this power series is* $\geq r_0 = \operatorname{dist}(z_0, \partial G)$ $(r_0 = \infty$ *if* $G = \mathbb{C})$.

VI. Theorem. *Let the function* $p : G \to \mathbb{R}$ *be continuous and positive. Then the set* $H_p(G)$ *of all functions* $f \in H(G)$ *such that* $\sup_G p(z)|f(z)| < \infty$ *with the norm defined by*

$$\|f\| = \sup\{p(z)|f(z)| : z \in G\}$$

is a Banach space.

Proof. We refer to 5.III for the proof that $H(G)$ is a normed space. To prove completeness of this space, we note first that for each disk B with $\bar{B} \subset G$ there exist positive constants α, β such that $0 < \alpha \leq p(z) \leq \beta$ in $\bar{B}$. Thus the inequalities $|f(z)| \leq \|f\|/p(z) \leq \|f\|/\alpha$ hold in $\bar{B}$. From this it follows that a Cauchy sequence (f_n) with respect to the norm $\|\cdot\|$ converges uniformly in $\bar{B}$, hence converges locally uniformly in G. The limit function f is then holomorphic in G by IV.(b). A Cauchy sequence is bounded, say, $\|f_n\| \leq C$ for all n. Since $p(z)|f_n(z)| \leq C$ in G, the limit $p(z)|f(z)|$ is also bounded by C, which implies that $f \in H_p(G)$. From the inequality $\|f_n - f_{n+k}\| < \varepsilon$ for $n \geq n_0$, $k \geq 1$ we obtain $\|f_n - f\| \leq \varepsilon$ by letting $k \to \infty$. Thus $\lim \|f_n - f\| = 0$. ∎

One sees from the proof that the continuity of p is not important. What is needed is an estimate of the form $0 < \alpha \leq p(z) \leq \beta$ in each closed disk $\bar{B} \subset G$ (α, β depend on B).

Theorem V is used, in particular, in the proof of Theorem D.VI. Banach spaces of holomorphic functions of the kind described in Theorem VI arise in the existence theorems in 8.II, 10.X, and 21.II.

D. Functional Analysis

We bring together here some ideas and theorems from functional analysis that play a role in the subject matter of this text. The topics in this section are selected to convey a deeper insight into functional-analytic methods of proof related to the contraction principle. The section concludes with a proof of the Schauder fixed point theorem.

In the following, X generally stands for a real or complex Banach space with the norm $\|\cdot\|$.

I. Convergence of Series. The convergence of a series $\sum x_n$ with $x_n \in X$ is defined in 5.IV. Such a series is said to be *absolutely convergent* if $\sum \|x_n\| < \infty$. As in the real case, the following hold:

(a) If $\sum x_n$ is absolutely convergent, then $\sum x_n$ is convergent and $\|\sum x_n\| \le \sum \|x_n\|$.

The comparison criterion remains valid. It states that $\|x_n\| \le \alpha_n$, where $\sum \alpha_n < \infty$ implies that the series $\sum x_n$ is absolutely convergent. The classical proof for $x_n \in \mathbb{R}$ carries over. Comparison with the geometric series leads to the

(b) *Root Criterion.* The series is absolutely convergent if $\|x_n\|^{1/n} \le q < 1$ for n sufficiently large and divergent if $\|x_n\|^{1/n} \ge 1$ for infinitely many n.

II. Equivalent and Monotone Norms. Two norms $\|\cdot\|$, $\|\cdot\|'$ in a vector space X are called *equivalent* if there exist positive constants α, β such that

$$\alpha \le \frac{\|x\|}{\|x\|'} \le \beta \quad \text{for} \quad x \ne 0. \tag{1}$$

In this case we write $\|\cdot\| \sim \|\cdot\|'$. This relation is an equivalence relation in the set of all norms defined in X; i.e., it is reflexive ($\|\cdot\| \sim \|\cdot\|$), symmetric ($\|\cdot\| \sim \|\cdot\|'$ implies $\|\cdot\|' \sim \|\cdot\|$) and transitive ($\|\cdot\| \sim \|\cdot\|'$ and $\|\cdot\|' \sim \|\cdot\|''$ implies $\|\cdot\| \sim \|\cdot\|''$).

Let X be a normed space of functions $f : D \to Y$, where D is a nonempty set and Y is the space $\mathbb{R}^n$ or $\mathbb{C}^n$ (or some other Banach space) with the norm $|\cdot|$. The norm $\|\cdot\|$ in X is called monotone if for f, $g \in X$

$$|f(x)| \le |g(x)| \text{ for } x \in D \text{ implies that } \|f\| \le \|g\|.$$

All of the norms that appear in the examples from 5.III are monotone.

III. Bounded Linear Operators. The set $\mathcal{L}(X)$ of all linear mappings $A : X \to X$ with finite,

$$\textit{operator norm} \quad \|A\| := \sum\{\|Ax\| : \|x\| = 1\} = \sup\left\{\frac{\|Ax\|}{\|x\|} : x \ne 0\right\}$$

is itself a Banach space. The mapping A is Lipschitz continuous, $\|Ax - Ay\| \le L\|y - x\|$, and $\|A\|$ is the smallest Lipschitz constant for A.

(a) For $A, B \in \mathcal{L}(X)$ we have $\|AB\| \leq \|A\|\|B\|$, in particular $\|A^n\| \leq \|A\|^n$.

(b) If two norms $\|\cdot\|$, $\|\cdot\|'$ in X are equivalent, then the corresponding operator norms are equivalent in $\mathcal{L}(X)$. It follows from (1) that

$$\frac{\alpha}{\beta} \leq \frac{\|A\|}{\|A\|'} \leq \frac{\beta}{\alpha} \quad \text{for} \quad A \neq 0. \tag{2}$$

IV. The Spectral Radius. The spectral radius $r(A)$ of $A \in \mathcal{L}(X)$ is defined by

$$r(A) := \lim_{n\to\infty} \|A^n\|^{1/n} = \inf_{n>0} \|A^n\|^{1/n}.$$

The existence of the limit and the equality of the two expressions are proved by first using III.(a) to establish the inequality $\alpha_{m+n} \leq \alpha_m \alpha_n$ for $\alpha_n = \|A^n\|$, from which if $\beta_n = \ln \alpha_n$, the inequality $(\star)$ $\beta_{m+n} \leq \beta_m + \beta_n$ follows. Then one applies the following theorem from real analysis: From $(\star)$ it follows that $\lim \beta_n/n = \inf \beta_n/n$. A proof for this can be found in Pólya–Szegő (1970, Chap. I. Problem 98).

The operator A is called *nilpotent* if $A^p = 0$ for some power p and *quasinilpotent* if $r(A) = 0$.

From II.(b), we obtain in a simple manner the following:

(a) For two equivalent norms, the spectral radius has the same value. In other words, the spectral radius does not change when the norm is changed to an equivalent norm.

V. Power Series. Let $f(s) = \sum_0^\infty c_n s^n$ be a real or complex power series with positive radius of convergence r. For $A \in \mathcal{L}(X)$, $f(A)$ is defined to be the series $\sum_0^\infty c_n A^n$. Here $A^0 = I$, the identity mapping.

Theorem. *If the series for f has the radius of convergence $r > 0$, then the series*

$$f(a) = \sum_{n=0}^\infty c_n A^n \quad \text{with} \quad A \in \mathcal{L}(X)$$

is absolutely convergent if $r(A) < r$ and divergent if $r(A) > r$.

The proof follows immediately from the root criterion I.(b). Let $r(A) < s < r$. For large n we have $\|A^n\|^{1/n} \leq s$; hence $\|c_n A^n\| \leq c_n s^n$ and $\sum |c_n| s^n < \infty$.

On the other hand, it follows from $r < s < r(A)$ that $\|A^n\|^{1/n} > s$ and $|c_n| s^n \geq 1$ for infinitely many n (Cauchy–Hadamard formula for the radius of convergence of a power series); hence $|c_n| \|A^n\| \geq 1$ for these n. ∎

(a) The series obtained, when this process is applied to the geometric series $(1-s)^{-1} = \sum s^n$ with $A \in \mathcal{L}(X)$, is called the *Neumann series*. If $r(A) < 1$, then

$$(I-A)^{-1} = \sum_{n=0}^{\infty} A^n \in \mathcal{L}(X). \tag{3}$$

The theorem guarantees the absolute convergence of the series, and we have $(I-A)\sum A^n = \sum A^n - \sum A^{n+1} = I$, which shows that $\sum A^n$ is indeed the inverse of $I - A$ (the index n runs from 0 to ∞).

As a first application of the Neumann series we show that the spectral radius is the radius of the smallest circle about the origin that contains the spectrum (this is the property that gives it its name). We confine ourselves here to the case $X = \mathbb{C}^n$.

VI. Spectrum and Spectral Radius in the Finite Dimensional Case. Let A be a complex $n \times n$ matrix, which we identify with the linear mapping in $\mathbb{C}^n$ generated by A. The spectrum $\sigma(A)$ of A is the set of eigenvalues of A, that is, the set of all $\lambda \in \mathbb{C}$ with the property that $\det(A - \lambda I) = 0$.

Theorem. *For $A \in \mathcal{L}(\mathbb{C}^n)$, $r(A) = \max\{|\lambda| : \lambda \in \sigma(A)\}$.*

Proof. Let $\rho = \max\{|\lambda| : \lambda \in \sigma(A)\}$, let λ be an eigenvalue with $|\lambda| = \rho$, and let x be a corresponding eigenvector. From $Ax = \lambda x$ it follows that $A^n x = \lambda^n x$ and hence $\|A^n\| \geq |\lambda|^n = \rho^n$. This proves the inequality $\rho \leq r(A)$.

The matrix $A - \lambda I$ is invertible for $|\lambda| > \rho$. Thus the matrix $R(z) = (zA - I)^{-1}$ exists for $|z| < 1/\rho$ (if $\rho = 0$, this is true for all $z \in \mathbb{C}$). By Cramer's rule $R(z)$ is a rational function of the form $Q(z)/P_n(z)$ (P_n, Q polynomials), where $P_n(z) = \det(zA - I)$, which has no zeros in the disk $|z| < 1/\rho$. By Theorem V, the Neumann series

$$R(z) = (zA - I)^{-1} = -\sum_{n=0}^{\infty} z^n A^n$$

diverges for $r(zA) = |z| r(A) > 1$, and thus its radius of convergence is $\leq 1/r(A)$. On the other hand, $R(z)$ is a holomorphic function for $|z| < 1/\rho$, and from Theorem C.V it follows that the radius of convergence of the series (we are dealing here with n^2 scalar series) is $\geq 1/\rho$. Thus the reverse inequality $\rho \geq r(A)$ holds. ∎

The next theorem is of fundamental importance for the application of the contraction principle.

VII. Theorem. *Let $A \in \mathcal{L}(X)$ and $r(A) < \alpha$. Then there exists an equivalent norm $\|\cdot\|'$ (in a Hilbert space, an equivalent Hilbert norm $\|\cdot\|'$) with the following properties:*

(a) $\|A\|' \leq \alpha$.

(b) *If B commutes with A ($AB = BA$), then $\|B\|' \leq \|B\|$.*

Proof. We choose an n such that $\|A^n\|^{1/n} < \alpha$ and define

$$\|x\|' = \|x\| + \frac{\|Ax\|}{\alpha} + \frac{\|A^2x\|}{\alpha^2} + \cdots + \frac{\|A^{n-1}x\|}{\alpha^{n-1}}. \tag{4}$$

Clearly, $\|\cdot\|$ is a norm and $\|x\| \le \|x\|' \le K\|x\|$ for a suitable K. Moreover we have

$$\frac{\|Ax\|'}{\alpha} = \frac{\|Ax\|}{\alpha} + \cdots + \frac{\|A^{n-1}x\|}{\alpha^{n-1}} + \frac{\|A^nx\|}{\alpha^n}.$$

Under our assumptions, $\|A^nx\| \le \alpha^n\|x\|$, and therefore $\|Ax\|' \le \alpha\|x\|'$, which implies (a).

In the case of a Hilbert space with inner product $(\cdot,\cdot)$, one uses the inner product

$$(x,y)' = (x,y) + \frac{(Ax,Ay)}{\alpha^2} + \cdots + \frac{(A^{n-1}x, A^{n-1}y)}{\alpha^{2n-2}}. \tag{5}$$

The proof of (a) then runs along similar lines. Part (b) follows from the estimate $\|A^kBx\| \le \|B\|\,\|A^kx\|$ without difficulty. ∎

We draw some additional conclusions.

(c) *The spectral radius $r(A)$ is the infimum of $\|A\|'$, where all norms $\|\cdot\|'$ on X that are equivalent to $\|\cdot\|$ are admitted.*

(d) *If A and B commute and if $r(A) < \alpha$, $r(B) < \beta$, then there exists an equivalent norm $\|\cdot\|'$ with $\|A\|' \le \alpha$ and $\|B\|' \le \beta$.*

To prove this result one applies (4) twice obtaining first $\|A\|' \le \alpha$ and then, beginning again with $\|\cdot\|'$, a norm $\|\cdot\|''$ such that $\|B\|'' \le \beta$. Now (b) shows that $\|A\|'' \le \|A\|' \le \alpha$. ∎

(e) *If A and B commute, then $r(A+B) \le r(A) + r(B)$ and $r(AB) \le r(A)r(B)$.*

(f) *The spectral radius is an upper semicontinuous function; i.e., for $A \in \mathcal{L}(X)$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that $r(A+B) \le r(A) + \varepsilon$ for all $B \in \mathcal{L}(X)$ with $\|B\| < \delta$.*

Both theorems follow immediately from (c) and (d). Thus, for example, $r(A+B) \le \|A+B\|' \le \|A\|' + \|B\|'$ from which one obtains the first inequality in (e) as well as (f). ∎

In conclusion, let it be noted that in place of (4) and (5) one could also use a change of norm of the form

$$\|x\|' = \sum_{n=0}^{\infty} \|A^nx\|\alpha^{-n}, \quad \text{or} \quad (x,y)' = \sum_{n=0}^{\infty} (A^nx, A^ny)\alpha^{-2n}, \tag{4'}$$

respectively. The series are convergent because $r(A/\alpha) < 1$. Again (a) and (b) hold (Exercise!). Part (a) of the theorem was proved in a similar form for a Banach space by R.B. Holmes (1968).

VIII. A Fixed Point Theorem. *Let $A \in \mathcal{L}(X)$ with $r(A) < 1$ and $b \in X$. Then the fixed point equation*

$$x = Ax + b \tag{6}$$

has exactly one solution $x = (I - A)^{-1}b$. The fixed point depends continuously on A and b, i.e., for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for the solution of a nearby equation $y = By + c$ with $\|A - B\| < \delta$, $\|b - c\| < \delta$ the estimate $\|x - y\| < \varepsilon$ holds.

Proof. Clearly, it is sufficient to prove the continuous dependence with respect to an equivalent norm. We choose positive numbers ρ, σ, δ and a norm $\|\cdot\|'$ such that $\|A\|' \le \rho - \delta < \rho < 1$ and $\|b\|' \le \sigma - \delta$. Then for B, c with $\|A - B\|' < \delta$, $\|b - c\|' < \delta$ we have $\|B\|' < \rho$, $\|c\|' < \sigma$, and from $\|y\|' \le \|By\|' + \|c\|' \le \rho\|y\|' + \sigma$ it follows that $\|y\|' \le \sigma/(1-\rho) =: \beta$. From the identity

$$x - y = A(x - y) + (A - B)y + b - c,$$

one obtains the following estimate for the difference $z = x - y$:

$$\|z\|' \le \rho\|z\|' + \delta\beta + \delta \Longrightarrow \|z\|' \le \frac{\delta(1+\beta)}{1-\rho}.$$

One can now make δ smaller in order to get the inequality $\|z\|' < \varepsilon$. ∎

IX. The Integration Operator. Let X be the space of continuous functions $u : J \to Y$, where $J = [a, b]$, $Y = \mathbb{R}^n$ or $\mathbb{C}$, and let $K : X \to X$ be the *integration operator*

$$(Ku)(x) = \int_a^x u(s)\, ds.$$

It has the operator norm $\|K\| = b - a$ with respect to the maximum norm in X (exercise!). However, if one uses the equivalent norm

$$\|u\|_\alpha := \max\{|u(x)|\mathrm{e}^{-\alpha x} : x \in J\} \quad \text{with} \quad \alpha > 0,$$

then $\|K\|_\alpha < 1/\alpha$. Both statements are contained implicitly in the proof of the existence theorem 6.I. For the record:

Theorem. *The integration operator is quasinilpotent in the space $C(J)$; that is, $r(K) = 0$.*

Incidentally, this theorem also follows from the well-known representation of K^n,

$$(K^n u)(x) = \frac{1}{(n-1)!} \int_a^x (x - s)^{n-1} u(s)\, ds,$$

that appears in the remainder term of Taylor's theorem.

Exercise. Calculate the norm of K in the space $L_1(a, b)$. Estimate the norm of K relative to the weighted L_1-norm $\|u\|_\alpha := \int_a^b e^{-\alpha t}|u(t)|\, dt$ and show that K is quasinilpotent in $L_1(a, b)$.

X. The Initial Value Problem. We consider the initial value problem (6.1) (or (10.1–2))

$$y' = f(x, y), \quad y(\xi) = \eta \Longleftrightarrow y(x) = \eta + \int_\xi^x f(t, y(t))\, dt \tag{7}$$

in the interval $J = [\xi, \xi + a]$. Here we take $X = C(J)$.

Using the *Nemytskii operator* $F : X \to X$, defined by

$$(Fu)(x) := f(x, u(x)), \tag{8}$$

we can write the initial value problem in the abbreviated form

$$y = \eta + KFy =: Ty. \tag{7'}$$

Now, if f satisfies a Lipschitz condition $|f(x, y) - f(x, z)| \le L|y - z|$, then $|(Fu)(x) - (Fv)(x)| \le L|u(x) - v(x)|$ clearly holds. Because of the monotonicity of the norm $\|\cdot\|_\alpha$, it follows that $\|Fu - Fv\|_\alpha \le L\|u - v\|_\alpha$. After these preliminary observations, the proof of the existence–uniqueness theorem 10.VI takes only one line

$$\|Tu - Tv\|_\alpha = \|I(Fu - Fv)\|_\alpha \le \|K\|_\alpha \|Fu - Fv\|_\alpha \le \frac{L}{\alpha}\|u - v\|_\alpha, \tag{9}$$

in particular, $\le \frac{1}{2}\|u - v\|_\alpha$ if $\alpha = 2L$.

Remark. An estimation theorem similar to that in 12.V can be obtained from this result together with Theorem VIII.

Our next topic is the fixed point theorem discovered in 1930 by the Polish mathematician Juliusz Pavel Schauder (1899–1943). We derive it using the Brouwer fixed point theorem; cf. B.V. First some preliminaries.

XI. Convex and Compact Sets. Let X be a Banach space with the norm $\|\cdot\|$. A set $A \subset X$ is called *convex* if for arbitrary $a, b \in A$, the connecting segment $\overline{ab} = \{\lambda a + (1 - \lambda)b : 0 \le \lambda \le 1\}$ belongs to A. The *convex hull* $\operatorname{conv} A$ of A is the intersection of all convex supersets of A, hence the smallest convex superset of A.

(a) For a finite set $F = \{x_1, \ldots, x_p\} \subset X$,

$$\operatorname{conv} F = \{\lambda_1 x_1 + \cdots + \lambda_p x_p : \lambda_i \ge 0,\ \lambda_1 + \cdots + \lambda_p x_p = 1\}.$$

Compactness and relative compactness of sets were defined in 7.X. For compact sets the following result is true.

(b) *Borel Covering Theorem.* From a covering of a compact set A by open sets, it is possible to choose finitely many sets that cover A.

Let $U \subset X$ be a finite dimensional subspace of X of dimension p and $\{e_1, \ldots, e_p\}$ a basis for U. The elements of U have a unique representation as linear combinations of the basis elements, which defines a bijective linear mapping L from U onto $\mathbb{R}^p$: The image of $x \in U$ with the representation

$$x = \xi_1 x_1 + \cdots + \xi_p x_p \quad \text{is} \quad L(x) = \xi = (\xi_1, \ldots, \xi_p). \in \mathbb{R}^p$$

If $\xi = L(x)$, then $|\xi|' := \|x\|$ defines a norm on $\mathbb{R}^p$, and L is a norm isomorphism of U to $(\mathbb{R}^p, |\cdot|')$. Since L is linear and all norms in $\mathbb{R}^p$ are equivalent, a set $A \subset U$ is closed, compact, or convex if and only if the image $L(A)$ has these properties.

(c) A closed, bounded subset of a subspace $U \subset X$ with $\dim U < \infty$ is compact. In particular, the convex hull of a finite set $F \subset X$ is compact.

These statements also hold for complex Banach spaces (where then $\xi \in \mathbb{C}^p$). Finally,we recall a definition from 7.X: A mapping $T : D \subset X \to X$ is called compact if the image $T(D)$ is relatively compact.

XII. The Schauder Fixed Point Theorem. *A continuous, compact mapping T of a convex, closed set $D \subset X$ into D has at least one fixed point (X is a Banach space).*

Proof. By the fixed point theorem 7.X, it is sufficient to find, for every $\varepsilon > 0$, a point $x \in D$ with $\|x - Tx\| < \varepsilon$. Thus let $\varepsilon > 0$ be given. The set $B = \overline{T(D)}$ is compact by hypothesis. From the set of all balls $B_\varepsilon(b)$ with $b \in B$ we can select, by the Borel covering theorem XI.(b), a finite number $B_\varepsilon(b_i)$ $(i = 1, \ldots, p)$ that cover B. Let $F = \{b_1, \ldots, b_p\} \subset B$ and $C = \operatorname{conv} F$. By XI.(c) and the convexity of D, the set C is a compact, convex subset of D. Now we define a continuous mapping $\phi : B \to C$ by setting (all sums run from $i = 1$ to $i = p$)

$$\phi(x) = \sum \lambda_i(x) b_i \quad \text{with} \quad \lambda_i(x) = \frac{\mu_i(x)}{\mu(x)}.$$

Here $\mu_i(x) = (\varepsilon - \|x - b_i\|)_+$; that is,

$$\mu_i(x) = 0 \quad \text{if} \quad \|x - b_i\| \geq \varepsilon, \quad \mu_i(x) = \varepsilon - \|x - b_i\| \quad \text{if} \quad \|x_i - b\| < \varepsilon,$$

and $\mu(x) = \sum \mu_i(x)$. Since for every $x \in B$ there exists a b_k with $\|x - b_k\| < \varepsilon$, we have $\mu(x) > 0$ for $x \in B$, and therefore ϕ is continuous. Clearly, $\lambda_i(x) \geq 0$ and $\sum \lambda_i(x) = 1$; hence $\phi(B) \subset C$ by XI.(a). Further, because $x = \sum \lambda_i(x) x$,

$$\|\phi(x) - x\| = \left\| \sum \lambda_i(x)(b_i - x) \right\| \leq \sum \lambda_i(x) \|b_i - x\| < \varepsilon \tag{$\star$}$$

for $x \in B$, since here only summands with $\|b_i - x\| < \varepsilon$ appear (if $\|b_i - x\| \geq \varepsilon$, then $\lambda_i = 0$). The mapping $S = \phi \circ T$ maps D into C; its restriction to C is thus a continuous mapping of C into itself. Since C is convex and compact, there exists, by Corollary 1 and Corollary 3 to the Brouwer fixed point theorem in B.V., a fixed point $x_0 = S(x_0) = \phi(Tx_0) \in C$. From ($\star$) one now obtains

$$\|x_0 - Tx_0\| = \|\phi(Tx_0) - Tx_0\| < \varepsilon,$$

i.e., x_0 is the desired ε-fixed point of T. ∎

Solutions and Hints for Selected Exercises

Exercises in § 1

XII. (a) $y = (x-c)^3$ for $x \le c$, $= 0$ for $c < x < d$, $= (x-d)^3$ for $x \ge d$, assuming $c \le d$. The cases $c = -\infty$, $d = \infty$ are also allowed.

(b) $y = 0$ for $x < c$, $= (x-c)^3$ or $-(x-c)^3$ for $x \ge c$ $(-\infty < c \le \infty)$.

(c) $y = \frac{1}{2}(1 - \cosh(x-c))$ for $x \le c$, $= 0$ for $c < x < d$, $= \frac{1}{2}(1 - \cos(x-d))$ for $d \le x \le d + \pi$, $= 1$ for $x > \pi + d$ $(-\infty \le c \le d \le \infty)$.

(d) For $z = y^2$ one obtains $z' = 2\mathrm{e}^{-z}/(2x + x^2)$, $z(2) = 0$ and hence $z(x) = \ln(1 + \ln(2x/(x+2)))$, $y = \pm\sqrt{z}$ for $x > 2$ (since $z \ge 0$). Here (and occasionally elsewhere) it turns out that f is not defined at the point (ξ, η).

(e) $y = \exp(\mathrm{e} \cdot \sin x/(1 + \cos x))$, $0 < x < \pi$.

(f) $y = \frac{1}{2}\phi^{-1}\left(1 + \frac{1}{2}\pi + 4\sin x\right)$ for $\pi - \alpha < x < 2\pi + \alpha$, where $\alpha = \arctan\left(\frac{1}{8}\pi - \frac{1}{4}\right) \approx 0.1432$ and ϕ^{-1} is the function inverse to $\phi(s) = s + \sin s$. Note that ϕ is strictly increasing in $\mathbb{R}$, but $\phi'((2k+1)\pi) = 0$, i.e., $(\phi^{-1})'((2k+1)\pi) = \infty$ for $k \in \mathbb{N}$. The function y is continuous in $\mathbb{R}$, and the solution exists in $\mathbb{R}$, if we agree to allow the value $y' = \infty$ at the points where $y = \frac{1}{2}\pi$.

(g) $y = x + 2$, $y = x + 4$, $y = x + 3 - \tanh(x - c)$, $y = x + 3 - \coth(x - c)$ $(c \in \mathbb{R})$.

(h) $\int dy(2-y)/(y(y-1)) = 2\int dx/x$ implies $\log(|y-1|/y^2) = \log Cx^2$. By direct calculation one sees that for any solution $y(x)$ the functions $y(-x)$ and $y(\alpha x)$ $(\alpha > 0)$ are also solutions. We choose, e.g., $C = 1/2$ and get $|y-1|/y^2 = x^2/2$. These are two quadratic equations for y, and each has two solutions. Obviously, $y = 0$, $y = 1$, and $y = 2$ are critical values. One obtains with $a = \sqrt{2}/2$,

$$\begin{aligned}
y_1 &= (1 + \sqrt{1 - 2x^2})/x^2 && \text{for} \quad 0 < x < a \quad (y_1 > 2),\\
y_2 &= (1 - \sqrt{1 - 2x^2})/x^2 && \text{for} \quad 0 < x < a \quad (1 < y_2 < 2),\\
y_3 &= -(1 - \sqrt{1 + 2x^2})/x^2 && \text{for} \quad x > 0 \quad (0 < y_3 < 1),\\
y_4 &= -(1 + \sqrt{1 + 2x^2})/x^2 && \text{for} \quad x > 0 \quad (y_4 < 0).
\end{aligned}$$

For $x > 0$ small, $y_1(x) \approx 2/x^2$, $y_2 \approx 1 + \frac{1}{2}x^2$, $y_3(x) \approx 1 - \frac{1}{2}x^2$, $y_4(x) \approx -2/x^2$.

In addition, there are solutions $y \equiv 0$ and $y \equiv 1$. By Theorems VII and VIII, exactly one solution goes through each point (ξ, η) with $\xi \neq 0$, $\eta \neq 2$. On the other hand, from the solutions given above (with x replaced by $\pm\alpha x$) we can find a solution through the point (ξ, η), which shows that all solutions are found.

(i) $y = x \sinh(\ln C|x|)$ $(x \neq 0)$ with $C > 0$.

(j) $y' = 2y/x$.

(k) $y' = 2xy/(x^2+1)$.

(l) $y' = x^3(\operatorname{sgn} y)(\sqrt{1+4|y|/x^4} - 1)$.

Exercises in § 2

V. (a) The solution $y = x$ exists in $\mathbb{R}$. All other solutions

$$y = x + e^{x^2} \Big/ \left(C - \int_0^x e^{t^2}\, dt\right), \quad C \in \mathbb{R},$$

exist in intervals that are bounded on one side.

(b) $y = ce^{\cos x} + 2(\cos x + 1)$ and $y = (c + \sin x - \frac{1}{3}\sin^3 x)/\cos^3 x$ $(c \in \mathbb{R})$.

(c) $y = [(1+1/\eta^3)e^{-3x^5/5} - 1]^{-1/3}$ for $\eta \neq 0$, $y \equiv 0$ for $\eta = 0$. Here $s = t^{-1/3}$ is defined as the inverse function of $t = s^{-3}$, i.e., $t^{-1/3} := (\operatorname{sgn} t)/\sqrt[3]{|t|}$.

VI. Let $F(x) = \int_x^1 f(t)\, dt$. A necessary and sufficient condition for (a) is $F(x) \to \infty$ as $x \to 0+$ and for (b) is $F(x) + \log x \to \infty$ as $x \to 0+$.

The general (positive) solution for the second equation is $y = \exp(ce^{F(x)})$; since $y(1) \leq 1/\mathrm{e}$, only $c \leq -1$ is permitted. Part (a) holds for $F(x) \to \infty$, while $\mathrm{e}^{F(x)} + \log x \to \infty$ as $x \to 0+$ is sufficient for (b).

VII. The equation $u'' - 2u' + 5u = 0$ has $u = a\mathrm{e}^x \cos 2(x - c)$ as its general solution, and from

$$u' = \mathrm{e}^x yu = a\mathrm{e}^x(\cos 2(x-c) - 2\sin 2(x-c))$$

it follows that $y = \mathrm{e}^{-x}(1 - 2\tan 2(x-c))$. The initial value $y(0) = \eta$ is obtained when $c = \frac{1}{2}\arctan(\eta - 1)/2$.

Exercises in § 3

VIII. (a) $F(x,y) = \sin(x+y^2) + 3xy = C$. Since $F_x(0,0) = 1$ and $F_y(0,0) = 0$, for $C = 0$ there exists a solution of the form $x = \phi(y)$ in $U(0,0)$, and $\phi'(0) = 0$ as well as $\phi''(0) = -2$. (Differentiate $F(\phi(y), y) \equiv 0$ twice.) Since $F(0,y) > 0$ for $0 < |y| < \sqrt{\pi}$, the solution remains in the half-plane $x < 0$ for these values of y.

Incidentally, the inequality $-y^2 < \phi(y) < 0$ $(y > 0$ small) is obtained merely by noting that $\operatorname{sgn} \sin(x+y^2) = 1$, since $\operatorname{sgn} xy = -1$ in the second quadrant. Moreover, from $\sin(x+y^2) + 3xy \approx x + y^2 + 3xy = 0$ the approximation $\phi(y) \approx -y^2/(1+3y)$ follows for y close to 0.

Let S_k denote the set $k\pi < x + y^2 < (k+1)\pi$ $(k \in \mathbb{N})$. Then the points $(x, y) \in S_k$ with $xy > 0$ (k even) and the points $(x, y) \in S_k$ with $xy < 0$ (k odd) are forbidden as well as all points with $|xy| > 1/3$ (sketch!). The solution curve for $y > 0$ can be represented graphically in the form $x = \phi(y)$ with $\phi(\sqrt{k\pi}) = 0$, $\phi \to 0$ for $y \to \infty$, and it oscillates around the positive y-axis. Similarly, it oscillates around the negative x-axis for $x \to -\infty$.

(b) $F(x, y) \equiv \frac{1}{2}x^2 - xy - \dfrac{1}{y} = C$. In the form (2c) one gets $y' = y^2(y - x)/(1 - xy^2)$ with the additional solution $y \equiv 0$.

(c) $y = 2x/(C - x^2)$ and $y \equiv 0$.

Exercises in § 4

VIII. (d) $x(p) = \dfrac{2}{(1-p)^2}\left(C - \dfrac{1}{p} - \ln p\right)$, $y(p) = p^2 x(p) + 2\ln p$ $(p > 0)$ as well as $y = x$.

Exercises in § 5

XII. (a) $\|T\|_0 = \frac{1}{2}a^2$, $\|T\|_1 = 1 - (1 - \mathrm{e}^{-a^2})/a^2$, $\|T\|_2 = \frac{1}{2}(1 - \mathrm{e}^{-a^2})$.

(c) One must find a sequence (f_n) of functions $f_n \in C^1(J)$, uniformly convergent in J, but with $\lim f_n(x) = f(x) \notin C^1(J)$; for instance $f_n(x) = |x|^{1+1/n}$ if $0 \in J^\circ$.

Exercises in § 6

IX. If the assertion is false, then there exist two points $(x_n, y_n), (x_n, y_n') \in A$ for each n with

$$|f(x_n, y_n) - f(x_n, y_n')| > n|y_n - y_n'|. \tag{$*$}$$

A subsequence of $((x_n, y_n))$, denoted again by $((x_n, y_n))$, converges; thus $\lim(x_n, y_n) = (x, y) \in A$. If $|f| \le K$ in A, it follows from $(*)$ that $2K > n|y_n - y_n'|$, and thus $\lim(x_n, y_n') = (x, y)$. The function $f(x, y)$ satisfies a Lipschitz condition with respect to y in a neighborhood U of (x, y). On the other hand, (x_n, y_n) and (x_n, y_n') are in U for n large, which contradicts inequality $(*)$.

Exercises in § 7

IX. (a) $\phi(\alpha) = 1 - \sqrt{\alpha_+}$, $\quad \bar{\alpha} = \frac{1}{2}(3 - \sqrt{5}) = 0.382$.

(b) $\psi(\beta) = 1 + \sqrt{\beta_+}$, $\quad \bar{\beta} = \frac{1}{2}(3 + \sqrt{5}) = 2.618$. Since $z' \ge 2x$, one has $z \ge x^2$; hence $z' \ge 2x + 2\sqrt{x^2} = 4x$, i.e., $z \ge 2x^2$. Similarly, one obtains $z_2(x) \ge 2x^2$ independent of z_0. In the differential equation, z_+ can therefore be replaced by $\max\{z, 2x^2\}$, and thus one obtains the condition of Rosenblatt with $k = 1/\sqrt{2}$.

XV. $a_k = 1/k!$. For $\alpha = 0$, the series reduces to $\sum x^k/k! = e^x$, and this is also the solution to the initial value problem. In a bounded interval $0 \le x \le a$ the limit as $\alpha \to 0$ is uniform for each summand (why?), hence also for each finite partial sum. Because of $\{(x - k\alpha)_+\}^k \le a^k$, the remainders become uniformly small in α.

Exercises in § 8

IV. (a) $y = x + x^2 + \frac{1}{6}x^3 - \frac{1}{12}x^4 + \cdots$. For $x, y \in \mathbb{C}$, one has $|\mathrm{e}^x| \le e^{|x|}$ and $|\cos y| \le \cosh |y|$, as seen from the power series. For the cylinder $Z : |x| \le a$, $|y| \le b$ it follows that $|f| \le e^a + a \cosh b =: M$. For example, the choice $a = 1/2$, $b = 2$ yields $b/M = 0.57$ and $\alpha = \min(a, b/M) = 1/2$. Setting $a = 0.53$ and $b = 2$, one gets $\alpha = 0.53$; therefore, the convergence radius of the series is $> 1/2$.

Remark. A better result is obtained with the aid of Theorem 8.V. In the present case, it leads to the estimation $|y(x)| \le \phi(|x|)$ $(x \in \mathbb{C})$, where ϕ is the solution of $\phi'(t) = \mathrm{e}^t + t \cosh \phi$, $\phi(0) = 0$. Using the Lohner algorithm, which gives exact bounds for the solution of an initial value problem, one obtains $\phi(0.8228\ldots) < 28.06$. Hence the power series converges at least for $|z| \le 0.8228$. Details on the Lohner algorithm are described in 9.XVI.

(b) $y = 1 + x + \frac{3}{2}x^2 + \frac{5}{2}x^3 + \frac{37}{8}x^4 + \cdots$. The formulae for calculating the coefficients immediately give $0 \le b_k \le a_k$. From $u(x) = 1/\sqrt{1-2x}$ it follows that $a \le 1/2$.

Exercises in § 9

XI. $y = \eta + x^2$ for $\eta > 0$, $y = \eta - x^2$ for $\eta < 0$ (both unique), $y = cx^2$ with $|c| \le 1$ for $\eta = 0$; hence $y^* = x^2$, $y_* = -x^2$.

XII. (a) In Exercise 8.IV.(b) it was shown that $u = 1/\sqrt{1-2x}$ is a lower solution, and since $u' = u^3 < x^3 + u^3$, this also follows from Theorem VIII. The ansatz $w = 1/\sqrt{1-bx}$ (with $b > 2$) for an upper solution gives the following condition:

$$\frac{b}{2\sqrt{1-bx}^{\,3}} > x^3 + \frac{1}{\sqrt{1-bx}^{\,3}} \iff \frac{b}{2} - 1 > x^3\sqrt{1-bx}^{\,3} \quad \left(0 \le x < \frac{1}{b}\right).$$

The maximum of $x^2(1-bx)$ is attained for $bx = \frac{2}{3}$ and equals $4/(27b^2)$. Therefore the condition is equivalent to

$$\frac{b}{2} - 1 > \frac{1}{3\sqrt{3}}\left(\frac{2}{3b}\right)^2, \text{ or } \frac{81}{8}\sqrt{3}\,b^3\left(\frac{b}{2} - 1\right) > 1.$$

It is satisfied for $b = 2.015$. Hence

$$\frac{1}{\sqrt{1-2x}} < y(x) < \frac{1}{\sqrt{1-bx}} \text{ with } b = 2.015,$$

and $1/b = 0.4963 \le a \le 0.5$.

These surprisingly good bounds were easily obtained. One gets a much more precise estimation by the algorithm by Lohner, mentioned in Exercise XV, which gives exact bounds for the initial value problem. One calculates the solution at some point c where the value of the function is already large, say, $y(c) \in [\underline{y}, \bar{y}]$. One then determines for $x > c$ a lower solution v and an upper solution w, and their asymptotes a_1 and a_0. The latter yield $a_0 \le a \le a_1$. In the present case we use the ansatz

$$\begin{aligned} v(c) = \underline{y},\ v' = v^3 \quad &\Longrightarrow \quad a_1 - c = 1/2\underline{y}^2, \\ w(c) = \bar{y},\ w' = \alpha w^3 \quad &\Longrightarrow \quad a_0 - c = 1/2\alpha\bar{y}^2. \end{aligned}$$

Here $\alpha = 1 + (a_1/\bar{y})^3$, and thus $\alpha w^3 \ge a_1^3 + w^3 \ge x^3 + w^3$. Information concerning the accuracy of the estimation is given by

$$a_1 - a_0 = \frac{1}{2\underline{y}^2} - \frac{1}{2\bar{y}^2} + \frac{1}{2\bar{y}^2}\left(1 - \frac{1}{\alpha}\right) \approx \frac{\bar{y} - \underline{y}}{y(c)^3}.$$

All calculations have to be done by interval arithmetic and intermediate results rounded on the safe side.

One gets, e.g., for $c = 0.49829\ 04344\ 79713$ the following reliable bounds:

$$y(c) \in 3.4005_{04}^{13} \cdot 10^4, \qquad \text{hence } a \in 0.49829\ 04349\ 121_{09}^{12}.$$

(Here $c = 0.49829\ 04344\ 79713$ is the decimal representation of a binary number; the program uses the binary system.)

(b) Obviouly, $y > 1$ for $x > 0$. By use of the estimate

$$y < \sqrt{1 + y^2} < y + \alpha \ \text{ for } \ y > \eta \ \text{ with } \ \alpha = \sqrt{1 + \eta^2} - \eta < 1/2\eta$$

one obtains for $\eta = 1$ a lower solution v and an upper solution w from the linear problems

$$v' = x + v,\ v(0) = 1 \ \text{ and } \ w' = x + \sqrt{2} - 1 + w,\ w(0) = 1.$$

The linear differential equation $u' = x + \alpha + u$ has the solutions $u = \lambda e^x - x - (1 + \alpha)$; hence

$$v = 2e^x - x - 1 < y < w = (1 + \sqrt{2})e^x - x - \sqrt{2} \ \text{ for } \ x > 0.$$

For a better estimation of the order of magnitude of y one calculates the solution y at some c, as in (a). For $c = 10$ one gets

$$y(10) \in [\underline{y}, \bar{y}] = 48180.4_{31}^{51}.$$

The equations $v' = x + v$, $v(10) = \underline{y}$ give a lower solution $v = \underline{\lambda}e^x - x - 1$, where $\underline{\lambda}$ satisfies $\underline{\lambda}e^{10} - 11 = \underline{y}$. Analogously with $w' = x + \alpha + w$, $\alpha = 1/20$, one gets the lower solution $w = \bar{\lambda}e^x - x - 21/20$; again $\bar{\lambda}$ is obtained by using the initial condition $w(10) = \bar{y}$. Therefore,

$$\lambda = \lim_{x \to \infty} e^{-x} y(x) \in [\underline{\lambda}, \bar{\lambda}] = 2.19242_{76}^{85}.$$

Exercises in § 11

IX. (b) $D = 58.469$. (c) $L = 205.237$. (d) $L = 2.371$ m.

XI. (b) (i) periodic; (ii) $x(t) \to -\infty$ for $t \to \pm\infty$; (iii) $x(t) \to \pm\infty$ for $t \to \pm\infty$ if $\eta > 0$ and $x(t) \to \mp\infty$ for $t \to \pm\infty$ if $\eta < 0$.

(c) For $A = B$.

(h) For $h(x) = x^\alpha$ one has $V(r) = \gamma r^\beta$ with $\beta = 1 - \frac{1}{2}(\alpha + 1)$.

(j) $a_1 = \frac{1}{4}$, $a_2 = \frac{9}{64}$.

Exercise in § 15

VI. $Y' = AY$ implies $(Y^*)' = (Y')^* = Y^*A^* = -Y^*A$, since $(AB)^* = B^*A^*$. Hence for $Z(t) = Y(t)Y^*(t)$ one gets

$$Z' = Y'Y^* + Y(Y^*)' = AZ - ZA, \quad Z(\tau) = I.$$

This is a homogeneous linear system of n^2 differential equations for n^2 functions $z_{ij}(t)$. Since $Z(t) = I$ is a solution and the solution is unique, $Z(t) \equiv I$ in J.

Exercises in § 16

IV. One has $c(t) = a(t) + ib(t)$; for $v = z\bar{z}$ it follows that $v' = 2at$. In the example we have $c = \mathrm{e}^{\mathrm{i}t}$, therefore $z' = \mathrm{e}^{\mathrm{i}t}z$ with the solutions $z = c \cdot \exp(-\mathrm{i}\mathrm{e}^{\mathrm{i}t})$. In particular, the solution with $Z(0) = 1$ is

$$z(t) = \exp(\mathrm{i} - \mathrm{i}\mathrm{e}^{it}) = \mathrm{e}^{\sin t}(\cos(1 - \cos t) + \mathrm{i}\sin(1 - \cos t)).$$

For this solution $v(t) = \mathrm{e}^{2\sin t}$, hence $\mathrm{e}^{-2} \le v(t) \le \mathrm{e}^2$. If z is written as a column vector $\begin{pmatrix} x \\ y \end{pmatrix}$, then $X(t) = (z, \mathrm{i}z)$ is a fundamental system with $X(0) = I$ and $\det X(t) = \mathrm{e}^{2\sin t}$.

VI. Since A is periodic, $Y(t + p)$ is also a solution, in fact, a fundamental system (Corollary 15.III). By 15.II.(h) one has $Y(t + p) = Y(t)C$, and $t = 0$ yields $C = Y(0)^{-1}Y(p)$. Hence $Y(t+2p) = Y(t+p)C = Y(t)C^2$, $Y(t+3p) = \cdots$.

(c) follows from a simple calculation. For the proof of (d), let $c \neq 0$ be an eigenvector of C for the eigenvalue λ, thus $Cc = \lambda c$. Then it is easily seen that the solution $y(t) = Y(t)c$ satisfies $y(t + p) = \lambda y(t)$.

Exercises in § 20

VII. The ansatz $x = y = \phi$ (the pendulums swing in phase) and the ansatz $x = -y = \psi$ (the pendulums swing 180° in opposite phase) lead to

$$m\ddot{\phi} = -\alpha\phi \quad \text{und} \quad m\ddot{\psi} = -\alpha\psi - 2k\psi,$$

from which we get four linearly independent solutions,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\beta t, & \sin\beta t, & \cos\gamma t, & \sin\gamma t \\ \cos\beta t, & \sin\beta t, & -\cos\gamma t, & -\sin\gamma t \end{pmatrix}$$

with $\beta = \sqrt{\alpha/m}$, $\gamma = \sqrt{(\alpha + 2k)/m}$. The solution for the pushed pendulum reads

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \beta' \sin \beta t + \gamma' \sin \gamma t \\ \beta' \sin \beta t - \gamma' \sin \gamma t \end{pmatrix} \quad \text{with } \beta' = 1/2\beta,\ \gamma' = 1/2\gamma.$$

Exercises in § 22

VIII. For $w = w_1$ one gets $w'' = \alpha w/z^2$. It follows from the ansatz $w = z^c$ and an additional consideration of the case $\alpha = -1/4$ that

$$w = z^{c_1} \ \text{ and } \ w = z^{c_2} \ \text{ with } \ c_{1,2} = \tfrac{1}{2} \pm \sqrt{\alpha + \tfrac{1}{4}} \ \text{ for } \ \alpha \neq -\tfrac{1}{4},$$

$$w = z^{1/2} \ \text{ and } \ w = z^{1/2} \log z \ \text{ for } \ \alpha = -\tfrac{1}{4}.$$

For $\alpha = n(n-1)$ $(n = 1, 2, 3, \ldots)$ all solutions are rational functions.

Exercises in § 25

XII. $z^2u'' + (3z+1)u' + u = (z^2u' + (z+1)u)' = 0$. The point $z = 0$ is strongly singular; the point $z = \infty$ weakly singular. The equation $z^2u' + (z+1)u = c =$ const is to be solved. A solution of the homogeneous equation with $c = 0$ is $u_1(t) = (1/z)e^{1/z}$, and a solution of an inhomogeneous equation with $c = 1$ is

$$u_2(z) = u_1(z) \int \frac{1}{z^2 u_1(z)}\, dz.$$

Term-by-term integration of the integrand $(1/z)e^{-1/z} = \sum\limits_0^\infty (-1)^n/(n!\, z^{n+1})$ gives

$$u_2(z) = u_1(z)(\log z + h(z)) \ \text{ with } \ h(z) = \sum_{n=1}^\infty \frac{(-1)^{n+1}}{n \cdot n!\, z^n}.$$

The functions u_1, u_2 are a fundamental system of solutions of the original equation.

The transformation $\zeta = 1/z$, $w(\zeta) = u(1/\zeta)$ leads to

$$\zeta^2 w'' - \zeta(\zeta + 1)w' + w = 0, \quad \text{index equation} \quad P(\lambda) = (\lambda - 1)^2 = 0.$$

For indices $\lambda_1 = \lambda_2 = 1$, $\lambda_1 - \lambda_2$ is an integer; hence a log term may occur (24.XIII). The power-series ansatz $w = \sum\limits_0^\infty w_k \zeta^{k+1}$ leads to $k^2 w_k - k w_{k-1} = 0$ $(k \geq 0)$ with $w_{-1} = 0$. The choice $w_0 = 1$ yields $w_k = 1/k!$. The result is the above solution $w = \zeta e^\zeta = u_1(1/\zeta)$.

The second solution is obtained by the transformation $v(s) = w(e^s)$ (24.VII), by which the differential equation becomes $v'' - (2 + e^s)v' + v = 0$. The ansatz

$$v(s) = \sum_{k=0}^\infty \frac{a_k + b_k s}{k!} e^{(k+1)s} \qquad \text{(compare (24.19))}$$

leads to the recursion formulae

$$k^2 b_k - k^2 b_{k-1} = 0 \quad \text{and} \quad k^2 a_k - k^2 a_{k-1} + 2kb_k - kb_{k-1} = 0$$

($k \geq 0$, $a_{-1} = b_{-1} = 0$). With $b_0 = -1$ and $a_0 = 0$ one gets $b_k = -1$ for all k and $a_k = 1 + \dfrac{1}{2} + \cdots + \dfrac{1}{k}$. Hence

$$w(\zeta) = -\zeta e^{\zeta} \log \zeta + \sum_{k=1}^{\infty} \frac{a_k}{k!} \zeta^{k+1}, \quad a_k = 1 + \frac{1}{2} + \cdots + \frac{1}{k}.$$

We chose b_0, a_0 such that $w(1/z) = u_2(z)$; the log-free term in u_2 starts with $1/z^2$. An independent proof that $u_1 \cdot h$ equals the above sum leads to the following interesting relation, which can be proved by induction:

$$a_k = \sum_{i=1}^{k} \frac{(-1)^{i+1}}{i} \binom{k}{i}.$$

Exercises in § 26

XVII. (a_1) From the general solution $u = a \cos x + b \sin x + \frac{1}{2} e^x$ one easily obtains

$$u = -\frac{1}{2} \cos x + \frac{\cos 1 - e}{2 \sin 1} \sin x + \frac{1}{2} e^x.$$

(a_2) Using Green's function from Exercise XVI one gets

$$(\sin 1) u(x) = \sin(x-1) \textstyle\int_0^x e^{\xi} \sin \xi \, d\xi + \sin x \int_x^1 e^{\xi} \sin(\xi - 1) \, d\xi.$$

With $2 \int e^{\xi} \sin(\xi - \alpha) \, d\xi = e^{\xi}(\sin(\xi - \alpha) - \cos(\xi - \alpha))$ it follows that

$$\begin{aligned} 2(\sin 1) u(x) &= \sin(x-1)[e^x(\sin x - \cos x) + 1] \\ &\quad - \sin x [e^x(\sin(x-1) - \cos(x-1)) - e]. \end{aligned}$$

Since $\sin x \cdot \cos(x-1) - \cos x \cdot \sin(x-1) = \sin 1$, we get

$$u(x) = \frac{1}{2} e^x + \frac{\sin(x-1)}{2 \sin 1} - \frac{e \sin x}{2 \sin 1}.$$

The addition theorem for the sine gives the solution in the form (a_1).

(b) For $v(t) = u(e^t)$ one gets $4\ddot{v} - 4\dot{v} = 0$, whence $v = e^{t/2}$ and $v = te^{t/2}$, i.e., $u = \sqrt{x}$ and $u = \sqrt{x} \log x$. In the construction given in V, we can choose $u_1 = \sqrt{x} \log x$, $u_2 = \sqrt{x} \log(x/2)$ to get $c = \log 2$,

$$(\log 2) \Gamma(x, \xi) = \begin{cases} \sqrt{\xi} \log \xi \cdot \sqrt{x} \log \dfrac{x}{2} & \text{for} \quad 1 \leq \xi \leq x \leq 2, \\[2ex] \sqrt{x} \log x \cdot \sqrt{\xi} \log \dfrac{\xi}{2} & \text{for} \quad 1 \leq x \leq \xi \leq 2. \end{cases}$$

(c) $u_1 = 1$, $u_2 = 1 - x$, $c = -1$.

(d) Use Theorem IX.

XXIV. (b) Formula (10) with $u_1 = x^{1-\alpha}$, $u_2 = 1$, $c = \alpha - 1$ or $u_1 = 1$, $u_2 = 1 - x^{1-\alpha}$, $c = \alpha - 1$.

Exercises in § 27

XV. For $\gamma > \alpha^2/4$ in the case $\alpha = \beta$ and for $\gamma > 0$ in the case $\alpha < \beta$.

XVI. (a) $u(1) = 0$ yields $u = c \cdot \sin\sqrt{\lambda}\,(x-1)$, and $u(0) = u'(0)$ yields $-\sin\sqrt{\lambda} = \sqrt{\lambda}\cos\sqrt{\lambda}$ or $\sqrt{\lambda} = -\tan\sqrt{\lambda}$. For the equation $-s = \tan s$, the set of positive solutions $s_0, s_1, \dots$ with $(n+\frac{1}{2})\pi < s_n < (n+1)\pi$ and $s_n - (n+\frac{1}{2})\pi \searrow 0$ is countable, as seen from a sketch of the tangent function. From the equivalent fixed-point equation $s = -\arctan s + (n+1)\pi$ the number s_n is obtained by iteration [contraction principle 5.IX; note that the map $s \to \arctan s$ is contracting in the interval $(\frac{\pi}{2}, \infty)$]. Therefore, the nth eigenvalue and the nth eigenfunction are given by

$$\lambda_n = s_n^2, \quad u_n = \sin s_n(x-1) \quad (n = 0, 1, 2, \dots).$$

You should convince yourself that there are no eigenvalues for $\lambda \le 0$.

(b) $u(0) = u'(0)$ is satisfied by $u = \sin(\sqrt{\lambda}\,x + a)$ with $a = \arctan\sqrt{\lambda}$, and the corresponding condition at $x = 1$ leads to the equation $\tan a = \tan(\sqrt{\lambda} + a)$. Hence $\sqrt{\lambda} = n\pi$, and accordingly

$$\lambda_n = n^2\pi^2 \text{ and } u_n = \sin(n\pi x + a_n) \text{ with } a_n = \arctan n\pi \ (n = 1, 2, \dots).$$

The function $u_1 = \sin(\pi x + \arctan\pi)$ has a zero at $(0,1)$; thus the numbering of the λ_n matches that in Theorem II. However, λ_0 and u_0 are still missing. As is easily seen, $u_0 = \mathrm{e}^x$ satisfies the boundary conditions and yields $\lambda_0 = -1$. Additional eigenvalues $\lambda \le 0$ do not occur.

(c) For $v(t) = u(\mathrm{e}^t)$ the eigenvalue problem is $\ddot{v} + \lambda v = 0$, $\dot{v}(0) = \dot{v}(2\pi) = 0$ with the solution $v_n(t) = \cos\frac{1}{2}nt$, $\lambda_n = \frac{1}{4}n^2$ for $n = 0, 1, 2, \dots$. Hence

$$\lambda_n = \frac{1}{4}n^2 \text{ and } u_n(x) = \cos\left(\frac{n}{2}\log x\right) \ (n = 0, 1, 2, \dots).$$

In particular, $\lambda_0 = 0$ and $u_0(x) \equiv 1$.

Exercise in § 28

XIV. (c) The problem of the vibrating string is

$$\begin{aligned} \phi_{tt} &= \phi_{xx} \text{ for } 0 < x < \pi,\ t > 0, \\ \phi(t,0) &= \phi(t,\pi) = 0,\ \phi(0,x) = f(x),\ \phi_t(0,x) = g(x). \end{aligned}$$

It has the solution

$$\phi(t,x) = \sum_{n=1}^{\infty} (c_n \cos nt + d_n \sin nt) \sin nx$$

with

$$c_n = \frac{2}{\pi}\int_0^\pi f(x)\sin nx\,dx, \quad d_n = \frac{2}{n\pi}\int_0^\pi g(x)\sin nx\,dx.$$

Literature

Cited Articles

Arenstorf, R.F.: Periodic solutions of the restricted three body problem representing analytic continuations of Keplerian elliptic motions. *Amer. J. Math.* **85**, 27–35 (1963).

Banach, S.: Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fund. math.* **3**, 133–181 (1922).

Bessaga, C.: On the converse of the Banach "fixed point principle". *Colloq. Math.* **7**, 41–43 (1959).

Bompiani, E.: Un teorema di confronto ed un teorema di unicità per l'equazione differenziale $y' = f(x, y)$. *Atti Accad. Naz. Lincei Rend. Classe Sci. Fis. Mat. Nat.* (6) **1**, 298–302 (1925).

Bulirsch, R., Stoer, J.: Numerical treatment of ordinary differential equations by extrapolation methods. *Numerische Math.* **8**, 1–13 (1966).

Četaev, N.G.: Un théorème sur l'instabilité. *Dokl. Akad. Nauk SSSR* **2**, 529–534 (1934).

Grobman, D.: Homeomorphisms of systems of differential equations. *Dokl. Akad. Nauk SSSR* **129**, 880–881 (1959).

Gronwall, T.H.: Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Ann. Math.* **20**, 292–296 (1918).

Harris, W.A.Jr., Sibuya, Y., Weinberg, L.: Holomorphic solutions of linear differential systems at singular points. *Arch. Rat. Mech. Anal.* **35**, 245–248 (1969).

Hartman, P.: On the local linearization of differential equations. *Proc. Amer. Math. Soc.* **14**, 568–573 (1963).

Hirsch, M.: Systems of differential equations that are competitive or cooperative, II. Convergence almost everywhere. *SIAM J. Math. Anal.* **16**, 423–439 (1985).

Holmes, R.B.: A formula for the spectral radius of an operator. *Amer. Math. Monthly* **75**, 163–166 (1968).

Hopf, E.: A remark on linear elliptic differential equations of second order. *Proc. Amer. Math. Soc.* **3**, 791–793 (1952).

Kamke, E.: Zur Theorie der Systeme gewöhnlicher Differentialgleichungen, II. *Acta Math.* **58**, 57–85 (1932).

Krasnosel'skii, M.A., Krein, S.G.: On a class of uniqueness theorems for the equation $y' = f(x, y)$. *Usp. Mat. Nauk (N.S.)* **11**, No. 1, (67), 209–213 (1956).

LaSalle, J.P.: Stability theory for ordinary differential equations. *J. Diff. Eqs.* **4**, 57–65 (1968).

Lazer, A.C., McKenna, P.J.: Large-amplitude periodic oscillations in suspension bridges: Some new connections with nonlinear analysis. *SIAM Review* **32**, 537–578 (1990).

Lohner, R.: *Einschliesung der Lösung gewöhnlicher Anfangs- und Randwertaufgaben und Anwendungen.* Dissertation, Universität Karlsruhe 1988.

Lohner, R.: Enclosing the solutions of ordinary initial and boundary value problems. *Computer Arithmetic—Scientific Computation and Programming Languages.* Ed. E. Kaucher, U. Kulisch, Ch. Ullrich, pp. 255–286. Stuttgart: Teubner 1987.

McKenna, P.J., Walter, W.: Nonlinear oscillations in a suspension bridge. *Arch. Rat. Mech. Anal.* **98**, 167–177 (1987).

McNabb, A.: Strong comparison theorems for elliptic equations of second order. *J. Math. Mech.* **10**, 431–440 (1961).

Morgenstern, D.: *Beiträge zur nichtlinearen Funktionalanalysis.* Dissertation, Technische Universität Berlin 1952.

Müller, M.: Über das Fundamentaltheorem in der Theorie der gewöhnlichen Differentialgleichungen. *Math. Z.* **26**, 619–645 (1926).

Müller, M.: Über die Eindeutigkeit der Integrale eines Systems gewöhnlicher Differentialgleichungen und die Konvergenz einer Gattung von Verfahren zur Approximation dieser Integrale. *Sitz.-Ber. Heidelberger Akad. Wiss. Math.-Naturwiss. Kl.* 1927, 9. Abh.

Ni, W.-M.: On the elliptic equation $\Delta u + K(x)u^\sigma = 0$, its generalisations, and application in geometry. *Indiana Univ. Math. J.* **31**, 493–529 (1982).

Olech, C.: Remarks concerning criteria for uniqueness of solutions of ordinary differential equations. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astron. Phys.* **8**, 661–666 (1960).

Osgood, W.F.: Beweis der Existenz einer Lösung der Differentialgleichung $dy/dx = f(x, y)$ ohne Hinzunahme der Cauchy–Lipschitzschen Bedingung. *Monatsh. Math. Phys.* **9**, 331–345 (1898).

Ostrowski, A.: The round-off stability of iterations. *Zeitschr. Angew. Math. Mech.* **47**, 77–82 (1967).

Peano, G.: Démonstration de l'integrabilité des équations différentielles ordinaires. *Math. Ann.* **37**, 182–228 (1890).

Perron, O.: Ein neuer Existenzbeweis für die Integrale der Differentialgleichung $y' = f(x, y)$. *Math. Ann.* **76**, 471–484 (1915).

Perron, O.: Über Ein- und Mehrdeutigkeit des Integrals eines Systems von Differentialgleichungen. *Math. Ann.* **95**, 98–101 (1926).

Prüfer, H.: Neue Herleitung der Sturm–Liouvilleschen Reihenentwicklung stetiger Funktionen. *Math. Ann.* **95**, 499–518 (1926).

Rosenblatt, A.: Über die Existenz von Integralen gewöhnlicher Differentialgleichungen. *Arkiv Mat. Astron. Fys.* **5**, Nr. 2 (1909).

Sauvage, L.: Sur les solutions régulières d'un system. *Ann. Sci. École Norm. Super.* (3) **3**, 391–404 (1886).

Walter, W.: There is an elementary proof of Peano's existence theorem. *Amer. Math. Monthly* **78**, 170–173 (1971).

Walter, W.: A note on contraction. *SIAM Review* **18**, 107–111 (1976).

Walter, W.: Bemerkungen zum Kontraktionsprinzip. *Elemente d. Math.* **31**, 90–91 (1976).

Walter. W.: A theorem on elliptic differential inequalities with an application to gradient bounds. *Math. Z.* **200**, 293–299 (1990).

Walter, W.: Minimum principles for weak solutions of second order differential equations. In: *General Inequalities 6*, Birkhäuser Verlag 1992, pp. 369–376.

Walter, W.: A useful Banach algebra. *Elemente d. Math.* **47**, 27-32 (1992).

Walter, W.: A new approach to minimum and comparison principles for nonlinear ordinary differential operators of second order. *Nonlinear Analysis* **25**, 1071–1078 (1995).

Walter, W.: Enclosure methods for capricious solutions of ordinary differential equations. *LAM 32 "The Mathematics of Numercial Analysis"*, ed. J. Renegar, M. Shub, S. Smale. Amer. Math. Soc. 1996, 867–878.

Walter, W.: A note on Sturm type comparison theorems for nonlinear operators. *J. Diff. Eqs.* **135**, 358–365 (1997).

Walter, W.: On strongly monotone flows. *Ann. Polon. Math.* **66**, 269–274 (1997).

Textbooks and Monographs

Amann, H.: *Ordinary Differential Equations. An Introduction to Nonlinear Analysis.* Berlin–New York: deGruyter 1990.

Arnold, V. I.: *Ordinary Differential Equations.* Berlin–Heidelberg: Springer 1992.

Brauer, F., Nohel, J.A.: *The qualitative theory of ordinary differential equations. An introduction.* New York–Amsterdam: W. A. Benjamin, Inc. 1969.

Braun, M.: *Differential Equations and their Applications, 2nd ed.* Applied Math. Sciences 15. New York: Springer 1978.

Burckel, R. B.: *An Introduction to Classical Complex Analysis, Vol. 1.* Basel–Stuttgart: Birkhäuser 1979.

Carathéodory, C.: *Vorlesungen über reelle Funktionen.* Leipzig: Teubner 1918.

Cesari, L.: *Asymptotic behavior and stability problems in ordinary differential equations. 3rd ed.* Berlin–Göttingen–Heidelberg: Springer 1971.

Coddington, E.A., Levinson, N.: *Theory of ordinary differential equations.* New York–Toronto–London: McGraw-Hill Book Co. 1955.

Collatz, L.: *The numerical treatment of differential equations.* 3rd ed. Berlin–Heidelberg–New York: Springer 1966.

Collatz, L.: *Differentialgleichungen. Eine Einführung unter besonderer Berücksichtigung der Anwendungen. 7. Aufl.* Stuttgart: Teubner 1988.

Drazin, P.G.: Nonlinear Systems. Cambridge: Cambridge Univ. Press 1992.

Hahn, W.: *Stability of Motion.* Die Grundlehren der mathematischen Wissenschaften, Bd. 138. Berlin–Heidelberg–New York: Springer 1967.

Hale, J.K., Koçak, H.: *Dynamics and Bifurcations.* New York: Springer 1991.

Hartman, P.: *Ordinary differential equations.* New York–London–Sydney: John Wiley 1964.

Heuser, H.: *Gewöhnliche Differentialgleichungen. 3. Aufl.* Stuttgart: Teubner 1995.

Jordan, D.W., Smith, P.: *Nonlinear Ordinary Differential Equations. 2nd ed.* Oxford: Clarendon Press 1988.

Kamke, E.: *Differentialgleichungen reeller Funktionen. 2. Aufl.* Leipzig: Akademische Verlagsgesellschaft 1945.

Kamke, E.: *Differentialgleichungen. Lösungsmethoden und Lösungen. Bd. 1, 10. Aufl.* Stuttgart: Teubner 1983.

Kamke, E.: *Differentialgleichungen, I. Gewöhnliche Differentialgleichungen. 5. Aufl.* Leipzig: Akademische Verlagsgesellschaft 1964.

Krasovskii, N.N.: *Stability of Motion.* Palo Alto, California: Stanford Univ. Press 1963.

Pólya, G., Szegő, G.: *Aufgaben und Lehrsätze aus der Analysis I. 4. Aufl.* Berlin–Heidelberg–New York: Springer 1970.

Redheffer, R.: *Differential Equations, Theory and Applications.* Boston: Jones and Bartlett 1991.

Reissig, R., Sansone, G., Conti, R.: *Qualitative Theorie nichtlinearer Differentialgleichungen.* Roma: Edizioni Cremonese 1963.

Remmert, R.: *Theory of Complex Functions.* Graduate Texts in Mathematics 122. New York: Springer 1991.

Titchmarsh, E.C.: *Eigenfunction Expansions associated with Second-order Differential Equations, Part I.* Oxford: Clarendon Press 1962.

Walter, W.: *Differential- und Integral-Ungleichungen.* Springer Tracts in Natural Philosophy, Vol. 2. Berlin–Göttingen–Heidelberg–New York: Springer 1964.

Walter, W.: *Differential and Integral Inequalities.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Bd. 55. Berlin–Heidelberg–New York: Springer 1970.

Walter, W.: *Analysis 1 und 2.* Grundwissen Mathematik, Bd. 3 und 4. 4. Aufl. Berlin–Heidelberg–New York–Tokyo: Springer 1997, 1994. Cited as Walter 1 or 2.

Wiggins, S.: *Global Bifurcations and Chaos.* New York: Springer 1988.

Index

Notation

Sets. We denote by $\mathbb{N} = \{1, 2, 3, \ldots\}$ the set of positive integers, by $\mathbb{Z}$ the set of all integers, by $\mathbb{R}$ the set of real numbers, and by $\mathbb{C}$ the set of complex numbers. The symbol $\mathbb{R}^n$ stands for the set of all n-tuples of real numbers (n-dimensional Euclidean space) and $\mathbb{C}^n$ for the set of all n-tuples of complex numbers (n-dimensional complex or unitary space); cf. 5.III.(a), (b).

The boundary of a set A is denoted by ∂A, the interior by A°, and the closure by $\overline{A} = \partial A \cup A^\circ$.

Intervals. As usual, intervals of real numbers are denoted by $[a, b]$, (a, b), $[a, b)$, $(a, b]$. An interval without further specification can be open, closed, half-open, bounded, or unbounded. Thus $\mathbb{R}$ and $[a, \infty) = \{x \in \mathbb{R} : x \geq a\}$ are also intervals.

Functions. The *graph* of a function $f : D \to E$ is denoted by $\operatorname{graph} f$. Thus $\operatorname{graph} f$ is the set of all pairs $(x, f(x)) \in D \times E$ with $x \in D$. If $A \subset D$, then $g = f|A$ is the *restriction* of f to A. Thus $\operatorname{Dom} g = A$ and $g(x) = f(x)$ for $x \in A$. Any function $h : B \to E$ with $B \supset D$ and $h|D = f$ is called an *extension* of f. The image of A under f is

$$f(A) := \{y \in E : \text{ there is an } x \in A \text{ with } y = f(x)\}.$$

Classes of Functions. For $M \subset \mathbb{R}^n$ the class of continuous functions on M is denoted by $C(M)$. Depending on context, the functions in $C(M)$ are real-valued, complex-valued, or vector-valued. The symbol $C^k(J)$ represents the class of functions that are k-times differentiable on the interval J with the convention that $C^0(J) = C(J)$. If G is an open set in $\mathbb{R}^n$, then $C^k(G)$ is the class of functions that together with all partial derivatives of order $\leq k$ are continuous in G. We set $C^0(G) = C(G)$. Further classes are listed under the abbreviations below; they are explained in the text.

Trajectories. Some trajectories are marked with dots that correspond to equidistant t-values; cf. 3.V, VI.

Abbreviations

$AC(J)$	(absolutely continuous on J)	121
B_r	(ball in $\mathbb{R}^n$ or in a Banach space)	
$C^k(\overline{G})$	(continuously differentiable in $\overline{G}$)	153
D_-, D^-	(Dini derivative)	89, 342

Δ	(Laplace operator)	71
Δ_p	(p-Laplacian)	141
$\operatorname{dist}(\mathbf{x}, A)$	(distance from the point $\mathbf{x}$ to the set A)	117, 323
$\operatorname{dist}(A, B)$	(distance between the sets A and B)	323
$\mathbf{e}_i$	(unit vector)	160
$\operatorname{Ext}(C)$	(exterior of the closed curve C)	337
(H)		16
$H(G)$	(holomorphic in G)	84, 213, 348
$H_0(G)$	(holomorphic and bounded in G)	55
H_δ		225
H_δ^q		230
H_r	(Hilbert space with weighted norm)	296
I	(identity matrix)	160
I_α		70
$\operatorname{Int}(C)$	(interior of the closed curve C)	337
K_r	(disk $\|z\| < r$)	222
K_r^0	(punctured disk $0 < \|z\| < r$)	217
K_r^-	(disk minus a radius)	221
L_α		70
$L(J)$	(integrable over J)	121
$L^2(J)$	(quadratically integrable over J)	
$L_{\mathrm{loc}}(J)$	(locally integrable in J)	
P_q		229
R_α	(half-plane $\operatorname{Re} z < \alpha$)	217
S	(class of functions)	279
(S)	(assumption for Sturmian theory)	246
(SL)	(assumption for Sturm–Liouville theory)	269
$\sigma(A)$	(spectrum of the matrix A)	176
$\operatorname{tr} A$	(trace of the matrix A)	166
(U)	(uniqueness condition)	67, 146
$\|\mathbf{x}\|_e$	(Euclidean norm)	55, 106
$\\|\cdot\\|_r$	(weighted norm)	296
$(x)_m$		239
$(.,.)$	(inner product)	286
$(.,.)_r$	(inner product)	295

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory. 2nd ed.
2 OXTOBY. Measure and Category. 2nd ed.
3 SCHAEFER. Topological Vector Spaces.
4 HILTON/STAMMBACH. A Course in Homological Algebra. 2nd ed.
5 MAC LANE. Categories for the Working Mathematician. 2nd ed.
6 HUGHES/PIPER. Projective Planes.
7 SERRE. A Course in Arithmetic.
8 TAKEUTI/ZARING. Axiomatic Set Theory.
9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.
10 COHEN. A Course in Simple Homotopy Theory.
11 CONWAY. Functions of One Complex Variable I. 2nd ed.
12 BEALS. Advanced Mathematical Analysis.
13 ANDERSON/FULLER. Rings and Categories of Modules. 2nd ed.
14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.
15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.
16 WINTER. The Structure of Fields.
17 ROSENBLATT. Random Processes. 2nd ed.
18 HALMOS. Measure Theory.
19 HALMOS. A Hilbert Space Problem Book. 2nd ed.
20 HUSEMOLLER. Fibre Bundles. 3rd ed.
21 HUMPHREYS. Linear Algebraic Groups.
22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.
23 GREUB. Linear Algebra. 4th ed.
24 HOLMES. Geometric Functional Analysis and Its Applications.
25 HEWITT/STROMBERG. Real and Abstract Analysis.
26 MANES. Algebraic Theories.
27 KELLEY. General Topology.
28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.
29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.
30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.
31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.
32 JACOBSON. Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.
33 HIRSCH. Differential Topology.
34 SPITZER. Principles of Random Walk. 2nd ed.
35 ALEXANDER/WERMER. Several Complex Variables and Banach Algebras. 3rd ed.
36 KELLEY/NAMIOKA et al. Linear Topological Spaces.
37 MONK. Mathematical Logic.
38 GRAUERT/FRITZSCHE. Several Complex Variables.
39 ARVESON. An Invitation to C^*-Algebras.
40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.
41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory. 2nd ed.
42 SERRE. Linear Representations of Finite Groups.
43 GILLMAN/JERISON. Rings of Continuous Functions.
44 KENDIG. Elementary Algebraic Geometry.
45 LOÈVE. Probability Theory I. 4th ed.
46 LOÈVE. Probability Theory II. 4th ed.
47 MOISE. Geometric Topology in Dimensions 2 and 3.
48 SACHS/WU. General Relativity for Mathematicians.
49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.
50 EDWARDS. Fermat's Last Theorem.
51 KLINGENBERG. A Course in Differential Geometry.
52 HARTSHORNE. Algebraic Geometry.
53 MANIN. A Course in Mathematical Logic.
54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.
55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.
56 MASSEY. Algebraic Topology: An Introduction.
57 CROWELL/FOX. Introduction to Knot Theory.
58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.
59 LANG. Cyclotomic Fields.
60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.

61 WHITEHEAD. Elements of Homotopy Theory.
62 KARGAPOLOV/MERLZJAKOV. Fundamentals of the Theory of Groups.
63 BOLLOBÁS. Graph Theory.
64 EDWARDS. Fourier Series. Vol. I 2nd ed.
65 WELLS. Differential Analysis on Complex Manifolds. 2nd ed.
66 WATERHOUSE. Introduction to Affine Group Schemes.
67 SERRE. Local Fields.
68 WEIDMANN. Linear Operators in Hilbert Spaces.
69 LANG. Cyclotomic Fields II.
70 MASSEY. Singular Homology Theory.
71 FARKAS/KRA. Riemann Surfaces. 2nd ed.
72 STILLWELL. Classical Topology and Combinatorial Group Theory. 2nd ed.
73 HUNGERFORD. Algebra.
74 DAVENPORT. Multiplicative Number Theory. 2nd ed.
75 HOCHSCHILD. Basic Theory of Algebraic Groups and Lie Algebras.
76 IITAKA. Algebraic Geometry.
77 HECKE. Lectures on the Theory of Algebraic Numbers.
78 BURRIS/SANKAPPANAVAR. A Course in Universal Algebra.
79 WALTERS. An Introduction to Ergodic Theory.
80 ROBINSON. A Course in the Theory of Groups. 2nd ed.
81 FORSTER. Lectures on Riemann Surfaces.
82 BOTT/TU. Differential Forms in Algebraic Topology.
83 WASHINGTON. Introduction to Cyclotomic Fields. 2nd ed.
84 IRELAND/ROSEN. A Classical Introduction to Modern Number Theory. 2nd ed.
85 EDWARDS. Fourier Series. Vol. II. 2nd ed.
86 VAN LINT. Introduction to Coding Theory. 2nd ed.
87 BROWN. Cohomology of Groups.
88 PIERCE. Associative Algebras.
89 LANG. Introduction to Algebraic and Abelian Functions. 2nd ed.
90 BRØNDSTED. An Introduction to Convex Polytopes.
91 BEARDON. On the Geometry of Discrete Groups.
92 DIESTEL. Sequences and Series in Banach Spaces.
93 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part I. 2nd ed.
94 WARNER. Foundations of Differentiable Manifolds and Lie Groups.
95 SHIRYAEV. Probability. 2nd ed.
96 CONWAY. A Course in Functional Analysis. 2nd ed.
97 KOBLITZ. Introduction to Elliptic Curves and Modular Forms. 2nd ed.
98 BRÖCKER/TOM DIECK. Representations of Compact Lie Groups.
99 GROVE/BENSON. Finite Reflection Groups. 2nd ed.
100 BERG/CHRISTENSEN/RESSEL. Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions.
101 EDWARDS. Galois Theory.
102 VARADARAJAN. Lie Groups, Lie Algebras and Their Representations.
103 LANG. Complex Analysis. 3rd ed.
104 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part II.
105 LANG. $SL_2(\mathbf{R})$.
106 SILVERMAN. The Arithmetic of Elliptic Curves.
107 OLVER. Applications of Lie Groups to Differential Equations. 2nd ed.
108 RANGE. Holomorphic Functions and Integral Representations in Several Complex Variables.
109 LEHTO. Univalent Functions and Teichmüller Spaces.
110 LANG. Algebraic Number Theory.
111 HUSEMÖLLER. Elliptic Curves.
112 LANG. Elliptic Functions.
113 KARATZAS/SHREVE. Brownian Motion and Stochastic Calculus. 2nd ed.
114 KOBLITZ. A Course in Number Theory and Cryptography. 2nd ed.
115 BERGER/GOSTIAUX. Differential Geometry: Manifolds, Curves, and Surfaces.
116 KELLEY/SRINIVASAN. Measure and Integral. Vol. I.
117 SERRE. Algebraic Groups and Class Fields.
118 PEDERSEN. Analysis Now.

119 ROTMAN. An Introduction to Algebraic Topology.
120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.
121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups. 2nd ed.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.
131 LAM. A First Course in Noncommutative Rings.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic K-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.
160 LANG. Differential and Riemannian Manifolds.
161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.
162 ALPERIN/BELL. Groups and Representations.
163 DIXON/MORTIMER. Permutation Groups.
164 NATHANSON. Additive Number Theory:
165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.
166 SHARPE. Differential Geometry: Cartan's
167 MORANDI. Field and Galois Theory.
168 EWALD. Combinatorial Convexity and Algebraic Geometry.
169 BHATIA. Matrix Analysis.
170 BREDON. Sheaf Theory. 2nd ed.
171 PETERSEN. Riemannian Geometry.
172 REMMERT. Classical Topics in Complex Function Theory.
173 DIESTEL. Graph Theory.
174 BRIDGES. Foundations of Real and Abstract Analysis.
175 LICKORISH. An Introduction to Knot Theory.
176 LEE. Riemannian Manifolds.
177 NEWMAN. Analytic Number Theory.
178 CLARKE/LEDYAEV/STERN/WOLENSKI. Nonsmooth Analysis and Control Theory.
179 DOUGLAS. Banach Algebra Techniques in Operator Theory. 2nd ed.
180 SRIVASTAVA. A Course on Borel Sets.
181 KRESS. Numerical Analysis.
182 WALTER. Ordinary Differential Equations.

183 MEGGINSON. An Introduction to Banach Space Theory.
184 BOLLOBÁS. Modern Graph Theory.
185 COX/LITTLE/O'SHEA. Using Algebraic Geometry.
186 RAMAKRISHNAN/VALENZA. Fourier Analysis on Number Fields.
187 HARRIS/MORRISON. Moduli of Curves.
188 GOLDBLATT. Lectures on Hyperreals.
189 LAM. Lectures on Rings and Modules.