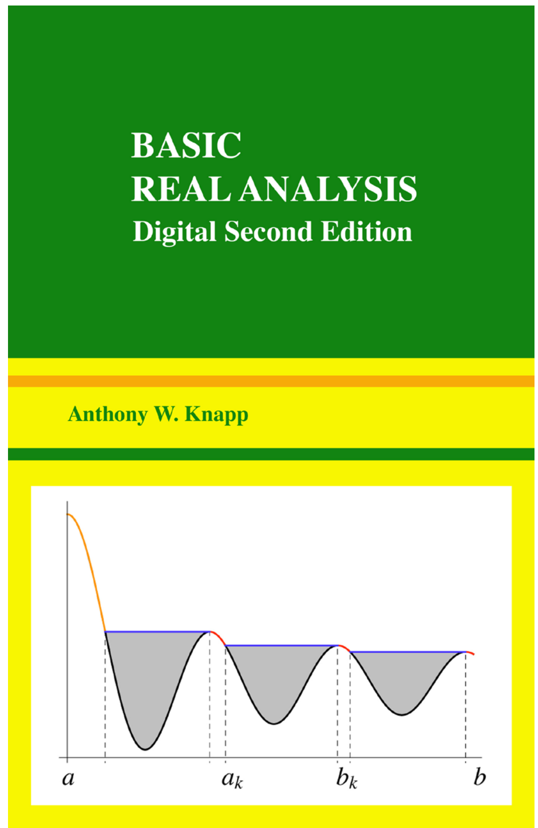
IV. Theory of Ordinary Differential Equations and Systems, 218-266

DOI: 10.3792/euclid/9781429799997-4

from

Basic Real Analysis
Digital Second Edition

Anthony W. Knapp



Full Book DOI: 10.3792/euclid/9781429799997
ISBN: 978-1-4297-9999-7

Anthony W. Knapp
81 Upper Sheep Pasture Road
East Setauket, N.Y. 11733–1729, U.S.A.
Email to: aknapp@math.stonybrook.edu
Homepage: www.math.stonybrook.edu/∼aknapp

Title: Basic Real Analysis, with an appendix "Elementary Complex Analysis"
Cover: An instance of the Rising Sun Lemma in Section VII.1.

Mathematics Subject Classification (2010): 28–01, 26–01, 42–01, 54–01, 34–01, 30–01, 32–01.

First Edition, ISBN-13 978-0-8176-3250-2

Published by Birkhäuser Boston

Digital Second Edition, not to be sold, no ISBN

Published by the author

CHAPTER IV

Theory of Ordinary Differential Equations and Systems

Abstract. This chapter treats the theory of ordinary differential equations, both linear and nonlinear.

Sections 1–4 establish existence and uniqueness theorems for ordinary differential equations. The first section gives some examples of first-order equations, mostly nonlinear, to illustrate certain kinds of behavior of solutions. The second section shows, in the presence of continuity for a vector-valued F satisfying a "Lipschitz condition," that the first-order system $y' = F(t, y)$ has a unique local solution satisfying an initial condition $y(t_0) = y_0$. Since higher-order equations can always be reduced to first-order systems, these results address existence and uniqueness for n^{th}-order equations as a special case. Section 3 shows that the solutions to a system depend well on the initial condition and on any parameters that are present in F. Section 4 applies these results to existence of integral curves for a vector field and to construction of coordinate systems from families of integral curves.

Sections 5–8 concern linear systems. Section 5 shows that local solutions of linear systems may be extended to global solutions and that in the homogeneous case the vector space of global solutions has dimension equal to the size of the system. The method of variation of parameters reduces the solution of any linear system to the solution of a homogeneous linear system. Sections 6–7 identify explicit solutions to n^{th}-order linear equations and first-order linear systems. The "Jordan canonical form" of a square matrix plays a role in the case of a system. Section 8 discusses power-series solutions to second-order homogeneous linear equations whose coefficients are given by convergent power series, as well as solutions that arise in the case of regular singular points. Two kinds of special functions are mentioned that result from this study—Legendre polynomials and Bessel functions.

1. Qualitative Features and Examples

To introduce the subject of ordinary differential equations, this section gives examples of some qualitative features and complicated phenomena that can occur in such equations.

If F is a complex-valued function of $n+2$ variables, a function $y(t)$ is said to be a **solution** of the **ordinary differential equation**

$$F(t, y, y', y'', \dots, y^{(m)}) = 0$$

of m^{th} **order** on the open interval (a, b) if

$$F(t, y(t), y'(t), \dots, y^{(m)}(t)) = 0$$

identically for $a < t < b$. The equation is "ordinary" in the sense that there is only one independent variable. The equation is said to be **linear** if it is of the form

$$a_m(t)y^{(m)} + a_{m-1}(t)y^{(m-1)} + \cdots + a_1(t)y' + a_0(t)y = q(t),$$

and it is **homogeneous linear** if in addition, q is the 0 function. A linear ordinary differential equation has **constant coefficients** if $a_m(t), \ldots, a_0(t)$ are all constant functions.

Let us come to examples, which will point toward the enormous variety of phenomena that can occur. We stick to the first-order case, and all the examples will have F real-valued. Let us look only for real-valued solutions. Pictures indicating the qualitative behavior of the solutions of each of the examples are in Figure 4.1.

EXAMPLES.

(1) Simple equations can have relatively complicated solutions. This is already true for the equation

$$y' = 1/t \qquad \text{on the interval } (0, +\infty).$$

Integration shows that all solutions are of the form $\log t + c$; on an interval of negative t's, the solutions are of the form $\log|t| + c$. The c comes from a corollary of the Mean Value Theorem that says that a real-valued function on an open interval with 0 derivative everywhere is necessarily constant.[1] Another example, but with no singularity, is $y' = ty$. To solve this equation on intervals where $y(t) \neq 0$, write $y'/y = t$, so that $\log|y| = \frac{1}{2}t^2 + a$ and $|y| = e^a e^{t^2/2}$. Thus $y(t) = ce^{t^2/2}$, with $c \neq 0$ constant, on any interval where $y(t)$ is nowhere 0. The function $y(t) = 0$ is a solution as well, and all real solutions on an interval are of the form $y(t) = ce^{t^2/2}$ with c real. See Figures 4.1a and 4.1b.

(2) Solutions may not be defined on obvious intervals. For the equation

$$ty' + y = \sin t,$$

we can recognize the two sides as $\frac{d}{dt}(ty)$ and $\frac{d}{dt}(-\cos t)$. Therefore $ty = c - \cos t$. Dividing by t, we obtain $y(t) = \frac{c - \cos t}{t}$ on any interval that does not contain 0. What about intervals containing $t = 0$? If we put $t = 0$ in the formula $ty = c - \cos t$, we see that c must be 1. In this case we can define $y(0) = 0$ there, and then $y'(0)$ exists. We obtain the additional solution

$$y(t) = \begin{cases} \dfrac{1 - \cos t}{t} & \text{for } t \neq 0, \\ 0 & \text{for } t = 0, \end{cases}$$

on any open interval containing 0. Figure 4.1c shows graphs of some solutions.

[1] See Section A2 of Appendix A for further information.

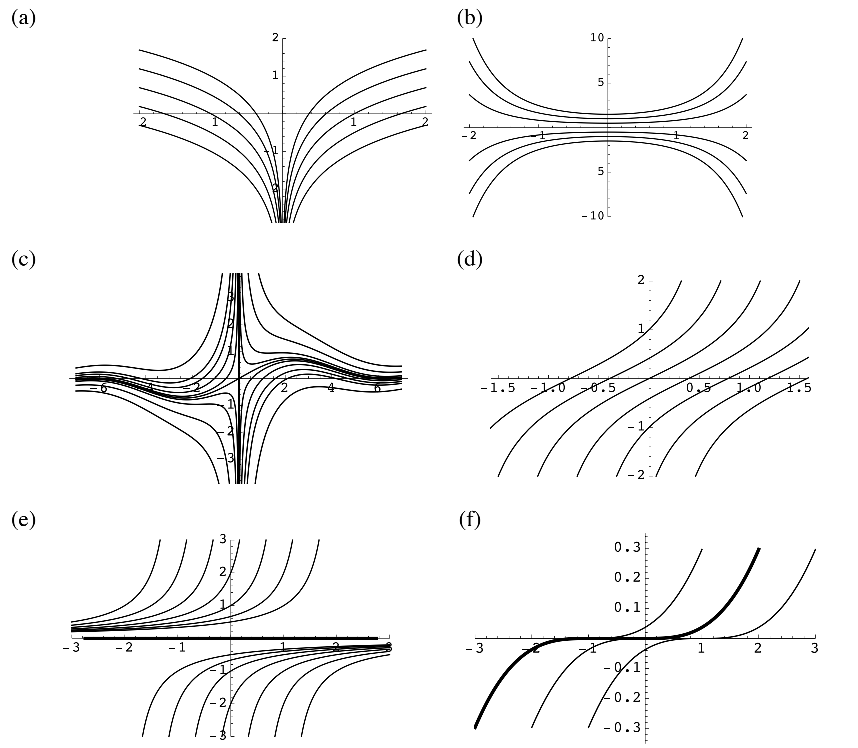


FIGURE 4.1. Graphs of solutions of some first-order ordinary differential equations: (a) $y' = 1/t$, (b) $y' = ty$, (c) $ty' + y = \sin t$, (d) $y' = y^2 + 1$, (e) $y' = y^2$, (f) $y' = y^{2/3}$.

(3) Even if the equation seems nice for all t, the solutions may not exist for all t. An example occurs with

$$y' = y^2 + 1,$$

which we solve by the steps $\frac{d}{dt}(\arctan y) = 1$, $\arctan y = t + c$, $y = \tan(t + c)$. The solutions behave badly when $t + c$ is any odd multiple of $\pi/2$. Solutions are defined at most on intervals of length π. Figure 4.1d shows graphs of some solutions for this example.

(4) Some solutions may look quite different from all the others. For example, with

$$y' = y^2,$$

we solve by $-1/y = t + c$ for $y \neq 0$, so that $y(t) = -\dfrac{1}{t + c}$. Also, $y(t) = 0$ is a solution. Here the solutions of the form $y(t) = -\frac{1}{t+c}$ are not defined for all t, but the solution $y(t) = 0$ *is* defined for all t. We might think of $y(t) = 0$ as the limiting case with c tending to $\pm\infty$. Figure 4.1e shows graphs of some of the solutions for this example.

(5) New solutions can sometimes be pieced together from old ones. For example, the equation

$$y' = y^{2/3}$$

is solved where $y \neq 0$ by the steps $y^{-2/3}y' = 1$, $3y^{1/3} = t + c$, and $y(t) = \frac{1}{27}(t+c)^3$. But also $y(t) = 0$ is a solution. In fact, we can piece solutions of these types together. For example, the function

$$y(t) = \begin{cases} \frac{1}{27}(t+1)^3 & \text{for } t < -1, \\ 0 & \text{for } -1 \leq t \leq 0, \\ \frac{1}{27}t^3 & \text{for } 0 < t, \end{cases}$$

is a solution on $(-\infty, +\infty)$. Figure 4.1f shows graphs of some of the solutions for this example.

One thing that stands out in the above examples is that the set of solutions seems to depend, more or less, on a single parameter c. The inference is that nothing much worse than the c occurs because somewhere an integration is taking place and the Mean value Theorem is controlling how many indefinite integrals there can be. One way of trying to quantify this statement about how the number of solutions is limited is to say that for any fixed $t = t_0$ and given real number y_0, there is only one solution $y(t)$ near t_0 with $y(t_0) = y_0$. This statement is not quite accurate, however, as Example 5 shows. The uniqueness theorem in Section 2 will give a precise result. The data (t_0, y_0) are called an **initial condition**.

Something else that stands out, although perhaps not without the visual aid of the graphs of solutions as in Figure 4.1, is that the graphed solutions appear to fill the entire part of the plane corresponding to the t's under study. In the framework of the previous paragraph, the statement is that for any fixed $t = t_0$ and given real number y_0, there exists a solution $y(t)$ near t_0 with $y(t) = y_0$. The existence theorem in Section 2 will give a precise result.

WEAK VERSION OF EXISTENCE AND UNIQUENESS THEOREMS. Let D be a nonempty convex open set in $\mathbb{R}^2$, and let (t_0, y_0) be in E. If $F : D \to \mathbb{R}$ is a continuous function such that $\frac{\partial}{\partial y}F(t, y)$ exists and is continuous in D, then for any sufficiently small open interval of t's containing t_0, the equation $y' = F(t, y)$ has a unique solution $y(t)$ with $y(t_0) = y_0$ such that the graph of $t \mapsto y(t)$ lies in D.

An improved theorem, together with a proof, will be given in Section 2. The proof of existence uses "Picard iterations," and the idea is as follows. First we convert the differential equation into an equivalent integral equation

$$y(t) = \int_{t_0}^{t} F(s, y(s))\,ds + y_0.$$

Second we use the right side as input and the left side as output to define successive approximations to a solution:

$$\begin{aligned} y_0(t) &= y_0, \\ y_1(t) &= \int_{t_0}^{t} F(s, y_0(s))\,ds + y_0, \\ &\ \vdots \\ y_{n+1}(t) &= \int_{t_0}^{t} F(s, y_n(s))\,ds + y_0. \end{aligned}$$

Third we use the Weierstrass M test to show that the series with partial sums $y_N(t) = y_0 + \sum_{n=1}^{N} (y_n(t) - y_{n-1}(t))$ is uniformly convergent. If the limiting function is denoted by $y(t)$, we check that $y(t)$ satisfies the integral equation from which we started. Hence $y(t)$ is a solution of the differential equation.

2. Existence and Uniqueness

In this section we state and prove the main existence and uniqueness theorems for solutions of ordinary differential equations. First let us establish an appropriate setting more general than the one in Section 1.

The examples in Section 1 were all of the first order. They could all have been written in the form $y = F(t, y)$ with F real-valued, and we considered real-valued solutions $y(t)$. From equations as simple as $y'' + y' + y = 0$, whose real-valued solutions are

$$y(t) = a_1 e^{-t/2} \cos(t\sqrt{3}/2) + a_2 e^{-t/2} \sin(t\sqrt{3}/2),$$

we know that it can be easier to work, at least initially, with complex-valued solutions. In this particular case, it is easier as a first step to find all complex-valued solutions, namely

$$y(t) = c_1 \exp\left(\tfrac{1}{2}(-1 + i\sqrt{3})t\right) + c_2 \exp\left(\tfrac{1}{2}(-1 - i\sqrt{3})t\right),$$

and then to extract the real-valued solutions from them. The solution method, which will be discussed in more detail in Section 6 below, involves finding all complex solutions of a certain polynomial equation with real coefficients, and the method is more natural if the coefficients of the polynomial equation are allowed to be complex.

Thus right away, it is natural to consider first-order equations $y' = F(t, y)$ with F complex-valued and to look for complex-valued solutions. The theory in Chapter III avoided working with functions of several variables in which some of the variables are complex, and we can update the theory of Chapter III here. The technique, which is to consider the complex variable y as two real variables $\operatorname{Re} y$ and $\operatorname{Im} y$, is again applicable. Thus we have only to think of $F(t, y)$ as a function of three real variables, even if we do not separate y into its two components in writing $F(t, y)$, and the theory of Chapter III applies directly. In adopting the point of view that y is actually two real variables, we need to apply the same consideration to y', and we are led to view $y' = F(t, y)$ as a *system* of two simultaneous equations, namely $\operatorname{Re} y' = \operatorname{Re} F(t, y)$ and $\operatorname{Im} y' = \operatorname{Im} F(t, y)$. This viewpoint merely makes our functions conform to the prescriptions of Chapter III. It is not necessary to work with the expanded notation; all we have to remember is that in this part of the theory we never differentiate a function with respect to a *complex* variable.

The utility of allowing $y' = F(t, y)$ to represent a *system* of ordinary differential equations has, in any event, been thrust upon us. Let us consider the notion of a system a bit more. With a little trick the second-order equation $y'' + y' + y = 0$ can itself be transformed into a system, quite apart from the issue of real vs. complex variables. The trick is to introduce two unknown functions u_1 and u_2 to play the roles of y and y'. Then u_1 and u_2 satisfy $u_2 = u_1'$ and $u_2' = u_1'' = y'' = -y' - y = -u_2 - u_1$. In other words, u_1 and u_2 satisfy the system

$$
\begin{aligned}
u_1' &= u_2,\\
u_2' &= -u_1 - u_2.
\end{aligned}
$$

Conversely if $u_1(t)$ and $u_2(t)$ satisfy this system of equations, then $y(t) = u_1(t)$ is a solution of $y'' + y' + y = 0$. In this way, the given second-order equation is completely equivalent to a certain system of two first-order equations with two unknown functions.

Let F be a function defined on an open set D of $\mathbb{R} \times \mathbb{C}^{km}$ and taking values in $\mathbb{C}^k$. A $\mathbb{C}^k$-valued function $y(t) = (y_1(t), \dots, y_k(t))$ is said to be a **solution** of the **system** $F(t, y, y', \dots, y^{(m)}) = 0$ of k ordinary differential equations of **order** m in the open interval (a, b) if $F(t, y(t), y'(t), \dots, y^{(m)}) = 0$ identically for $a < t < b$.

We saw that the single second-order equation $y'' + y' + y = 0$ is equivalent to a certain first-order system of two equations, and the technique for exhibiting this equivalence works more generally: a system of k equations of order m that has been solved for the m^{th}-order derivatives is equivalent to a system of km equations of first order.

We shall consider first-order systems of the form $y' = F(t, y)$, where F is continuous on an open subset D of $\mathbb{R} \times \mathbb{C}^n$ and takes values in $\mathbb{C}^n$. The example $y' = y^{2/3}$ in Section 1 fits these hypotheses, and we saw that the hoped-for uniqueness fails for this equation. In the weak theorem stated at the end of Section 1, an additional hypothesis was imposed in order to address this problem: for $y' = F(t, y)$ with only real-valued solutions of interest, the hypothesis is that $\partial F/\partial y$ exists and is continuous on the domain D of F. Generalizing this condition presumably means saying something about partial derivatives in each of the directions y_j for $1 \le j \le n$. In addition, we must remember the injunction against differentiating with respect to complex variables. Thus we really expect a condition concerning $2n$ first-order derivatives. Fortunately there is an easily stated less-stringent condition that is nevertheless good enough. The condition is that F satisfy a **Lipschitz condition** in its y variable, i.e., that there exist a real number k such that

$$|F(t, y_1) - F(t, y_2)| \le k|y_1 - y_2|$$

for all pairs of points (t, y_1) and (t, y_2) in the domain D of F.

If F is a real-valued continuous function of two real variables with a continuous partial derivative in the second variable, then the Mean Value Theorem gives

$$F(t, y_1) - F(t, y_2) = (y_1 - y_2)\,\frac{\partial F}{\partial y}(t, \xi)$$

with ξ between y_1 and y_2, *provided* the line segment from (t, y_1) to (t, y_2) lies in the domain D of F. The partial derivative is bounded on any compact subset of D, and thus F satisfies, on any compact convex subset of D, a Lipschitz condition in the second variable.

Theorem 4.1 (Picard–Lindelöf Existence Theorem). Let D be a nonempty open set in $\mathbb{R}^1 \times \mathbb{C}^n$, let (t_0, y_0) be in D, and suppose that $F : D \to \mathbb{C}^n$ is a continuous function such that $F(t, y)$ satisfies a Lipschitz condition in the y variable and has $|F(t, y)| \le M$ on D. Let R be a compact set in $\mathbb{R}^1 \times \mathbb{C}^n$ of the form

$$R = \{(t, y) \bigm| |t - t_0| \le a \text{ and } |y - y_0| \le b\},$$

and suppose that R is contained in D. Put $a' = \min\{a,\ b/M\}$. Then there exists a solution $y(t)$ of the system

$$y' = F(t, y)$$

on the open interval $|t - t_0| < a'$ satisfying the initial condition

$$y(t_0) = y_0.$$

REMARKS. A variant of Theorem 4.1 takes D to be in $\mathbb{R}^1 \times \mathbb{C}^n$ but insists only on continuity of F, not on the Lipschitz condition. Then a local solution still exists for $|t - t_0| < a'$. This better result, known as the "Cauchy–Peano Existence Theorem," appears in Problems 20–25 at the end of the chapter and is proved by an argument using Ascoli's Theorem. However, Example 5 in Section 1 shows that there is no corresponding uniqueness theorem, and within the text we omit the proof of the better existence theorem. Another variant of Theorem 4.1 assumes that the domain D of a given $F_{\mathbb{R}}$ lies in $\mathbb{R}^1 \times \mathbb{R}^n$, $F_{\mathbb{R}}$ takes values in $\mathbb{R}^n$, and y_0 is in $\mathbb{R}^n$. Then $y' = F_{\mathbb{R}}(t, y)$ has a solution $y(t)$ such that $y(t_0) = y_0$ and the range of y is $\mathbb{R}^n$. In fact, when $F_{\mathbb{R}}$ satisfies a Lipschitz condition in the y variable, this variant is a consequence of Theorem 4.1 as stated. To derive this variant, one extends the given function $F_{\mathbb{R}}$ from the subset of $\mathbb{R}^1 \times \mathbb{R}^n$ to a subset of $\mathbb{R}^1 \times \mathbb{C}^n$ by making it constant in $\operatorname{Im} y$. Specifically the new system is $y' = F(t, y)$ with $F(t, y) = F_{\mathbb{R}}(t, \operatorname{Re} y)$, and the initial condition remains as $y(t_0) = y_0$. The part of the system corresponding to equations for $\operatorname{Im} y'$ is just $\operatorname{Im} y' = 0$, since F is real-valued, and therefore $\operatorname{Im} y(t)$ is constant. Since y_0 is real, $\operatorname{Im} y(t)$ must be 0. Thus Theorem 4.1 yields a solution $y(t)$ with range $\mathbb{R}^n$ under these special hypotheses.

PROOF. The first step is to see that the set of differentiable functions $t \mapsto y(t)$ on $|t - t_0| < a'$ satisfying $y' = F(t, y)$ and $y(t_0) = y_0$ is the same as the set of continuous functions $t \mapsto y(t)$ on $|t - t_0| < a'$ satisfying the integral equation $y(t) = \int_{t_0}^t F(s, y(s))\, ds + y_0$.

If y is differentiable and satisfies the differential equation and the initial condition, then y is certainly continuous and hence $s \mapsto F(s, y(s))$ is continuous. Then $\int_{t_0}^t F(s, y(s))\, ds$ is differentiable by the Fundamental Theorem of Calculus (Theorem 1.32), and the differential equation shows that $y(t)$ and $\int_{t_0}^t F(s, y(s))\, ds$ have the same derivative for $|t - t_0| < a'$. Thus they differ by a constant. The constant is checked by putting $t = t_0$, and indeed y satisfies the integral equation.

Conversely if y is continuous and satisfies the integral equation, then $s \mapsto F(s, y(s))$ is continuous, and the Fundamental Theorem of Calculus shows that $\int_{t_0}^t F(s, y(s))\, ds$ is differentiable. This function equals $y(t) - y_0$ by the integral equation, and hence y is differentiable. Differentiating the two sides of the integral equation, we see that y satisfies the differential equation. Also, if we put $t = t_0$ in the integral equation, we see that y satisfies the initial condition $y(t_0) = y_0$.

Thus it is enough to prove existence for a continuous solution of the integral equation. For $t_0 - a' \leq t \leq t_0 + a'$, define inductively

$$\begin{aligned} y_0(t) &= y_0, \\ y_1(t) &= y_0 + \int_{t_0}^t F(s, y_0(s))\, ds, \end{aligned}$$

$$\vdots$$

$$y_n(t) = y_0 + \int_{t_0}^{t} F(s, y_{n-1}(s))\, ds,$$

with the usual convention that $\int_{t_0}^{t} = -\int_{t}^{t_0}$. Let us see inductively that the graph of $y_n(t)$ lies in the set

$$R' = \big\{(t, y) \,\big|\, |t - t_0| \le a' \text{ and } |y - y_0| \le b\big\},$$

for $|t - t_0| \le a'$. The graph of $y_0(t) = y_0$ is just $\big\{(t, y_0) \,\big|\, |t - t_0| < a'\big\}$, and this lies in R'. The inductive hypothesis is that $(t, y_{n-1}(t))$ lies in R' for $\big\{(t, y_0) \,\big|\, |t - t_0| \le a'\}$. Then

$$|y_n(t) - y_0| = \Big| \int_{t_0}^{t} F(s, y_{n-1}(s)\, ds \Big| \le M|t - t_0| \le Ma' \le b,$$

and therefore $(t, y_n(t))$ lies in R' for $|t - t_0| \le a'$. This completes the induction, and hence the graph of $y_n(t)$ lies in R' for $|t - t_0| \le a'$.

Now write

$$y_N(t) = y_0(t) + \sum_{n=1}^{N} [y_n(t) - y_{n-1}(t)]$$

for $N \ge 0$. We shall use the Weierstrass M test (Proposition 1.20), adapted to a series of functions with values in $\mathbb{C}^n$, to prove uniform convergence of this series. Thus we are to bound $|y_n(t) - y_{n-1}(t)|$, and we shall do so inductively for $n \ge 1$. We start from the inequality $|F(t, y)| \le M$ on R' and the Lipschitz condition

$$|F(t, y_j(t) - F(t, y_{j-1})| \le k|y_j(t) - y_{j-1}(t)| \quad \text{for } j \ge 1.$$

Say that $t_0 \le x \le t_0 + a'$ for definiteness. Then

$$|y_1(t) - y_0(t)| = \Big| \int_{t_0}^{t} F(s, y_0(s))\, ds \Big| \le M(t - t_0)$$

and

$$\begin{aligned} |y_2(t) - y_1(t)| &= \Big| \int_{t_0}^{t} [F(s, y_1(s)) - F(s, y_0(s))]\, ds \Big| \\ &\le \int_{t_0}^{t} |F(s, y_1(s)) - F(s, y_0(s))|\, ds \end{aligned}$$

$$\begin{aligned} &\leq \int_{t_0}^{t} k|y_1(s) - y_0(s)|\,ds \\ &\leq \int_{t_0}^{t} kM(s-t_0)\,ds \quad \text{from the previous display} \\ &= \frac{Mk(t-t_0)^2}{2!}. \end{aligned}$$

Now we carry out an induction. The base case is the estimate carried out above for $|y_1(t)-y_0(t)|$. The estimate for $|y_2(t)-y_1(t)|$ suggests the inductive hypothesis, namely the inequality

$$|y_{n-1}(t) - y_{n-2}(t)| \leq \frac{Mk^{n-2}(t-t_0)^{n-1}}{(n-1)!}.$$

Then we have

$$\begin{aligned} |y_n(t) - y_{n-1}(t)| &\leq \int_{t_0}^{t} |F(s, y_{n-1}(s) - F(t, y_{n-2}(s))|\,ds \\ &\leq \int_{t_0}^{t} k|y_{n-1}(s) - y_{n-2}(s)|\,ds \\ &\leq Mk^{n-1}\int_{t_0}^{t} \frac{(s-t_0)^{n-1}}{(n-1)!}\,ds \qquad \text{by inductive hypothesis} \\ &= \frac{Mk^{n-1}(t-t_0)^n}{n!}, \end{aligned}$$

and the induction is complete. The argument when $t_0 - a' \leq t \leq t_0$ is completely similar, and the form of the estimate for the two cases combined is

$$|y_n(t) - y_{n-1}(t)| \leq \frac{Mk^{n-1}|t-t_0|^n}{n!} \qquad \text{for } |t-t_0| \leq a'.$$

There is no harm in assuming that k is > 0, and consequently

$$|y_n(t) - y_{n-1}(t)| \leq \frac{M}{k}\,\frac{k^n(a')^n}{n!}$$

independently of t. Since $\sum_{n=0}^{\infty} (n!)^{-1}k^n(a')^n = e^{ka'}$ is finite, the M test applies and shows that our series converges uniformly.

Thus $y_N(t)$ converges uniformly for $|t-t_0| \leq a'$, necessarily to a continuous function. We call this function $y(t)$. For $|t-t_0| \leq a'$, we have

$$\begin{aligned} \int_{t_0}^{t} F(s, y(s))\,ds &= \int_{t_0}^{t} [F(s, y(s)) - F(t, y_N(s))]\,ds + \int_{t_0}^{t} F(s, y_N(s))\,ds \\ &= \int_{t_0}^{t} [F(s, y(s)) - F(s, y_N(s))]\,ds + y_{N+1}(t) - y_0. \end{aligned}$$

On the right side, we have $\lim_N [y_{N+1}(t) - y_0] = y(t) - y_0$. Because of the Lipschitz condition the absolute value of the first term on the right side is

$$\leq a'k \sup_{|t-t_0|\leq a'} |y(t) - y_N(t)|,$$

and this tends to 0 as n tends to infinity. Thus

$$\int_{t_0}^{t} F(s, y(s))\, ds = y(t) - y_0,$$

and $y(t)$ is a continuous solution of the integral equation. □

Theorem 4.2 (uniqueness theorem). Let D be a nonempty open set in $\mathbb{R}^1 \times \mathbb{C}^n$, let (t_0, y_0) be in D, and suppose that $F : D \to \mathbb{C}^n$ is a continuous function such that $F(t, y)$ satisfies a Lipschitz condition in the y variable. For any $a'' > 0$, there exists at most one solution $y(t)$ to the system

$$y' = F(t, y)$$

on the open interval $|t - t_0| < a''$ satisfying the initial condition

$$y(t_0) = y_0.$$

PROOF. As in the proof of Theorem 4.1, it is enough to prove uniqueness for the integral equation. Suppose that $y(t)$ and $z(t)$ are two solutions for $|t - t_0| < a''$. Fix $\epsilon > 0$. Then $|y(t)-z(t)|$ is bounded by some constant C for $|t-t_0| \leq a''-\epsilon$, and F is assumed to satisfy a Lipschitz condition $|F(t, y_1)-F(t, y_2)| \leq k|y_1-y_2|$ on D.

We argue as in the proof of Theorem 4.1, working first for $t_0 \leq t$ and starting from

$$|y(t) - z(t)| \leq C$$

and from

$$\begin{aligned} |y(t) - z(t)| &= \Big| \int_{t_0}^{t} [F(s, y(s)) - F(s, z(s))]\, ds \Big| \\ &\leq \int_{t_0}^{t} |F(s, y(s)) - F(s, z(s))|\, ds \\ &\leq \int_{t_0}^{t} k|y(s) - z(s)|\, ds \\ &\leq Ck(t - t_0). \end{aligned}$$

Inductively we suppose that

$$|y(t) - z(t)| \le \frac{Ck^{n-1}(t-t_0)^{n-1}}{(n-1)!}.$$

Then

$$\begin{aligned} |y(t) - z(t)| &\le \int_{t_0}^{t} |F(s, y(s)) - F(s, z(s))|\, ds \\ &\le \int_{t_0}^{t} k|y(s) - z(s)|\, ds \\ &\le Ck^n \int_{t_0}^{t} \frac{(s-x_0)^{n-1}}{(n-1)!}\, ds = \frac{Ck^n(t-t_0)^n}{n!}, \end{aligned}$$

and thus $|y(t) - z(t)| \le C(n!)^{-1}k^n(t-t_0)^n$ for all n. A similar estimate is valid for $t \le t_0$, and the combined estimate is

$$|y(t) - z(t)| \le \frac{Ck^n|t-t_0|^n}{n!}.$$

Since $\sum C(n!)^{-1}k^n|t-t_0|^n$ converges, the individual terms tend to 0. Therefore $y(t) = z(t)$ for $|t - t_0| \le a'' - \epsilon$. Since ϵ is arbitrary, $y(t) = z(t)$ for $|t - t_0| < a''$. □

3. Dependence on Initial Conditions and Parameters

In abstract settings where the existence and uniqueness theorems play a role, it is frequently of interest to know how the unique solution depends on the initial data (t_0, y_0) such that $y(t_0) = y_0$. To quantify this dependence, let us write the unique solution corresponding to $y' = F(t, y)$ as $y(t, t_0, y_0)$ rather than $y(t)$. We continue to use y' to indicate the derivative in the t variable even though the differentiation is now actually a partial derivative.

Theorem 4.3. Let D be a nonempty open set in $\mathbb{R}^1 \times \mathbb{C}^n$, let (t, y^*) be in D, and suppose that $F : D \to \mathbb{C}^n$ is a continuous function such that $F(t, y)$ satisfies a Lipschitz condition in the y variable. Let R be a compact set in $\mathbb{R}^1 \times \mathbb{C}^n$ of the form

$$R = \big\{(t, y) \,\big|\, |t - t^*| \le a \text{ and } |y - y^*| \le b\big\},$$

suppose that R is contained in D, and let M be an upper bound for $|F|$ on R. Put $a' = \min\{a, b/M\}$. If $|t_0 - t^*| < a'/2$ and $|y_0 - y^*| < b/2$, then there exists a unique solution $t \mapsto y(t, t_0, y_0)$ on the interval $|t - t_0| < a'/2$ to the system and initial data

$$y' = F(t, y) \quad \text{and} \quad y(t_0, t_0, y_0) = y_0,$$

and the function $(t, t_0, y_0) \mapsto y(t, t_0, y_0)$ is continuous on the open set

$$U = \big\{(t, t_0, y_0) \,\big|\, |t - t_0| < a'/2,\ |t_0 - t^*| < a'/2,\ |y_0 - y^*| < b/2\big\}.$$

If F is smooth on D, then $(t, t_0, y_0) \mapsto y(t, t_0, y_0)$ is smooth on U.

REMARK. It is customary to summarize the result about continuity qualitatively by saying that the unique solution depends continuously on the initial data.

PROOF OF CONTINUITY. Let us first check that there is indeed a unique solution for each pair (t_0, y_0) in question and that its graph, as a function of t, lies in

$$R' = \big\{(t, y) \,\big|\, |t - t^*| \le a' \text{ and } |y - y^*| \le b\big\}.$$

For this purpose, fix t_0 and y_0 with $|t_0 - t^*| \le a'/2$ and $|y_0 - y^*| \le b/2$. Use of the triangle inequality shows that the closed set with $|t - t_0| < a'/2$ and $|y - y_0| < b/2$ lies within R. Thus $|F| \le M$ on this set. Theorem 4.1 shows that there exists a solution with graph in this smaller set for $|t - t_0| < a''$, where $a'' = \min\{a'/2, (b/2)/M\}$. Now

$$\min\{a'/2, b/(2M)\} = \tfrac{1}{2}\min\{a', b/M\} = \tfrac{1}{2}a',$$

and hence there exists a solution for $|t - t_0| < a'/2$ with graph in R. This solution $y(t, t_0, y_0)$ is unique by Theorem 4.2, and it is the result of the construction in the proof of Theorem 4.1.

The idea is to trace through the construction in the proof of Theorem 4.1 and to see that the function $(t, t_0, y_0) \mapsto y(t, t_0, y_0)$ is the uniform limit of explicit continuous functions on U. Imitating a part of the proof of Theorem 4.1, we define, for (t, t_0, y_0) in U,

$$\begin{aligned}
y_0(t, t_0, y_0) &= y_0,\\
y_1(t, t_0, y_0) &= y_0 + \int_{t_0}^{t} F(s, y_0(s, t_0, y_0))\, ds,\\
&\;\vdots\\
y_m(t, t_0, y_0) &= y_0 + \int_{t_0}^{t} F(s, y_{m-1}(s, t_0, y_0))\, ds.
\end{aligned}$$

We shall show by induction that $y_n(t, t_0, y_0)$ is continuous on U. Certainly $y_0(t, t_0, y_0)$ is continuous on U.

For the inductive step we need a preliminary calculation. Let I_1 be the closed interval between t_0 and t, and let I_2 be the closed interval between t_0' and t'. Suppose we have two functions f_1 and f_2 of a variable s such that

(i) f_1 is defined for s between t_0 and t with $|f_1| \le M$ there,
(ii) f_2 is defined for s between t_0' and t' with $|f_2| \le M$ there, and
(iii) $|f_1(s) - f_2(s)| \le \epsilon$ on their common domain.

If a' is $\ge$ the maximum distance among t_0, t, t_0', t', let us show that

$$\Big|\int_{t_0}^{t} f_1(s)\, ds - \int_{t_0'}^{t'} f_2(s)\, ds\Big| \le M(|t_0 - t_0'| + |t - t'|) + a'\epsilon. \qquad (*)$$

To show this for all possible order relations on the set $\{t_0, t_0', t, t'\}$, we observe that there is no loss of generality in assuming that t_0 is the smallest member of the set. There are then six cases.

Case 1. $t_0 \le t_0' \le t' \le t$, so that (iii) applies on $[t_0', t']$. Then

$$\int_{t_0}^{t} f_1(s)\,ds - \int_{t_0'}^{t'} f_2(s)\,ds = \int_{t_0}^{t_0'} f_1(s)\,ds + \int_{t_0'}^{t'} (f_1(s) - f_2(s))\,ds + \int_{t'}^{t} f_1(s)\,ds$$

and hence

$$\Big| \int_{t_0}^{t} f_1(s)\,ds - \int_{t_0'}^{t'} f_2(s)\,ds \Big| \le M|t_0' - t_0| + \epsilon|t' - t_0'| + M|t - t'|.$$

Therefore $(*)$ holds in this case.

Case 2. $t_0 \le t_0' \le t \le t'$, so that (iii) applies on $[t_0', t]$. Then

$$\int_{t_0}^{t} f_1(s)\,ds - \int_{t_0'}^{t'} f_2(s)\,ds = \int_{t_0}^{t_0'} f_1(s)\,ds + \int_{t_0'}^{t} (f_1(s) - f_2(s))\,ds - \int_{t}^{t'} f_2(s)\,ds,$$

and hence

$$\Big| \int_{t_0}^{t} f_1(s)\,ds - \int_{t_0'}^{t'} f_2(s)\,ds \Big| \le M|t_0' - t_0| + \epsilon|t - t_0'| + M|t' - t|.$$

Therefore $(*)$ holds in this case.

Case 3. $t_0 \le t \le t' \le t_0'$. Then

$$\Big| \int_{t_0}^{t} f_1(s)\,ds \Big| \le M|t - t_0| \le M(|t_0' - t_0| - |t_0' - t'|)$$

and

$$\Big| \int_{t_0'}^{t'} f_2(s)\,ds \Big| \le M|t_0' - t'|,$$

so that $(*)$ holds in this case.

Case 4. $t_0 \le t' \le t_0' \le t$. Then

$$\Big| \int_{t_0}^{t} f_1(s)\,ds \Big| \le M|t - t_0| = M(|t_0' - t_0| + |t - t_0'|)$$

and

$$\Big| \int_{t_0'}^{t'} f_2(s)\,ds \Big| \le M|t' - t_0'| = M(|t - t'| - |t - t_0'|),$$

so that $(*)$ holds in this case.

Case 5. $t_0 \le t \le t_0' \le t'$. Then

$$\Big|\int_{t_0}^{t} f_1(s)\,ds\Big| \le M|t-t_0| \le M|t_0'-t_0|$$

and
$$\Big|\int_{t_0'}^{t'} f_2(s)\,ds\Big| \le M|t'-t_0'| \le M|t'-t|,$$

so that $(*)$ holds in this case.

Case 6. $t_0 \le t' \le t \le t_0'$. Then

$$\Big|\int_{t_0}^{t} f_1(s)\,ds\Big| \le M|t-t_0| = M(|t_0'-t_0| - |t_0'-t|)$$

and
$$\Big|\int_{t_0'}^{t'} f_2(s)\,ds\Big| \le M|t_0'-t'| = M(|t_0'-t| + |t'-t|),$$

so that $(*)$ holds in this case.

With $(*)$ proved we can now proceed with the inductive step to show that $y_n(t, t_0, y_0)$ is continuous on U. Thus assume that $y_{n-1}(t, t_0, y_0)$ is continuous on U. If (t, t_0, y_0) and (t', t_0', y_0') are in U, then

$$\begin{aligned} &y_n(t, t_0, y_0) - y_n(t', t_0', y_0') \\ &\quad = (y_0 - y_0') + \int_{t_0}^{t} F(s, y_{n-1}(s, t_0, y_0))\,ds - \int_{t_0'}^{t'} F(s, y_{n-1}(s, t_0', y_0'))\,ds \\ &\quad = (y_0 - y_0') + \int_{t_0}^{t} f_1(s)\,ds - \int_{t_0'}^{t'} f_2(s)\,ds, \end{aligned}$$

where $f_1(s) = F(s, y_{n-1}(s, t_0, y_0))$ and $f_2(s) = F(s, y_{n-1}(s, t_0', y_0'))$. Thus $(*)$ gives

$$|y_n(t, t_0, y_0) - y_n(t', t_0', y_0')| \le |y_0 - y_0'| + M(|t_0 - t_0'| + |t - t'|) + a'\epsilon \quad (**)$$

if ϵ is chosen such that $|f_1(s) - f_2(s)| \le \epsilon$ on the common domain of f_1 and f_2.

Let $\epsilon > 0$ be given, and choose some $\delta > 0$ for uniform continuity of F on the set R. By uniform continuity of y_{n-1}, choose $\eta > 0$ such that

$$|y_{n-1}(s, t_0, y_0) - y_{n-1}(s, t_0', y_0')| < \delta \quad \text{whenever } |(s, t_0, y_0) - (s, t_0', y_0')| < \eta.$$

Then $|(s, t_0, y_0) - (s, t_0', y_0')| < \eta$ implies $|f_1(s) - f_2(s)| \le \epsilon$ on the common domain of f_1 and f_2, and hence $(**)$ holds. Therefore y_n is continuous as a function on U. This completes the induction.

We know that $y_n(t, t_0, y_0)$ converges to a solution $y(t, t_0, y_0)$ uniformly in t if (t_0, y_0) is fixed. Let us see that the convergence is in fact uniform in (t, t_0, y_0). The proof of Theorem 4.1 yielded the estimate

$$|y_n(t, t_0, y_0) - y_{n-1}(t, t_0, y_0)| \leq \frac{M}{k} \frac{k^n (a')^n}{n!},$$

and this is independent of (t, t_0, y_0). Therefore the Weierstrass M test shows that $y_n(t, t_0, y_0)$ converges to $y(t, t_0, y_0)$ uniformly on U. The uniform limit of continuous functions is continuous by Proposition 2.21, and hence $y(t, t_0, y_0)$ is continuous. □

PROOF OF SMOOTHNESS. Under the assumption that F is smooth on D, we are to prove that $y(t, t_0, y_0)$ is smooth on U. We return to the earlier proof of continuity of $y(t, t_0, y_0)$ and show that each $y_n(t, t_0, y_0)$ is smooth. This smoothness is trivial for $n = 0$, we assume inductively that $y_{n-1}(t, t_0, y_0)$ is smooth, and we form

$$y_n(t, t_0, y_0) = y_0 + \int_{t_0}^{t} F(s, y_{n-1}(s, t_0, y_0))\, ds.$$

The function on the right side is the composition of $(t, t_0, y_0) \mapsto (t, t_0, t_0, y_0)$ followed by $(t, t_0, s_0, y_0) \mapsto \int_{t_0}^{t} F(s, y_{n-1}(s, s_0, y_0))\, ds$. The chain rule (Theorem 3.10), the Fundamental Theorem of Calculus (Theorem 1.32), and Proposition 3.28 allow us to compute partial derivatives of this function, and another argument with $(*)$ allows us to see that the partial derivatives are continuous. There is no difficulty in iterating this argument, and we conclude that $y_n(t, t_0, y_0)$ is smooth.

The same argument in the proof of Theorem 4.1 that enabled us to estimate the size of $y_n(t, t_0, y_0) - y_{n-1}(t, t_0, y_0)$ allows us to estimate any iterated partial derivative of this difference. New constants enter the estimate, but the qualitative result is the same, namely that any iterated partial derivative of $y_n(t, t_0, y_0)$ converges uniformly to that same iterated partial derivative of $y(t, t_0, y_0)$. Applying Theorem 1.23, we see that $y(t, t_0, y_0)$ is smooth. □

CONCLUDING REMARK. Sometimes a given system $y' = F(t, y)$ with initial condition $y(t_0) = y_0$ involves parameters in the definition of F, so that effectively the system is $y' = F(t, y, \lambda_1, \dots, \lambda_k)$. A natural problem is to find conditions under which the dependence of the solution on the k parameters is continuous or smooth. The answer is that this problem can be reduced to the problem addressed by Theorem 4.3. We simply introduce k additional variables z_j, one for each parameter λ_j, together with new equations $z_j' = 0$ and new initial conditions $z_j(t_0) = \lambda_j$.

4. Integral Curves

If U is an open subset of $\mathbb{R}^n$, then a **vector field** on U may be defined as a function $X : U \to \mathbb{R}^n$. The vector field is **smooth** if X is a smooth function. In classical notation, X is written $X = \sum_{j=1}^n a_j(x_1, \dots, x_n) \frac{\partial}{\partial x_j}$, and the function carries $(x_1, \dots, x_n)$ to $(a_1(x_1, \dots, x_n), \dots, a_n(x_1, \dots, x_n))$. The traditional geometric interpretation of X is to attach to each point p of U the vector $X(p)$ as an arrow based at p. This interpretation is appropriate, for example, if X represents the velocity vector at each point in space of a time-independent fluid flow.

In Chapter II we defined the term "path" in a metric space to mean a continuous function from a closed bounded interval of $\mathbb{R}^1$ into the metric space. Then in Chapter III we used the term "curve" to refer to any continuous function from an interval, not necessarily closed, into $\mathbb{R}^n$. In this chapter the term **curve** in a metric space will be used to refer to a continuous function from an open interval of $\mathbb{R}^1$ into the metric space.

A standard problem in connection with vector fields on an open subset U of $\mathbb{R}^2$ is to try to draw curves within U with the property that the tangent vector to the curve at any point matches the arrow for the vector field. An illustration occurs in Figure 4.2. This section abstracts and generalizes this kind of curve.

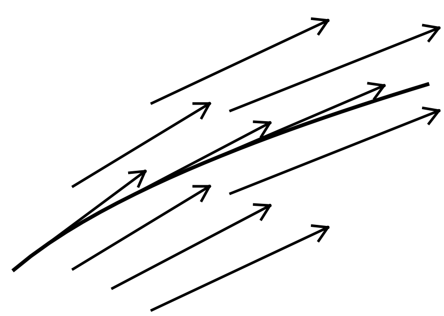


FIGURE 4.2. Integral curve of a vector field.

Let $X : U \to \mathbb{R}^n$ be a smooth vector field on U. A curve $c(t)$ is an **integral curve** for X if c is smooth (i.e., of class C^∞) and $c'(t) = X(c(t))$ for all t in the domain of c. Depending on one's interpretation of the informal wording in the previous paragraph, the present definition is perhaps more demanding than the definition given for $\mathbb{R}^2$ above: the expression $c'(t)$ involves both magnitude and direction, and the present definition insists that both ingredients match with $X(c(t))$, not just the direction.

Proposition 4.4. Let $X : U \to \mathbb{R}^n$ be a smooth vector field on an open subset U of $\mathbb{R}^n$, and let p be in U. Then there exist an $\varepsilon > 0$ and an integral curve

$c : (-\varepsilon, \varepsilon) \to U$ such that $c(0) = p$. Any two integral curves c and d for X having $c(0) = d(0) = p$ coincide on the intersection of their domains.

PROOF. Apart from the smoothness the first conclusion is just a restatement of a special case of Theorem 4.1 in different notation. The conditions on c are that c be a solution of $c' = X(c)$ and that $c(0) = p$. The existence of a solution is immediate from Theorem 4.1 if we put $F = X$, $c = y$, $t_0 = 0$, and $y_0 = p$. The way in which this application of Theorem 4.1 is a special case and not the general case is that F is independent of t here. The smoothness of c follows from Theorem 4.3, and the uniqueness follows from Theorem 4.2. □

The interest is not only in Proposition 4.4 in isolation but also in what happens to the integral curves when X is part of a family of vector fields.

Proposition 4.5. Let $X^{(1)}, \ldots, X^{(m)}$ be smooth vector fields on an open subset U of $\mathbb{R}^n$, let p be in U, and let V be a bounded open neighborhood of 0 in $\mathbb{R}^m$. For λ in V, put $X_\lambda = \sum_{j=1}^m \lambda_j X^{(j)}$. Then there exist an $\varepsilon > 0$ and a system of integral curves $c(t, \lambda)$, defined for $t \in (-\varepsilon, \varepsilon)$ and $\lambda \in V$, such that $c(\,\cdot\,, \lambda)$ is an integral curve for X_λ with $c(0, \lambda) = p$. Each curve $c(t, \lambda)$ is unique, and the function $c : (-\varepsilon, \varepsilon) \times V \to U$ is smooth. If $m = n$, if the vectors $X^{(1)}(p), \ldots, X^{(n)}(p)$ are linearly independent, and if δ is any positive number less than ε, then the Jacobian matrix of $\lambda \mapsto c(\delta, \lambda)$ at $\lambda = 0$ is nonsingular.

REMARK. In the final conclusion of this proposition, the open neighborhood of 0 within V is allowed to depend on δ. It follows from the final conclusion that the Inverse Function Theorem (Theorem 3.17) and its corollary (Corollary 3.21) are applicable to the mapping $\lambda \mapsto c(\delta, \lambda)$ at $\lambda = 0$. These results produce a smooth inverse function carrying an open subneighborhood of 0 within V onto an open subneighborhood of p of U. In effect the inverse function assigns locally defined coordinates in λ space to a neighborhood of U.

PROOF. We set up the system of equations $c' = X_\lambda \circ c$, i.e.,

$$c_i' = \sum_{j=1}^{m} \lambda_j X_i^{(j)}(c),$$

with initial condition $c(0) = p$. This is a smooth system of the kind considered in Theorem 4.3, and the λ_j with $1 \le j \le m$ are parameters. The parameters are handled by the concluding remark in Section 3: we obtain unique solutions $c(t, \lambda)$ for t in some open interval $(-\varepsilon, \varepsilon)$, and $(t, \lambda) \mapsto c(t, \lambda)$ is smooth.

Now suppose that $m = n$, that the vectors $X^{(1)}(p), \ldots, X^{(n)}(p)$ are linearly independent, and that $0 < \delta < \varepsilon$. The function c satisfies

$$c_i'(t, \lambda) = \sum_{j=1}^{n} \lambda_j X_i^{(j)}(c(t, \lambda)), \tag{$*$}$$

and we use this information to compute the Jacobian matrix of $\lambda \mapsto c(\delta, \lambda)$ at $\lambda = 0$. The Fundamental Theorem of Calculus, Proposition 3.28, and $(*)$ give

$$\begin{aligned}
\frac{\partial c_i}{\partial \lambda_j}(\delta, \lambda) &= \frac{\partial c_i}{\partial \lambda_j}(0, \lambda) + \int_0^\delta \frac{\partial c_i'}{\partial \lambda_j}(t, \lambda)\, dt \\
&= \frac{\partial c_i}{\partial \lambda_j}(0, \lambda) + \frac{\partial}{\partial \lambda_j} \int_0^\delta c_i'(t, \lambda)\, dt \\
&= \frac{\partial c_i}{\partial \lambda_j}(0, \lambda) + \int_0^\delta X_i^{(j)}(c(t, \lambda))\, dt + \sum_{k=1}^n \lambda_k \frac{\partial}{\partial \lambda_j} \int_0^\delta X_i^{(k)}(c(t, \lambda))\, dt.
\end{aligned}$$

Now $c_i(0, \lambda) = p_i$ for all λ, and hence $\frac{\partial c_i}{\partial \lambda_j}(0, \lambda)\big|_{\lambda=0} = 0$. Also, $c(t, 0)$ is constant in t by $(*)$, and the constant is $c(0, 0) = p$. Finally when λ is set equal to 0 in the term $\sum_{k=1}^n \lambda_k \frac{\partial}{\partial \lambda_j} \int_0^\delta X_i^{(k)}(c(t, \lambda))\, dt$, each λ_k becomes 0, and thus the whole term becomes 0. Thus the above equation specializes at $\lambda = 0$ to

$$\frac{\partial c_i}{\partial \lambda_j}(\delta, \lambda)\Big|_{\lambda=0} = 0 + \delta X_i^{(j)}(p) + 0.$$

The vectors $X^{(j)}(p)$ are by assumption linearly independent, and hence the determinant of the matrix $[X_i^{(j)}(p)]$ is not 0. Consequently the Jacobian matrix $\lambda \mapsto c(\delta, \lambda)$ at $\lambda = 0$ is nonsingular if $\delta \neq 0$. □

5. Linear Equations and Systems, Wronskian

Recall from Section 1 that a **linear ordinary differential equation** is defined to be an equation of the type

$$a_n(t) y^{(n)} + a_{n-1}(t) y^{(n-1)} + \cdots + a_1(t) y' + a_0(t) y = q(t)$$

with real or complex coefficients. The equation is **homogeneous** if q is the 0 function, **inhomogeneous** in general. In order for the existence and uniqueness theorems of Section 1 to apply, we need to be able to solve for $y^{(n)}$ and have all coefficients be continuous afterward. Thus we assume that $a_n(t) = 1$ and that $a_{n-1}(t), \ldots, a_0(t)$ and $q(t)$ are continuous on some open interval.

Even in simple cases, the theory is helped by converting a single equation to a system of first-order equations. In Section 1 we saw an indication that a way to make this conversion is to put

$$\begin{aligned} y_1 &= y \\ y_2 &= y' \\ &\vdots \\ y_{n-1} &= y^{(n-2)} \\ y_n &= y^{(n-1)} \end{aligned} \qquad \text{and get} \qquad \begin{aligned} y_1' &= y_2 \\ y_2' &= y_3 \\ &\vdots \\ y_{n-1}' &= y_n \\ y_n' &= -a_0(t)y_1 - \cdots - a_{n-1}y_n + q(t). \end{aligned}$$

If we change the meaning of the symbol y from a scalar-valued function to the vector-valued function $y = (y_1, \ldots, y_n)$, then we arrive at the system

$$y' = A(t)y + Q(t),$$

where $A(t)$ is the n-by-n matrix of continuous functions given by

$$A(t) = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0(t) & -a_1(t) & -a_2(t) & \cdots & -a_{n-1}(t) \end{pmatrix}$$

and $Q(t)$ is the n-component column vector of continuous functions given by

$$Q(t) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ q(t) \end{pmatrix}.$$

In a general linear first-order system of the kind we shall study, $A(t)$ can be any n-by-n matrix of continuous functions and $Q(t)$ can be any column vector of continuous functions; thus the first-order system obtained by conversion of a single n^{th}-order equation is of quite a special form among all first-order linear systems.

For a system $y' = A(t)y + Q(t)$ as above, the Lipschitz condition for the function $F(t, y) = A(t)y + Q(t)$ is automatic, since

$$|F(t, y) - F(t, y^*)| = |A(t)(y - y^*)| \le \|A(t)\||y - y^*|$$

and since the function $t \mapsto \|A(t)\|$ is bounded on any compact subinterval of our domain interval. By the uniqueness theorem (Theorem 4.2), a unique solution to the system is determined by data (t_0, y_0), the local solution corresponding to (t_0, y_0) being the one satisfying the initial condition that the vector $y(t_0)$ equal the vector y_0. If we track down what these data correspond to in the case of a single n^{th}-order equation, we see that a unique solution to a single n^{th}-order equation of the kind described above is determined by initial values at a point t_0 for the scalar-valued solution and all its derivatives through order $n - 1$.

First-order linear systems of size one can be solved explicitly in terms of known functions and integrations. Specifically the single homogeneous first-order equation $y' = a(t)y$ is solved by $y(t) = c\exp\big(\int^t a(s)\,ds\big)$, and the solution of a single inhomogeneous first-order equation can be reduced to the homogeneous case by the variation-of-parameters formula that appears later in this section. However, there need not be such an elementary solution of a first-order linear system of size two, not even a system that comes from a single second-order equation. Elementary solutions exist when the coefficient matrix has constants as entries, and we shall address that case in the next two sections. Sometimes one can write down tidy power-series solutions when the coefficient matrix has nonconstant entries, and we shall take up that matter later in the chapter. For now, we develop some general theory about first-order linear systems, beginning with the homogeneous case. The linearity implies that the set of solutions to the system $y' = A(t)y$ on an open interval is a vector space (of vector-valued functions) in the sense that it is closed under addition and scalar multiplication.

Theorem 4.6. Let $y' = A(t)y$ be a homogeneous linear first-order n-by-n system with $A(t)$ continuous for $a < t < b$. Then

(a) any solution on a subinterval (a', b') extends to a solution on the whole interval (a, b),

(b) the dimension of the vector space of solutions on any subinterval (a', b') is exactly n,

(c) if $v_1(t), \dots, v_r(t)$ are solutions on an interval (a', b') and if t_0 is in that interval, then $v_1, \dots, v_r$ are linearly independent functions if and only if the column vectors $v_1(t_0), \dots, v_n(t_0)$ are linearly independent.

PROOF. We begin by proving (c). If $c_1v_1(t) + \cdots + c_rv_r(t)$ is identically 0 for constants $c_1, \dots, c_r$ not all 0, then $c_1v_1(t_0) + \cdots + c_rv_r(t_0) = 0$ for the same constants. Conversely suppose that $c_1v_1(t_0) + \cdots + c_rv_r(t_0) = 0$ for constants not all 0. Put $v(t) = c_1v_1(t) + \cdots + c_rv_r(t)$. Then $v(t)$ and the 0 function are solutions of the system satisfying the same initial conditions—that they are 0 at t_0. By the uniqueness theorem (Theorem 4.2), $v(t)$ is the 0 function. This proves (c).

The upper bound in (b) is immediate from (c) since the dimension of the space of n-component column vectors is n.

Let us prove that n is a lower bound for the dimension in (b) if the interval containing t_0 is sufficiently small. By the existence theorem (Theorem 4.1), there exists a solution $v_j(t)$ on some interval $|t - t_0| < \varepsilon_j$ such that $v_j(t_0) = e_j$. The $v_j(t)$ are then solutions on $|t - t_0| < \varepsilon$ with $\varepsilon = \min\{\varepsilon_1, \dots, \varepsilon_n\}$, and they are linearly independent by (c). Hence the dimension of the space of solutions is at least n on the interval $|t - t_0| < \varepsilon$ or on any subinterval containing t_0.

We are not completely done with proving (b), but let us now prove (a). Let $v(t)$ be a solution on (a', b'). If we have a collection of solutions on different intervals containing (a', b') and each pair of solutions is consistent on their common domain, then the union of the solutions is a solution. Consequently we may assume that $v(t)$ does not extend to a solution on any larger interval. We are to prove that $(a', b') = (a, b)$. Suppose on the contrary that $b' < b$. We use $t_0 = b'$ in the previous paragraph of the proof; the result is that on some interval $|t - b'| < \varepsilon$ with ε sufficiently small and at least small enough so that $a' < b' - \varepsilon$, the space of solutions has dimension n with a basis $\{v_1, \dots, v_n\}$. By (c), the column vectors $v_1(b' - \varepsilon), \dots, v_n(b' - \varepsilon)$ are linearly independent, and thus the restrictions of $v_1, \dots, v_n$ to $(b' - \varepsilon,\ b')$ are linearly independent. The restriction of $v(t)$ to the interval $(b' - \varepsilon,\ b')$ is a solution, and thus there exist constants $c_1, \dots, c_n$ such that

$$v(t) = c_1 v_1(t) + \cdots + c_n v_n(t) \qquad \text{for } b' - \varepsilon < t < b'.$$

But then the function equal to $v(t)$ on (a', b') and equal to $c_1 v_1(t) + \cdots + c_n v_n(t)$ on $(b' - \varepsilon, b' + \varepsilon)$ extends $v(t)$ to a solution on a larger interval and contradicts the maximality of the domain of $v(t)$. This proves that $b' = b$. Similarly we find that $a' = a$. This proves (a).

We return to the unproved part of (b). Fix t_0 in (a', b'). On a subinterval about t_0, the space of solutions has dimension n, as we have already proved. Let $\{v_1, \dots, v_n\}$ be a basis. By (a), we can extend $v_1, \dots, v_n$ to solutions on (a', b'). Then the space of solutions on (a', b') has dimension at least n, and (b) is now completely proved. □

EXAMPLE. Let us illustrate the content of Theorem 4.6 by means of a single second-order equation, namely $y'' + y = 0$. We know that $c_1 \cos t + c_2 \sin t$ is a solution for every pair of constants c_1 and c_2. To convert the equation to a system, we introduce $y_1 = y$ and $y_2 = y'$. The system is then

$$\begin{aligned} y_1' &= y_2, \\ y_2' &= -y_1, \end{aligned}$$

and hence the matrix is $A(t) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, a matrix of constants. The scalar-valued solutions $\cos t$ and $\sin t$ of $y'' + y = 0$ correspond to the vector-valued solutions

$\begin{pmatrix} \cos t \\ -\sin t \end{pmatrix}$ and $\begin{pmatrix} \sin t \\ \cos t \end{pmatrix}$, respectively; each of these has a scalar-valued solution in its first entry and the derivative in the second entry. In either case, both solutions are defined on the interval $(-\infty, +\infty)$. The theorem says that the restrictions of these two functions to any subinterval span the solutions on that subinterval. According to (c), the linear independence of the scalar-valued solutions $\cos t$ and $\sin t$ is reflected by the linear independence of the column vectors $\begin{pmatrix} \cos t_0 \\ -\sin t_0 \end{pmatrix}$ and $\begin{pmatrix} \sin t_0 \\ \cos t_0 \end{pmatrix}$ for any t_0 in $(-\infty, +\infty)$. The latter independence we can see immediately by observing that the matrix $\begin{pmatrix} \cos t_0 & \sin t_0 \\ -\sin t_0 & \cos t_0 \end{pmatrix}$ has determinant equal to 1 and not 0.

The kind of matrix formed in the previous example is a useful tool when generalized to an arbitrary homogeneous linear system, and it has a customary name. Let $v_1(t), \dots, v_n(t)$ be solutions of an n-by-n homogeneous linear system $y' = A(t)y$ with $A(t)$ continuous. The **Wronskian matrix** of $v_1, \dots, v_n$ is the n-by-n matrix whose j^{th} column is v_j. If $v_{i,j}$ denotes the i^{th} entry of the j^{th} solution, then

$$W(t) = \begin{pmatrix} v_{1,1}(t) & \cdots & v_{1,n}(t) \\ \vdots & \ddots & \vdots \\ v_{n,1}(t) & \cdots & v_{n,n}(t) \end{pmatrix}.$$

Since each column of $W(t)$ is a solution, we obtain the matrix identity $W'(t) = A(t)W(t)$.

EXAMPLE, CONTINUED. In the case of the single second-order equation $y'' + y = 0$, we listed two linearly independent scalar-valued solutions as $\cos t$ and $\sin t$. When the equation is converted into a 2-by-2 homogeneous linear system, the Wronskian matrix is

$$W(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}.$$

For a general n^{th}-order equation with $v_1, \dots, v_n$ as scalar-valued solutions, the Wronskian matrix of the associated system is

$$W(t) = \begin{pmatrix} v_1(t) & \cdots & v_n(t) \\ v_1'(t) & \cdots & v_n'(t) \\ \vdots & \ddots & \vdots \\ v_1^{(n-1)}(t) & \cdots & v_n^{(n-1)}(t) \end{pmatrix}.$$

Proposition 4.7. If $v_1(t), \dots, v_n(t)$ are solutions on an interval of an n-by-n homogeneous linear system $y' = A(t)y$ with $A(t)$ continuous, then the following are equivalent:

(a) $v_1, \dots, v_n$ are linearly independent solutions,

(b) $\det W(t)$ is nowhere 0,
(c) $\det W(t)$ is somewhere nonzero.

PROOF. By Theorem 4.6c, (a) here is equivalent to the linear independence of $v_1(t_0), \ldots, v_n(t_0)$, no matter what t_0 we choose, hence is equivalent to the condition $\det W(t_0) \neq 0$, no matter what t_0 we choose. The proposition follows. □

We shall use the Wronskian matrix of a homogeneous system to analyze the solutions of any corresponding inhomogeneous system.

Proposition 4.8. For an inhomogeneous linear system $y' = A(x)y + Q(t)$ with $A(t)$ and $Q(t)$ continuous for $a < t < b$, any solution $y^*(t)$ on a subinterval (a', b') of (a, b) extends to be a solution on (a, b), and the most general solution $y(t)$ is of the form $y(t) = h(t) + y^*(t)$, where $y^*(t)$ is one solution of $y' = A(t)y + Q(t)$ and $h(t)$ is an arbitrary solution of the homogeneous system $y' = A(t)y$.

PROOF. If y^* and y^{**} are two solutions of $y' = A(t)y + Q(t)$ on (a', b'), then $(y^{**} - y^*)'(t) = (A(t)y^{**}(t) + Q(t)) - (A(t)y^*(t) + Q(t)) = A(t)(y^* - y^{**})(t)$, and $h = y^{**} - y^*$ solves $y' = A(t)y$ on (a', b'). Conversely if h solves $y' = A(t)y + Q(t)$ on (a', b'), then

$$\begin{aligned}(y^* + h)'(t) &= y^{*\prime}(t) + h'(t)\\ &= (A(t)y^*(t) + Q(t)) + A(t)h(t) = A(t)(y^* + h)(t) + Q(t),\end{aligned}$$

and $y^* + h$ is a solution of $y' = A(t)y + Q(t)$ on (a', b').

We are left with showing that any solution y^* of $y' = A(t)y + Q(t)$ on (a', b') extends to a solution on (a, b). As in the proof of Theorem 4.6a, we can form unions of functions and thereby assume that y^* cannot be extended to be a solution on a larger interval. The claim is that $(a', b') = (a, b)$. Assuming the contrary, suppose, for example, that $b' < b$. By the existence theorem (Theorem 4.1), there exists a solution $y^{**}(t)$ of $y' = A(t)y + Q(t)$ for $|t - b'| < \varepsilon$ if ε is small enough. By the result of the previous paragraph, $y^*(t) = y^{**}(t) + h(t)$ on $(b' - \varepsilon, b')$ for a suitable choice of h that solves the homogeneous system $y' = A(t)y$ on $(b' - \varepsilon, b')$. Since $y^{**}(t)$ is given as a solution of $y' = A(t)y + Q(t)$ on $(b' - \varepsilon, b' + \varepsilon)$ and since, by Theorem 4.6a, $h(t)$ extends to a solution of $y' = A(t)y$ on $(b' - \varepsilon, b' + \varepsilon)$, we see that $y^{**}(t) + h(t)$ extends to a solution of $y' = A(t)y + Q(t)$ on $(b' - \varepsilon, b' + \varepsilon)$. Then the function equal to $y^*(t)$ on (a', b') and to $y^{**}(t) + h(t)$ on $(b' - \varepsilon, b' + \varepsilon)$ extends $y^*(t)$ to a solution of $y' = A(t)y + Q(t)$ on a larger interval, namely $(a', b' + \varepsilon)$. We obtain a contradiction and conclude that b' must have equaled b. Similarly a' must equal a. Thus every solution of $y' = A(t)y + Q(t)$ on a subinterval extends to all of (a, b), and the proof is complete. □

Theorem 4.9 (variation of parameters). For an inhomogeneous linear system $y' = A(x)y + Q(t)$ with $A(t)$ and $Q(t)$ continuous for $a < t < b$, let $v_1, \ldots, v_n$ be linearly independent solutions of $y' = A(t)y$ on (a, b), and let $W(t)$ be their Wronskian matrix. Then a particular solution y^* of $y' = A(t)y + Q(t)$ on (a, b) is given by

$$y^*(t) = W(t)u(t), \qquad \text{where} \quad W(t)u'(t) = Q(t).$$

That is,

$$y^*(t) = W(t) \int^t W(s)^{-1} Q(s)\, ds.$$

REMARKS. Linearly independent solutions $v_1, \ldots, v_n$ as in the statement exist by Theorem 4.6.

PROOF. For any differentiable vector-valued function $u(t)$, $y^*(t) = W(t)u(t)$ has

$$(y^*)' = W'u + Wu' = AWu + Wu' = Ay^* + Wu'.$$

Thus y^* will have $(y^*)' = Ay^* + Q$ if and only if $Wu' = Q$. Since Proposition 4.7 shows that $W(t)^{-1}$ exists and is continuous, we can solve $Wu' = Q$ for u. $\square$

EXAMPLE, CONTINUED. Now consider the single second-order inhomogeneous linear equation $y'' + y = \tan t$ on the interval $|t| < \pi/2$. We saw that we can take $W(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$. We set up the system

$$\begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} u_1' \\ u_2' \end{pmatrix} = \begin{pmatrix} 0 \\ \tan t \end{pmatrix}$$

of algebraic linear equations and solve for u_1' and u_2':

$$\begin{pmatrix} u_1' \\ u_2' \end{pmatrix} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \begin{pmatrix} 0 \\ \tan t \end{pmatrix} = \begin{pmatrix} -\dfrac{\sin^2 t}{\cos t} \\ \sin t \end{pmatrix}.$$

A vector-valued function with derivative $\begin{pmatrix} u_1' \\ u_2' \end{pmatrix}$ for $|t| < \pi/2$ is

$$\begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix} = \begin{pmatrix} \sin t - \log(1 + \sin t) + \log \cos t \\ -\cos t \end{pmatrix},$$

and we thus take $y^*(t) = (\cos t)u_1(t) + (\sin t)u_2(t)$. The most general solution of the given inhomogeneous equation is therefore $y^*(t) + c_1 \cos t + c_2 \sin t$.

6. Homogeneous Equations with Constant Coefficients

In this section and the next, we discuss first-order homogeneous linear systems with constant coefficients. The system is of the form $y' = Ay$ with A a matrix of constants. A single homogeneous n^{th}-order linear equation with constant coefficients can be converted into such a first-order system and can therefore be handled by the method applicable to all first-order homogeneous linear systems with constant coefficients. But such an equation can be handled more simply in a direct fashion, and we therefore isolate in this section the case of a single n^{th}-order equation. This section and the next will make use of material on polynomials from Section A8 of Appendix A.

The equation to be studied in this section is of the form

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1 y' + a_0 y = 0$$

with coefficients in $\mathbb{C}$. Let us write this equation as $L(y) = 0$ for a suitable linear operator L defined on functions y of class C^n:

$$L = \Big(\frac{d}{dt}\Big)^n + a_{n-1}\Big(\frac{d}{dt}\Big)^{n-1} + \cdots + a_1\Big(\frac{d}{dt}\Big) + a_0.$$

The term a_0 is understood to act as a_0 times the identity operator. Since $\frac{d}{dt}\, e^{rt} = re^{rt}$, we immediately obtain

$$L(e^{rt}) = (r^n + a_{n-1}r^{n-1} + \cdots + a_1 r + a_0)e^{rt}.$$

The polynomial

$$P(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$$

is called the **characteristic polynomial** of the equation, and the formula $L(e^{rt}) = P(r)e^{rt}$ shows that $y(t) = e^{rt}$ is a solution of $L(y) = 0$ if and only if r is a root of the characteristic polynomial. From Section A8 of Appendix A, we know that the polynomial $P(\lambda)$ factors into the product of linear factors $\lambda - r$, the factors being unique apart from their order. Let us list the distinct roots, i.e., the distinct such complex numbers r, as $r_1, \dots, r_k$ with $k \le n$, and let us write m_j for the number of times that $\lambda - r_j$ occurs as a factor of $P(\lambda)$, i.e., the multiplicity of r_j as a root of P. Then we have $\sum_{j=1}^k m_j = n$ and

$$P(\lambda) = \prod_{j=1}^k (\lambda - r_j)^{m_j}.$$

Corresponding to this factorization of P is a factorization of L as

$$L = \prod_{j=1}^k \Big(\frac{d}{dt} - r_j\Big)^{m_j}.$$

On the right side the individual factors commute with each other because differentiation commutes with itself and with multiplication by constants. The following lemma therefore produces n solutions of the given equation $L(y) = 0$.

Lemma 4.10. For $m \geq 1$ and r in $\mathbb{C}$, all the functions $e^{rt}, te^{rt}, \ldots, t^{m-1}e^{rt}$ are solutions of the m^{th}-order differential equation

$$\Big(\frac{d}{dt} - r\Big)^m (y) = 0.$$

PROOF. Direct computation gives $\big(\frac{d}{dt} - r\big)(t^k e^{rt}) = kt^{k-1}e^{rt}$, and hence $\big(\frac{d}{dt} - r\big)^m (t^k e^{rt}) = k(k-1)\cdots(k-m+1)t^{k-m}e^{rt}$. The right side is 0 if $0 \leq k \leq m-1$, and the lemma follows. □

Lemma 4.11. Let $r_1, \ldots, r_N$ be distinct complex numbers, and let m_j be N integers ≥ 1. Then the $\sum_{j=1}^N m_j$ functions

$$e^{r_j t}, te^{r_j t}, \ldots, t^{m_j - 1}e^{r_j t}, \qquad 1 \leq j \leq N,$$

are linearly independent over $\mathbb{C}$.

PROOF. Let $k \geq 1$ be an integer, let r be a complex number, and let $P(t)$ be a polynomial of degree $\leq k-1$. We allow $P(t)$ to be the 0 polynomial. Then

$$\tfrac{d}{dt}[(t^k + P(t))e^{rt}] = r(t^k + P(t))e^{rt} + ((k-1)t^{k-1} + P'(t))e^{rt},$$

from which it follows that

$$\tfrac{d}{dt}[(t^k + P(t))e^{rt}] = (rt^k + Q(t))e^{rt} \tag{$*$}$$

with $Q(t)$ a polynomial of degree $\leq k-1$ or the 0 polynomial.

We shall prove by induction on N that if $P_1, \ldots, P_N$ are polynomials with complex coefficients such that $\sum_{j=1}^N P_j(t)e^{r_j t}$ is the 0 function, then all the P_j are 0 polynomials. For $N = 1$, if $P(t)e^{rt}$ is the 0 function, then $P(t)$ is the 0 function. Since a polynomial of degree $k \geq 0$ has at most k roots, we conclude that P has all coefficients 0. This disposes of the assertion for $N = 1$. Assume the result for $N-1$, and suppose that we are given that $\sum_{j=1}^{N-1} P_j(t)e^{r_j t} + P_N(t)e^{r_N t}$ is the 0 function, where $\{r_1, \ldots, r_{N-1}, r_N\}$ are distinct. Then

$$\sum_{j=1}^{N-1} P_j(t)e^{q_j t} + P_N(t) \tag{$**$}$$

is the 0 function when $q_j = r_j - r_N$ for $j \leq N-1$. If P_N is the 0 polynomial, the inductive hypothesis shows that all P_j with $j \leq N-1$ are 0 polynomials. Otherwise let P_N have degree d, and differentiate $(**)$ $d+1$ times. If $P_j(t)$ for $j \leq N-1$ is the sum of $a_{n_j}t^{n_j}$ plus lower-degree terms, then $(*)$ shows that the result of the differentiation is that

$$\sum_{j=1}^{N-1} \big(a_{n_j}(q_j)^{d+1} t^{n_j} + \text{lower-degree terms}\big) e^{q_j t}$$

is the 0 function. By the inductive hypothesis each a_{n_j} has to be 0, and hence all coefficients of each P_j have to be 0 for $j \le N-1$. Then $P_N(t)$ is identically 0 and must be the 0 polynomial. This completes the induction.

If we are given a linear combination of the functions in the statement of the lemma that equals the 0 function, then we obtain a relation of the form $\sum_{j=1}^{N} P_j(t)e^{r_j t} = 0$, and we have just seen that this relation forces all P_j to be 0 polynomials. This completes the proof. □

Proposition 4.12. Let the differential equation

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1 y' + a_0 y = 0,$$

with complex coefficients, have characteristic polynomial given by $P(\lambda) = \prod_{j=1}^{k} (\lambda - r_j)^{m_j}$ with $r_1, \ldots, r_k$ distinct complex numbers and with the m_j integers ≥ 0 such that $\sum_{j=1}^{k} m_j = n$. Then the n functions

$$e^{r_j t}, te^{r_j t}, \ldots, t^{m_j - 1}e^{r_j t}, \qquad 1 \le j \le k,$$

form a basis over $\mathbb{C}$ of the space of solutions of the given equation on any interval.

PROOF. Lemma 4.10 shows that the functions in question are solutions, Lemma 4.11 shows that they are linearly independent, and Theorem 4.6 shows that the dimension of the space of solutions on any interval is n. Since n linearly independent solutions have been exhibited, they must form a basis of the space of solutions. □

If the equation in Proposition 4.12 happens to have real coefficients, it is meaningful to ask for a basis over $\mathbb{R}$ of the space of real-valued solutions. Since the coefficients are real, we have $L(\bar{y}) = \overline{L(y)}$ for all complex-valued functions y of class C^n, and it follows that the complex conjugate of any complex-valued solution is again a solution. Thus the real and imaginary parts of any complex-valued solution are real-valued solutions. Meanwhile, the characteristic polynomial P of the equation has real coefficients, and it follows that the set of roots of P is closed under complex conjugation. In addition, the multiplicity of a root equals the multiplicity of its complex conjugate. For any integer $k \ge 0$ and complex number $a + bi$ with $b \ne 0$, we have

$$\mathbb{C}t^k e^{(a+bi)t} + \mathbb{C}t^k e^{(a-bi)t} = \mathbb{C}t^k e^{at}\cos bt + \mathbb{C}t^k e^{at}\sin bt.$$

Thus $t^k e^{at}\cos bt$ and $t^k e^{at}\sin bt$ form a basis over $\mathbb{C}$ of the space spanned by $t^k e^{(a+bi)t}$ and $t^k e^{(a-bi)t}$. The functions $t^k e^{at}\cos bt$ and $t^k e^{at}\sin bt$ are real-valued, and thus we obtain a basis over $\mathbb{C}$ consisting of the real-valued solutions of the given equation if we retain the solutions $t^k e^{rt}$ with r real and we replace any pair $t^k e^{(a+bi)t}$ and $t^k e^{(a-bi)t}$ of solutions, $b \neq 0$, by the pair $t^k e^{at}\cos bt$ and $t^k e^{at}\sin bt$.

Let us see that these resulting functions form a basis over $\mathbb{R}$ of the *real* vector space of real-valued solutions. In fact, we know that they are linearly independent over $\mathbb{R}$ because they are linearly independent over $\mathbb{C}$. To see that they span, we take any real-valued solution and expand it as a complex linear combination of these functions. The imaginary part of this expansion exhibits 0 as a linear combination of the given functions, and the coefficients must be 0 by linear independence. Thus the constructed functions form a basis over $\mathbb{R}$ of the space of real-valued solutions.

7. Homogeneous Systems with Constant Coefficients

Having discussed linear homogeneous equations with constant coefficients, let us pass to the more general case of first-order homogeneous linear systems with constant coefficients. We write the system as $y' = Ay$ with A an n-by-n matrix of constants. In principle we can solve the system immediately. Namely, Proposition 3.13c tells us that $\frac{d}{dt}(e^{tA}) = Ae^{tA}$, so that each of the n columns of e^{tA} is a solution of $y' = Ay$. At $t = 0$, e^{tA} reduces to the identity matrix, and thus these n solutions are linearly independent at $t = 0$. By Theorem 4.6 these n solutions form a basis of all solutions on any subinterval (a, b) of $(-\infty, +\infty)$. The solution satisfying the initial condition $y(t_0) = y_0$ is $y(t) = e^{tA}e^{-t_0A}y_0$, which is the particular linear combination $\sum_{j=1}^n c_j e^{tA} e_j$ of the columns of e^{tA} in which c_j is the number $c_j = (e^{-t_0A}y_0)_j$.

In practice it is not so obvious how to compute e^{tA} except in special cases in which the exponential series can be summed entry by entry. Let us write down three model cases of this kind, and ultimately we shall see that we can handle general A by working suitably with these cases.

MODEL CASES.

(1) Let

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ & 0 & 1 & 0 & \cdots & 0 & 0 \\ & & 0 & 1 & \cdots & 0 & 0 \\ & & & \ddots & \ddots & \vdots & \vdots \\ & & & & 0 & 1 & 0 \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{pmatrix}$$

be of size m-by-m with 0's below the main diagonal. Raising C to powers, we see that the $(i, j)^{\text{th}}$ entry of A^k is 1 if $j = i + k$ and is 0 otherwise. Hence

$$e^{tC} = \begin{pmatrix} 0 & t & \frac{1}{2!}t^2 & \frac{1}{3!}t^3 & \cdots & \frac{1}{(m-2)!}t^{m-2} & \frac{1}{(m-1)!}t^{m-1} \\ & 0 & t & \frac{1}{2!}t^2 & \cdots & \frac{1}{(m-3)!}t^{m-3} & \frac{1}{(m-2)!}t^{m-2} \\ & & 0 & t & \cdots & \vdots & \frac{1}{(m-3)!}t^{m-3} \\ & & & \ddots & \ddots & & \vdots \\ & & & & 0 & t & \frac{1}{2!}t^2 \\ & & & & & 0 & t \\ & & & & & & 0 \end{pmatrix}$$

with 0's below the main diagonal.

(2) Let

$$A = \begin{pmatrix} a & 1 & 0 & 0 & \cdots & 0 & 0 \\ & a & 1 & 0 & \cdots & 0 & 0 \\ & & a & 1 & \cdots & 0 & 0 \\ & & & \ddots & \ddots & \vdots & \vdots \\ & & & & a & 1 & 0 \\ & & & & & a & 1 \\ & & & & & & a \end{pmatrix},$$

so that $A = a1 + C$ with C as in the previous case. Since $a1$ and C commute, Proposition 3.13a shows that $e^{tA} = e^{at}e^{tC}$. In other words, e^{tA} is obtained by multiplying every entry of the matrix e^{tC} in the previous case by e^{at}. A matrix of this form A for some complex constant a and for some size m is said to be a **Jordan block**. Thus we know how to form e^{tA} if A is a Jordan block.

(3) Let A be block diagonal with each block being a Jordan block:

$$A = \begin{pmatrix} \text{block \#1} & & & \\ & \text{block \#2} & & \\ & & \ddots & \\ & & & \text{block \#}k \end{pmatrix}.$$

Then

$$e^{tA} = \begin{pmatrix} e^{t\,\text{block \#1}} & & & \\ & e^{t\,\text{block \#2}} & & \\ & & \ddots & \\ & & & e^{t\,\text{block \#}k} \end{pmatrix}.$$

Thus we know how to form e^{tA} if A is block diagonal with each block being a Jordan block. A matrix A of this kind is said to be in **Jordan form**.

The theorem reduces any computation of a matrix e^{tA} to this case.

Theorem 4.13 (Jordan normal form). For any square matrix A with complex entries, there exists a nonsingular complex matrix B such that $B^{-1}AB = J$ is in Jordan form.

REMARKS. This theorem comes from linear algebra, but knowledge of it is beyond the algebra prerequisites for this book. The proof is long and is not in the spirit of this text, and we shall omit it; however, the interested reader can find a proof in many algebra books that treat linear algebra. One such is the author's *Basic Algebra*. As a practical matter, the proof will not give us any additional information, since we already know that e^{tA} yields the solutions to $y' = Ay$ and the only remaining question is to convert the statement of the theorem into an explicit method of computation.

Let us see what Theorem 4.13 accomplishes. The solution of $y' = Ay$ with $y(t_0) = y_0$ is $y(t) = e^{(t-t_0)A}y_0$. Write $B^{-1}AB = J$ as in the proposition. Then Proposition 3.13d gives

$$\begin{aligned} y(t) &= e^{(t-t_0)A}y_0 = B(B^{-1}e^{(t-t_0)A}B)B^{-1}y_0 \\ &= Be^{(t-t_0)B^{-1}AB}B^{-1}y_0 = Be^{(t-t_0)J}B^{-1}y_0. \end{aligned}$$

If we can compute J, then Model Case 3 above tells us what $e^{(t-t_0)J}$ is. If we can compute B also, then we recover $y(t)$ explicitly.

The practical effect is that Theorem 4.13 gives us a method for calculating solutions. The idea behind the method is that the qualitative properties of B and J forced by the theorem are enough to lead us to explicit values of B and J. Let us go through the steps. A concrete example of J is

$$J = \begin{pmatrix} a & 1 & 0 & & & & & & & \\ 0 & a & 1 & & & & & & & \\ 0 & 0 & a & & & & & & & \\ & & & a & 1 & & & & & \\ & & & 0 & a & & & & & \\ & & & & & a & 1 & & & \\ & & & & & 0 & a & & & \\ & & & & & & & a & & \\ & & & & & & & & b & \\ & & & & & & & & & \ddots \end{pmatrix}.$$

It is helpful to know the extent of uniqueness in Theorem 4.13. The matrix J is actually unique up to permuting the order of the Jordan blocks. The matrix B is not at all unique but results from finding bases of certain subspaces of $\mathbb{C}^n$. The first step is to form the **characteristic polynomial**[2] $P(\lambda) = \det(\lambda 1 - A)$ of A. We have

$$\begin{aligned}\det(\lambda 1 - J) &= \det(\lambda 1 - B^{-1}AB) = \det(B^{-1}(\lambda 1 - A)B) \\ &= \det(B)^{-1}\det(\lambda 1 - A)\det(B) = \det(\lambda 1 - A),\end{aligned}$$

and thus J has the same characteristic polynomial as A. The characteristic polynomial of J is just the product of expressions $\lambda - d$ as d runs through the diagonal entries of J. According to Section A8 of Appendix A, the factorization of a polynomial with complex coefficients and with leading coefficient 1 into first-degree expressions $\lambda - c$ is unique up to order, and thus the factorization of $P(\lambda)$ tells us the diagonal entries of J. We still need to know the sizes of the individual Jordan blocks.

The sizes of the Jordan blocks come from computing dimensions of various null spaces—or kernels, in the terminology of linear functions. If a occurs as a diagonal entry of J, think of forming $J - a1$ and its powers, and consider the dimension of the kernel of each power. For example, with the explicit matrix J that is written above, we have

$$J - a1 = \begin{pmatrix} 0 & 1 & 0 & & & & & & \\ 0 & 0 & 1 & & & & & & \\ 0 & 0 & 0 & & & & & & \\ & & & 0 & 1 & & & & \\ & & & 0 & 0 & & & & \\ & & & & & 0 & 1 & & \\ & & & & & 0 & 0 & & \\ & & & & & & & 0 & \\ & & & & & & & & \boxed{\text{nonsingular}} \end{pmatrix},$$

and $\dim\ker(J - a1)$ is the number of Jordan blocks of size ≥ 1 with a on the diagonal, namely 4 in this case. Next we consider $(J - a1)^2$. In this case,

$$(J - a1)^2 = \begin{pmatrix} 0 & 0 & 1 & & & & & & \\ 0 & 0 & 0 & & & & & & \\ 0 & 0 & 0 & & & & & & \\ & & & 0 & 0 & & & & \\ & & & 0 & 0 & & & & \\ & & & & & 0 & 0 & & \\ & & & & & 0 & 0 & & \\ & & & & & & & 0 & \\ & & & & & & & & \boxed{\text{nonsingular}} \end{pmatrix},$$

[2] Many books write the characteristic polynomial as $\det(A - \lambda 1)$, which is the same as the present polynomial if n is even but is its negative if n is odd. The present notation has the advantage that the notions of characteristic polynomial here and in the previous section coincide when an n^{th}-order equation is converted into a first-order system.

and $\dim\ker(J-a1)^2 = 7$. This number arises as the sum of the previous number and the number of Jordan blocks of size ≥ 2 with a on the diagonal. Thus $\dim\ker(J-a1)^2 - \dim\ker(J-a1)$ in general is the number of Jordan blocks of size ≥ 2 with a on the diagonal. Finally we consider $(J-a1)^3$. In this case, the upper left part of $(J-a1)^3$ corresponding to diagonal entry a is all 0, and the lower right part is nonsingular; hence $\dim\ker(J-a1)^3 = 8$. This number arises as the sum of the previous number and the number of Jordan blocks of size ≥ 3 with a on the diagonal. Thus in general, $\dim\ker(J-a1)^3 - \dim\ker(J-a1)^2$ is the number of Jordan blocks of size ≥ 3 with a on the diagonal. In our example, the number $\dim\ker(J-a1)^k$ remains at 8 for all $k \geq 3$ because 8 is the multiplicity of a as a root of $P(\lambda)$, and we are therefore done with diagonal entry a; our computation has shown that the numbers of Jordan blocks of sizes $1, 2, 3, 4, \dots,$ are $1, 2, 1, 0, \dots,$ and a check on the computation is that $1(1)+2(2)+3(1) = 8$.

Of course, we do not have J at our disposal for these calculations, but A yields the same numbers. In fact, we have $B(J-a1)^k B^{-1} = (A-a1)^k$, from which we see that $x \in \ker(A-a1)^k$ if and only if $B^{-1}x \in \ker(J-a1)k$. Hence

$$B(\ker(J-a1)^k) = \ker(A-a1)^k.$$

Since B is nonsingular, the dimension of the kernel of $(J-a1)^k$ equals the dimension of the kernel of $(A-a1)^k$. Consequently

$$\begin{aligned}
\dim\ker(A-a1) &= \#\{\text{Jordan blocks of size} \geq 1 \text{ with } a \text{ on diagonal}\},\\
\dim\ker(A-a1)^2 - \dim\ker(A-a1)&\\
&= \#\{\text{Jordan blocks of size} \geq 2 \text{ with } a \text{ on diagonal}\},\\
\dim\ker(A-a1)^3 - \dim\ker(A-a1)^2&\\
&= \#\{\text{Jordan blocks of size} \geq 3 \text{ with } a \text{ on diagonal}\},\\
&\text{etc.}
\end{aligned}$$

Repeating this argument with the other roots of $P(\lambda)$, we find that we can determine J completely.

Calculating B requires working with vectors rather than dimensions. The columns of B are just $Be_1, \dots, Be_n$, and we seek a way of finding these. Fix attention on a root a of $P(\lambda)$. Consider an index i with $1 \leq i \leq n$, and suppose that the diagonal entry of J in column i is a. From the form of J, we see that either the i^{th} column of $J-a1$ is 0 or else it is e_{i-1}. In the latter case, index $i-1$ corresponds to the same Jordan block. Using the identity $(A-a1)B = B(J-a1)$, we see that either

$$(A-a1)(Be_i) = B(J-a1)e_i = 0$$

or

$$(A-a1)(Be_i) = B(J-a1)e_i = Be_{i-1},$$

and index $i-1$ corresponds to the same Jordan block as index i in the latter case. Thus the vectors Be_i corresponding to the columns with diagonal entry a and with smallest index for a Jordan block lie in $\ker(A-a1)$. They are linearly independent since B is nonsingular, and the number of them is the number of Jordan blocks corresponding to diagonal entry a. We saw that this number equals $\dim\ker(A-a1)$. Hence the vectors Be_i corresponding to the smallest indices going with each Jordan block form a basis of $\ker(A-a1)$.

Similarly

$$(A-a1)^2(Be_i) = B(J-a1)^2e_i = 0$$

or

$$(A-a1)^2(Be_i) = B(J-a1)^2e_i = Be_{i-2},$$

and index $i-2$ corresponds to the same Jordan block as index i in the latter case. Thus the vectors Be_i corresponding to the columns with diagonal entry a and with smallest or next smallest index for a Jordan block lie in $\ker(A-a1)^2$. They are linearly independent since B is nonsingular, and the number of them is the sum of the previously computed number, namely $\dim\ker(A-a1)$, plus the number of Jordan blocks of size ≥ 2 that correspond to diagonal entry a. We saw that this sum equals $\dim\ker(A-a1)^2$. Hence the vectors Be_i corresponding to the two smallest indices going with each Jordan block form a basis of $\ker(A-a1)^2$. The new vectors Be_i are therefore vectors that we adjoin to a basis of $\ker(A-a1)$ to obtain a basis of $\ker(A-a1)^2$.

In setting up these vectors properly, however, we have to correlate the indices studied at the previous step with those being studied now. The relevant formula is that the new indices i have the property $(A-a1)Be_i = Be_{i-1}$. To obtain vectors with this consistency property, we would take a basis S_1 of $\ker(A-a1)$, extend it to a basis S_2 of $\ker(A-a1)^2$, discard the members of S_1, apply $A-a1$ to the members of $S_2 - S_1$, and extend $(A-a1)(S_2-S_1)$ to a basis T_1 of $\ker(A-a1)$. Then $S_2' = (S_2 - S_1)\cup T_1$ is a new basis of $\ker(A-a1)^2$.

We can continue the argument in this way. It is perhaps helpful to read the general discussion of the argument side by side with the explicit example that appears below. We continue to find that the construction of new basis vectors gets in the way of the necessary consistency property with the earlier basis vectors. Thus we really must start with the largest index k such that $\ker(A-a1)^k \neq \ker(A-a1)^{k-1}$. We extend a basis S_{k-1} of $\ker(A-a1)^{k-1}$ to a basis S_k of $\ker(A-a1)^k$, and form

$$(S_k - S_{k-1}) \cup (A-a1)(S_k - S_{k-1}) \cup \cdots \cup (A-a1)^{k-1}(S_k - S_{k-1}).$$

These vectors will be the columns of B corresponding to the largest Jordan blocks with diagonal entry a. The vectors in

$$(A-a1)^2(S_k - S_{k-1}) \cup \cdots \cup (A-a1)^{k-1}(S_k - S_{k-1})$$

are linearly independent in $\ker(A - a1)^{k-2}$; we extend this set to a basis S'_{k-2} of $\ker(A - a1)^{k-2}$, and we extend $S'_{k-2} \cup (A - a1)(S_k - S_{k-1})$ to a basis S'_{k-1} of $\ker(A - a1)^{k-1}$. The adjoined vectors, together with the result of applying powers of $A - a1$ to them, will be the columns of B corresponding to the next largest Jordan blocks with diagonal entry a. The process continues until we obtain a basis of $\ker(A - a1)^k$ with the necessary consistency property throughout. Then we repeat the process for the other roots of $P(\lambda)$ and assemble the result.

EXAMPLE. Let

$$A = \begin{pmatrix} 4 & 1 & -1 \\ -8 & -2 & 2 \\ 8 & 2 & -2 \end{pmatrix}.$$

The characteristic polynomial is $P(\lambda) = \det(\lambda 1 - A) = \lambda^3$, whose factorization is evidently $P(\lambda) = (\lambda - 0)^3$. Computing the kernel of A, we find that $\dim \ker A = 2$, so that there are 2 Jordan blocks. Also, $A^2 = 0$, so that $\dim \ker A^2 = 3$ and the number of blocks of size ≥ 2 is $3 - 2 = 1$. Thus

$$J = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We form a basis of $\ker A$ by solving $A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$. The standard method of row reduction gives $x_1 = -\frac{1}{4}x_2 + \frac{1}{4}x_3$ with x_2 and x_3 arbitrary, so that a basis of $\ker A$ consists of $\begin{pmatrix} -\frac{1}{4} \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} \frac{1}{4} \\ 0 \\ 1 \end{pmatrix}$. We extend this to a basis of $\ker A^2 = \mathbb{C}^3$ by adjoining, for example, the vector $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. Then $Av_1 = \begin{pmatrix} 4 \\ -8 \\ 8 \end{pmatrix}$. The vector Av_1 is in $\ker A$, and we extend it to a basis of $\ker A$ by adjoining, for example, $v_2 = \begin{pmatrix} -1 \\ 4 \\ 0 \end{pmatrix}$. Then v_1, Av_1, v_2 form a basis of $\ker A^2 = \mathbb{C}^3$, and the above general method asks that these vectors be listed in the order Av_1, v_1, v_2. The matrix B is obtained by lining these vectors up as columns:

$$B = \begin{pmatrix} 4 & 1 & -1 \\ -8 & 0 & 4 \\ 8 & 0 & 0 \end{pmatrix}.$$

The result is easy to check. Computation shows that $B^{-1} = \begin{pmatrix} 0 & 0 & \frac{1}{8} \\ 1 & \frac{1}{4} & -\frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$, and then one can carry out the multiplications to verify that $B^{-1}AB = J$.

8. Series Solutions in the Second-Order Linear Case

In this section we shall consider, in some detail, series solutions for two kinds of ordinary differential equations.

The first kind is

$$y'' + P(t)y' + Q(t)y = 0,$$

where $P(t)$ and $Q(t)$ are given by convergent power-series expansions for $|t| < R$:

$$P(t) = a_0 + a_1 t + a_2 t^2 + \cdots ,$$
$$Q(t) = b_0 + b_1 t + b_2 t^2 + \cdots .$$

We seek power-series solutions of the form

$$y(t) = c_0 + c_1 t + c_2 t^2 + \cdots .$$

The same methods and theorem that handle this first kind of equation apply also to n^{th}-order homogeneous linear equations and to first-order homogeneous systems when the leading coefficient is 1 and the other coefficients are given by convergent power series. The second-order case, however, is by far the most important for applications and is sufficiently illustrative that we shall limit our attention to it.

The idea in finding the solutions is to assume that we have a convergent power-series solution $y(t)$ as above, to substitute the series into the equation, and to sort out the conditions that are imposed on the unknown coefficients. Our theorems on power series in Section I.7 guarantee us that the operations of differentiation and multiplication of power series maintain convergence, and thus the result of substituting into the equation is that we obtain an equality of a convergent power series with 0. Corollary 1.39 then shows that all the coefficients of this last power series must be 0, and we obtain recursive equations for the unknown coefficients. There is one theorem about the equations under study, and it tells us that the power series for $y(t)$ that we obtain by these manipulations is indeed convergent; we state and prove this theorem shortly.

Let us go through the steps of finding the solutions. These steps turn out to be clearer when done in complete generality than when done for an example. Thus we shall first make the computation in complete generality, then state and prove the theorem, and finally consider an important example. The expansions of $y(t)$ and its derivatives are

$$y(t) = c_0 + c_1 t + c_2 t^2 + \cdots ,$$
$$y'(t) = c_1 + 2c_2 t + 3c_3 t^2 + \cdots ,$$
$$y''(t) = 2c_2 + 3 \cdot 2c_3 t + 4 \cdot 3c_4 t^2 + \cdots .$$

Substituting all the series into the given equation yields

$$\begin{aligned}(2\cdot 1c_2 + 3\cdot 2c_3t + 4\cdot 3c_4t^2 + \cdots)\\ + (a_0 + a_1t + a_2t^2 + \cdots)(c_1 + 2c_2t + 3c_3t^2 + \cdots)\\ + (b_0 + b_1t + b_2t^2 + \cdots)(c_0 + c_1t + c_2t^2 + \cdots) = 0.\end{aligned}$$

If the series for $y(t)$ converges and if the left side is expanded out, then the coefficients of each power of t must be 0. Thus

$$\begin{aligned}&2\cdot 1c_2 + a_0c_1 + b_0c_0 = 0,\\ &3\cdot 2c_3 + (a_02c_2 + a_1c_1) + (b_0c_1 + b_1c_0) = 0,\\ &4\cdot 3c_4 + (a_03c_3 + a_12c_2 + a_2c_1) + (b_0c_2 + b_1c_1 + b_2c_0) = 0,\\ &\quad\vdots\\ &n(n-1)c_n + (a_0(n-1)c_{n-1} + a_1(n-2)c_{n-2} + \cdots + a_{n-2}c_1)\\ &\qquad + (b_0c_{n-2} + b_1c_{n-3} + \cdots + b_{n-2}c_0) = 0.\end{aligned}$$

These equations tell us that c_0 and c_1 are arbitrary and that $c_2, c_3, \ldots$ are each determined by the previous coefficients. Thus $c_2, c_3, \ldots$ may be computed inductively. Since $c_0 = y(0)$ and $c_1 = y'(0)$, this degree of flexibility is consistent with the existence and uniqueness theorems.

Theorem 4.14. If $P(t)$ and $Q(t)$ are given by convergent power series for $|t| < R$, then any formal power series that satisfies $y'' + P(t)y' + Q(t)y = 0$ converges for $|t| < R$ to a solution. Consequently every solution of this equation on the interval $-R < t < R$ is given by a power series convergent for $|t| < R$.

PROOF. Fix r with $0 < r < R$, and choose some R_1 with $r < R_1 < R$. Let the notation for the power series of P, Q, and y be as above. Theorem 1.37 shows that the series with terms $|a_nR_1^n|$ and $|b_nR_1^n|$ are convergent, and hence the terms are bounded as functions of n. Thus there exists a real number C such that $|a_n| \le C/R_1^n$ and $|b_n| \le C/R_1^n$ for all $n \ge 0$. We shall show that $|c_n| \le M/r^n$ for a suitable M and all $n \ge 0$.

The constant M will be fixed so that a large initial number of terms have $|c_n| \le M/r^n$, and then we shall see that all subsequent terms satisfy the same inequality. To find an M that works, we start from the formula computed above for c_n:

$$\begin{aligned}n(n-1)c_n = -(a_0(n-1)c_{n-1} + a_1(n-2)c_{n-2} + \cdots + a_{n-2}c_1)\\ - (b_0c_{n-2} + b_1c_{n-3} + \cdots + b_{n-2}c_0).\end{aligned}$$

If M works for $0, 1, \dots, n-1$, then

$$\begin{aligned} n(n-1)|c_n| &\le CM(n-1)(R_1^{-0}r^{-(n-1)} + R_1^{-1}r^{-(n-2)} + \cdots + R_1^{-(n-2)}r^{-1}) \\ &\quad + CM(R_1^{-0}r^{-(n-2)} + R_1^{-1}r^{-(n-3)} + \cdots + R_1^{-(n-2)}r^{-0}) \\ &= CM(n-1)r^{-n}r\Big(1 + \frac{r}{R_1} + \cdots + \Big(\frac{r}{R_1}\Big)^{n-2}\Big) \\ &\quad + CMr^{-n}r^2\Big(1 + \frac{r}{R_1} + \cdots + \Big(\frac{r}{R_1}\Big)^{n-2}\Big) \\ &\le r^{-n}(CM)(r(n-1) + r^2)\,\frac{1}{1-(r/R_1)} \end{aligned}$$

and therefore

$$|c_n| \le Mr^{-n}\Big(\frac{CR_1}{R_1 - r}\,\frac{r(n-1)+r^2}{n(n-1)}\Big).$$

For n sufficiently large, the factor in parentheses is ≤ 1. At that point we obtain $|c_n| \le Mr^{-n}$ if $|c_k| \le Mr^{-k}$ for $k < n$, and induction yields the asserted estimate. Thus $\sum c_n t^n$ converges for $|t| < r$. Since r can be arbitrarily close to R, $\sum c_n t^n$ converges for $|t| < R$.

Finally we saw above that c_0 and c_1 are arbitrary and can therefore be matched to any initial data for $y(0)$ and $y'(0)$. Consequently the vector space of power-series solutions convergent for $|t| < R$ has dimension 2. By Theorem 4.6, all solutions on the interval $-R < t < R$ are accounted for. This completes the proof. □

As a practical matter, the recursive expression for c_n becomes increasingly complicated as n increases, and a closed-form expression need not be available. However, in certain cases, something special happens that yields a closed-form expression for c_n. Here is an example.

EXAMPLE. **Legendre's equation** is

$$(1-t^2)y'' - 2ty' + p(p+1)y = 0$$

with p a complex constant. To apply the theorem literally, we should first divide the equation by $(1-t^2)$, and then the power-series expansions of the coefficients will be convergent for $|t| < 1$. The theorem says that we obtain two linearly independent power-series solutions of the equation for $|t| < 1$. To compute them, it is more convenient to work with the equation without making the preliminary division. Then the equation gives us

$$\begin{aligned} &(2c_2 + 3\cdot 2c_3t + 4\cdot 3c_4t^2 + \cdots) - (2c_2t^2 + 3\cdot 2c_3t^3 + 4\cdot 3c_4t^4 + \cdots) \\ &- 2(c_1t + 2c_2t^2 + 3c_3t^3 + 4c_4t^4 + \cdots) + p(p+1)(c_0 + c_1t + c_2t^2 + \cdots) = 0, \end{aligned}$$

which yields the following formulas for the coefficients:

$$
\begin{aligned}
&2c_2 + p(p+1)c_0 = 0,\\
&3\cdot 2c_3 - 2c_1 + p(p+1)c_1 = 0,\\
&4\cdot 3c_4 - 2\cdot 1c_2 - 2\cdot 2c_2 + p(p+1)c_2 = 0,\\
&\quad\vdots\\
&n(n-1)c_n - [(n-2)(n-3) + 2(n-2) - p(p+1)]c_{n-2} = 0.
\end{aligned}
$$

Thus we can write c_n explicitly as a product. We can verify convergence of $\sum c_n t^n$ directly by the ratio test: since

$$\frac{c_n t^n}{c_{n-2}t^{n-2}} = \frac{(n-2)(n-3) + 2(n-2) - p(p+1)}{n(n-1)}\, t^2,$$

we have convergence for $|t| < 1$. Observe that the numerator in the fraction on the right is equal to

$$(n-2)(n-3) + 2(n-2) - p(p+1) = (n-2)(n-1) - p(p+1),$$

and this is 0 when p is an integer ≥ 0 and $n - 2 = p$. Therefore one of the solutions is a polynomial of degree p if p is an integer ≥ 0. Such polynomials, when suitably normalized, are called **Legendre polynomials**.

The second kind of ordinary differential equation for which we shall seek series solutions is

$$t^2 y' + tP(t)y' + Q(t)y = 0,$$

where $P(t)$ and $Q(t)$ are given by convergent power-series expansions for $|t| < R$:

$$
\begin{aligned}
P(t) &= a_0 + a_1 t + a_2 t^2 + \cdots,\\
Q(t) &= b_0 + b_1 t + b_2 t^2 + \cdots.
\end{aligned}
$$

The existence and uniqueness theorems do not apply to this equation on an interval containing $t = 0$ unless t happens to divide $P(t)$ and t^2 happens to divide $Q(t)$. When this divisibility does not occur, the above equation is said to have a **regular singular point** at $t = 0$. The treatment of the corresponding n^{th}-order equation is no different, but we stick to the second-order case because of its relative importance in applications. For this kind of equation, the treatment of first-order systems is more complicated than the treatment of a single equation of n^{th} order.

Actually, the second-order equation above need not have power series solutions. The prototype for the above equation is the equation

$$t^2 y'' + tPy' + Qy = 0$$

with P and Q constant. This equation is known as **Euler's equation** and can be solved in terms of elementary functions. In fact, we make a change of variables by putting $t = e^x$ and $x = \log t$ for $t > 0$. Then we obtain

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt} = \frac{1}{t}\frac{dy}{dx}$$

and

$$\frac{d^2y}{dt^2} = -\frac{1}{t^2}\frac{dy}{dx} + \frac{1}{t}\frac{d}{dt}\Big(\frac{dy}{dx}\Big) = -\frac{1}{t^2}\frac{dy}{dx} + \frac{1}{t}\frac{d^2y}{dx^2}\frac{dx}{dt} = -\frac{1}{t^2}\frac{dy}{dx} + \frac{1}{t^2}\frac{d^2y}{dx^2},$$

and hence the equation becomes

$$\frac{d^2y}{dx^2} + (P-1)\,\frac{dy}{dx} + Qy = 0.$$

This is an equation of the kind considered in Section 6. A solution is e^{st}, where s is a root of the characteristic polynomial $s^2 + (P-1)s + Q = 0$. If the two roots of the characteristic polynomial are distinct, we obtain two linearly independent solutions for $x \in (-\infty, +\infty)$, and these transform back to two solutions t^s of the Euler equation for $t > 0$. If the characteristic equation has one root s of multiplicity 2, then we obtain the two linearly independent solutions e^{sx} and xe^{sx} for $x \in (-\infty, +\infty)$, and these transform back to two solutions x^s and $x^s \log x$ for $x > 0$.

In practice, the technique to solve the Euler equation $t^2y'' + tPy' + Qy = 0$ is to substitute $y(t) = t^s$ and obtain $s(s-1)t^s + sPt^s + Qt^s = 0$. This equation holds if and only if s satisfies

$$s(s-1) + sP + Q = 0,$$

which is called the **indicial equation**.

In the general case of a regular singular point, we proceed by analogy and are led to seek for $t > 0$ a series solution of the form

$$y(t) = t^s(c_0 + c_1 t + c_2 t^2 + \cdots) \qquad \text{with } c_0 \neq 0.$$

Suppose that the power-series part $\sum c_n t^n$ is convergent. We substitute and obtain

$$\begin{aligned} &t^s(c_0 s(s-1) + c_1(s+1)st + c_2(s+2)(s+1)t^2 + \cdots) \\ &\quad + t^s(a_0 + a_1 t + a_2 t^2 + \cdots)(sc_0 + (s+1)c_1 t + (s+2)c_2 t^2 + \cdots) \\ &\quad + t^s(b_0 + b_1 t + b_2 t^2 + \cdots)(c_0 + c_1 t + c_2 t^2 + \cdots) = 0. \end{aligned}$$

Dividing by t^s and setting the coefficient of each power of t equal to 0 gives the equations

$$
\begin{aligned}
&c_0 s(s-1) + sc_0 a_0 + c_0 b_0 = 0,\\
&c_1(s+1)s + ((s+1)c_1 a_0 + sc_0 a_1) + (c_1 b_0 + c_0 b_1) = 0,\\
&c_2(s+2)(s+1) + ((s+2)c_2 a_0 + \cdots) + (c_2 b_0 + \cdots) = 0,\\
&\quad\vdots\\
&c_n(s+n)(s+n-1) + ((s+n)c_n a_0 + \cdots) + (c_n b_0 + \cdots) = 0.
\end{aligned}
$$

Since c_0 is by assumption nonzero, we can divide the first equation by it, and we obtain

$$s(s-1) + a_0 s + b_0 = 0,$$

which is the **indicial equation** for $t^2 y' + tP(t)y' + Q(t)y = 0$. This determines the exponent s. Then c_0 is arbitrary, and all subsequent c_n's can be found recursively, provided the coefficient of c_n in the $(n+1)^{\text{st}}$ equation above is never 0 for $n \geq 1$, i.e., provided

$$(s+n)(s+n+1) + (s+n)a_0 + b_0 \neq 0 \qquad \text{for } n \geq 1.$$

In other words, we can solve recursively for all c_n in terms of c_0 provided $s+n$ does not satisfy the indicial equation for any $n \geq 1$. We summarize as follows.

Proposition 4.15. If $P(t)$ and $Q(t)$ are given by convergent power series for $|t| < R$, then the following can be said about formal series solutions of $t^2 y'' + tP(t)y' + Q(t)y = 0$ of the type $t^s(c_0 + c_1 t + c_2 t^2 + \cdots)$ with $c_0 \neq 0$:

(a) If the indicial equation has distinct roots not differing by an integer, then there are formal solutions of the type $x^s(c_0 + c_1 t + c_2 t^2 + \cdots)$ for each root s of the indicial equation.

(b) If the indicial equation has roots $r_1 \leq r_2$ with $r_2 - r_1$ equal to an integer, then there is a 1-parameter family of formal solutions of the type $t^{r_2}(c_0 + c_1 t + c_2 t^2 + \cdots)$ with $c_0 \neq 0$. If $r_1 < r_2$ in addition, there may be formal solutions $t^{r_1}(c_0 + c_1 t + c_2 t^2 + \cdots)$ with $c_0 \neq 0$, as there are for an Euler equation.

Theorem 4.16. If $P(t)$ and $Q(t)$ are given by convergent power series for $|t| < R$, then all formal series solutions of $t^2 y'' + tP(t)y' + Q(t)y = 0$ of the type $t^s(c_0 + c_1 t + c_2 t^2 + \cdots)$ with $c_0 \neq 0$ converge for $0 < t < R$ to a function that is a solution for $0 < t < R$.

PROOF. As in the proof of Theorem 4.14, fix r with $0 < r < R$, and choose some R_1 with $r < R_1 < R$. Let the series expansions of $P(t)$ and $Q(t)$ be as above, so that there is a number C with $|a_n| \le C/R_1^n$ and $|b_n| \le C/R_1^n$. Choose N large enough so that

$$\frac{Cr/R_1}{1-r/R_1}\left(\frac{|s|+n+1}{|(s+n)(s+n+1)+a_0(s+n)+b_0|}\right) \le 1 \qquad (*)$$

for $n \ge N$. Then choose M such that $|c_n| \le M/r^n$ for $n \le N$. We shall prove by induction on n that $|c_n| \le M/r^n$ for all n. The base case of the induction is $n = N$, where the inequality holds by definition of M. Suppose it holds for $1, \dots, n-1$. The formula for c_n is

$$\begin{aligned} &c_n\big((s+n)(s+n-1)+a_0(s+n)+b_0\big) \\ &\qquad = -[(s+n-1)a_1c_{n-1}+\cdots+sa_nc_0] - [b_1c_{n-1}+\cdots+b_nc_0]. \end{aligned}$$

Our inductive hypothesis gives

$$\begin{aligned} |c_n||(s+n)(s+n-1)&+a_0(s+n)+b_0| \\ &\le CM(|s|+n)(R_1^{-1}r^{-(n-1)}+\cdots+R_1^{-n}r^0) \\ &\quad + CM(R_1^{-1}r^{-(n-1)}+\cdots+R_n^{-n}r^0) \\ &= CM(|s|+n+1)r^{-n}\Big(\frac{r}{R_1}+\cdots+\frac{r^n}{R_1^n}\Big) \\ &\le Mr^{-n}\left[C(|s|+n+1)\left(\frac{r/R_1}{1-r/R_1}\right)\right]. \end{aligned}$$

Thus

$$|c_n| \le Mr^{-n}\left[\frac{Cr/R_1}{1-r/R_1}\left(\frac{|s|+n+1}{|(s+n)(s+n+1)+a_0(s+n)+b_0|}\right)\right] \le Mr^{-n},$$

the second inequality holding by $(*)$, and the induction is complete.

It follows that $\sum c_nt^n$ converges for $|t| < r$. Since r can be arbitrarily close to R, $\sum c_nt^n$ converges for $|t| < R$. This completes the proof. □

EXAMPLE. **Bessel's equation** of order p with $p \ge 0$. This is the equation

$$t^2y'' + ty' + (t^2 - p^2)y = 0.$$

It has $P(t) = 1$ and $Q(t) = t^2 - p^2$, both with infinite radius of convergence. The indicial equation in general is $s(s-1) + a_0s + b_0 = 0$ and hence is

$$s(s-1) + s - p^2 = 0$$

in this case. Thus $s = \pm p$. Theorem 4.16 shows that there is a solution of the form

$$J_p(t) = t^p\Big(\frac{1}{2^p p!} + c_1 t + c_2 t^2 + \cdots\Big),$$

and this is defined to be the **Bessel function** of order p. The theorem gives another solution of the form t^{-p} times a power series except possibly when p is an integer or a half integer. To determine all these solutions, we substitute the series $t^s \sum c_n t^n$ and get

$$\begin{aligned}
&s(s-1)c_0 + (s+1)sc_1 t + (s+2)(s+1)c_2 t^2 + \cdots \\
+\quad &sc_0 + (s+1)c_1 t + (s+2)c_2 t^2 + \cdots \\
+\quad &c_0 t^2 + c_1 t^3 + \cdots \\
-\quad &p^2 c_0 - p^2 c_1 t - p^2 c_2 t^2 - p^2 c_3 t^3 - \cdots = 0.
\end{aligned}$$

The resulting equations are

$$\begin{aligned}
&[s(s-1) + s - p^2]c_0 = 0 && \text{from } t^0, \\
&[(s+1)s + (s+1) - p^2]c_1 = 0 && \text{from } t^1, \\
&[(s+n)(s+n-1) + (s+n) - p^2]c_n + c_{n-2} = 0 && \text{from } t^n \text{ for } n \geq 2.
\end{aligned}$$

The first of these equations repeats the indicial equation, giving $s = \pm p$. The second says that either $c_1 = 0$ or that $s + 1$ solves the indicial equation. In the latter case $s = -\frac{1}{2}$ and $p = \frac{1}{2}$. The third says that $[(s+n)^2 - p^2]c_n = -c_{n-2}$. For the case that $s = +p$, we obtain

$$c_n = \frac{-c_{n-2}}{(p+n)^2 - p^2},$$

and there is no problem from the denominator. The result is that the Bessel function of order $p \geq 0$ is given by

$$\begin{aligned}
J_p(t) &= \frac{t^p}{2^p p!}\Big(1 - \frac{t^2}{2(2p+2)} + \frac{t^4}{2 \cdot 4(2p+2)(2p+4)} - \cdots\Big) \\
&= \Big(\frac{t}{2}\Big)^p \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+p)!}\Big(\frac{t}{2}\Big)^{2k}.
\end{aligned}$$

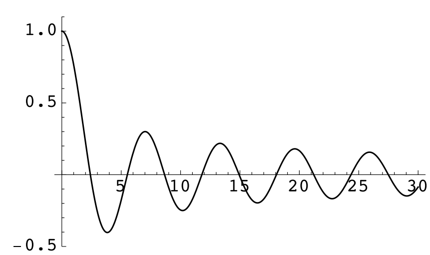


FIGURE 4.3. Graph of Bessel function $J_0(t)$.

For the case that $s = -p$, we obtain

$$c_n = \frac{-c_{n-2}}{(-p+n)^2 - p^2},$$

and the denominator gives a problem for $n = 2p$ and for no other value of n. If p is an integer, the problematic n is even and we must have $c_{n-2} = 0$, $c_{n-4} = 0$, $\dots$, $c_0 = 0$. The condition $c_0 = 0$ is a contradiction, and we conclude that there is no solution of the form t^{-p} times a nonzero power series; indeed, Problems 18–19 at the end of the chapter will identify a different kind of solution. If p is a half integer but not an integer, then the problematic n is odd, and we are led to conclude that $0 = c_{n-2} = \cdots = c_3 = c_1$, with c_0 and c_{2p} arbitrary. There is no contradiction, and we obtain a solution of the form t^{-p} times a nonzero power series.

9. Problems

1. For the differential equation $yy' = -t$:
   (a) Solve the equation.
   (b) Find all points (t_0, y_0) where the the existence theorem and the uniqueness theorem of Section 2 do not apply.
   (c) For each point (t_0, y_0) not in (b), give a solution $y(t)$ with $y(t_0) = y_0$.

2. Prove that the equation $y' = t + y^2$ has a solution satisfying the initial condition $y(0) = 0$ and defined for $|t| < 1/2$.

3. In classical notation, a particular vector field in the plane is given by $\sqrt{x}\,\frac{\partial}{\partial x} + \frac{1}{2}\frac{\partial}{\partial y}$. Find a parametric realization of an integral curve for this vector field passing through $(1, 1)$.

4. Evaluate $\dfrac{d}{dt}\displaystyle\int_0^{t^2} \frac{1}{s}(\sin st)\,ds$.

5. Find all solutions on $(-\infty, +\infty)$ to $y'' - 3y' + 2y = 4$.

6. (a) For each of these matrices A, find matrices B and J, with J in Jordan form, such that $A = BJB^{-1}$: $\quad A = \begin{pmatrix} 1 & 1 \\ 4 & -5 \end{pmatrix}$, $\quad A = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.
   (b) For each of the matrices A in (a), find a basis of solutions $y(t)$ to the system of differential equations $y' = Ay$.

7. The n^{th}-order equation $y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_0 y = 0$ with constant coefficients leads to a linear system $z' = Az$ with

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ & 0 & 1 & 0 & \cdots & 0 & 0 \\ & & 0 & 1 & \cdots & 0 & 0 \\ & & & \ddots & \ddots & \vdots & \vdots \\ & & & & 0 & 1 & 0 \\ & & & & & 0 & 1 \\ -a_0 & -a_1 & -a_2 & & \cdots & & -a_{n-1} \end{pmatrix}.$$

Prove that $\det(\lambda 1 - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0$ by expanding the determinant by cofactors.

8. (a) Let $\{f_n\}$ be a uniformly bounded sequence of Riemann integrable functions on $[0, 1]$. Define $F_n(t) = \int_0^t f_n(s)\,ds$. Prove that $\{F_n\}$ is an equicontinuous family of functions on $[0, 1]$.
   (b) Prove that the set of functions $y(t)$ on $[0, 1]$ with $y'' + y = f(t)$ and $y(0) = y'(0) = 0$ is equicontinuous as f varies over the set of continuous functions on $[0, 1]$ with $0 \le f(t) \le 1$ for all t.
   (c) Let $u(t)$ be continuous on $[a, b]$. Prove that the set of functions $y(t)$ on $[a, b]$ with $y'' + q(t)y = f(t)$ and $y(0) = y'(0) = 0$ is equicontinuous as $f(t)$ varies over the set of continuous functions on $[0, 1]$ with $0 \le f(t) \le 1$ for all t.

9. The differential equation $t^2y'' + (3t - 1)y' + y = 0$ has an irregular singular point at $t = 0$.
   (a) Verify that $\sum_{n=0}^{\infty}(n!)t^n$ is a formal power series solution of the equation even though the power series has radius of convergence 0.
   (b) Verify that $y(t) = t^{-1}e^{-1/t}$ is a solution for $t > 0$.

Problems 10–13 concern harmonic functions in the open unit disk, which were introduced in Problems 14–15 at the end of Chapter III. The first objective here is to use ordinary differential equations and Fourier series to show that all these functions may be expressed in a relatively simple form. The second objective is to use convolution, as defined in Problem 8 at the end of Chapter III, to relate this formula to the Poisson kernel, which was defined in Problems 27–29 at the end of Chapter I. Problems 10–12 here are an instance of the method of **separation of variables**, a beginning technique with partial differential equations; this topic is developed further in the companion volume, *Advanced Real Analysis*. In all problems in this set, let $u(x, y)$ be harmonic in the open unit disk.

10. Write $u(x, y)$ in polar coordinates as $u(r\cos\theta, r\sin\theta) = v(r, \theta)$. Using Fourier series, show for $0 \le r < 1$ and any $\delta > 0$ that $v(r, \theta)$ is the sum of an absolutely convergent Fourier series $\sum_{n=-\infty}^{\infty} c_n(r)e^{in\theta}$ with $|c_n(r)| \le M/n^2$ for $0 \le r \le 1 - \delta$ for some M depending on δ.

11. Let R_θ be the rotation matrix defined in Problem 15 at the end of Chapter III. That problem shows that $(u \circ R_\varphi)(x, y) = v(r, \theta + \varphi)$ is harmonic for each φ. Prove that $\frac{1}{2\pi} \int_{-\pi}^{\pi} (u \circ R_\varphi)(x, y) e^{-ik\varphi}\, d\varphi$ is harmonic and is given in polar coordinates by $c_k(r) e^{ik\theta}$.

12. By computing with the Laplacian in polar coordinates and showing that $c_k(r)$ is bounded as $r \downarrow 0$, prove that $c_k(r) = a_k r^{|k|}$ for some complex constant a_k. Conclude that every harmonic function in the open unit disk is of the form $v(r, \theta) = \sum_{n=-\infty}^{\infty} c_n r^{|n|} e^{in\theta}$, the sum being absolutely convergent for all r with $0 \le r < 1$.

13. Deduce from Problem 8 at the end of Chapter III that if $v(r, \theta)$ is as in the previous problem and if $0 < R < 1$, then $v(r, \theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_R(\varphi) P_{r/R}(\theta - \varphi)\, d\varphi$ for $0 \le r < R$, where P is the Poisson kernel and f_R is the C^∞ function $f_R(\theta) = \sum_{n=-\infty}^{\infty} c_n R^{|n|} e^{in\theta}$.

Problems 14–17 concern homogeneous linear differential equations. Except for the first of the problems, each works with a substitution in a second-order equation that simplifies the equation in some way.

14. If $a(t)$ is continuous on an interval and $A(t)$ is an indefinite integral, verify that all solutions of the single first-order linear homogeneous equation $y' = a(t)y$ are of the form $y(t) = ce^{A(t)}$.

15. (a) Suppose that $u(t)$ is a nowhere vanishing solution of

$$y'' + P(t)y' + Q(t)y = 0$$

on an interval, with P and Q assumed continuous. Look for a solution of the form $u(t)v(t)$, and derive the necessary and sufficient condition

$$v'(t) = cu(t)^{-2} e^{-\int P(t)\, dt}.$$

(b) For $y'' - ty' - y = 0$, one solution is $e^{t^2/2}$. Find a linearly independent solution.

16. Let $y'' + P(t)y' + Q(t)y = 0$ be given with P, P', and Q continuous on an interval. Write $y(t) = u(t)v(t)$, substitute, regard $u(t)$ as known, and obtain a second-order equation for v. Show how to choose $u(t)$ to make the coefficient of v' be 0, and thus reduce the given equation to an equation $v'' + R(t)v = 0$ with R continuous. Give a formula for R.

17. If $L(v) = (pv')' - qv + \lambda rv$, show that the substitution $u = v\sqrt{r}$ changes $L(v) = 0$ into $L_0(u) = 0$, where $L_0(u) = (p^*u')' - q^*u + \lambda u$ with $p^* = p/r$.

Problems 18–19 concern finding the form of the second solution to a second-order equation with a regular singular point. The first of the two problems amounts to a result in complex analysis but requires nothing beyond Chapter I of this book.

18. Suppose that $\sum_{n=0}^{\infty} c_n x^n$ is a power series with $c_0 = 1$.
 (a) Write down recursive formulas for the coefficients d_n of a power series $\sum_{n=0}^{\infty} d_n x^n$ with $d_0 = 1$ such that $\big(\sum_{n=0}^{\infty} c_n x^n\big)\big(\sum_{n=0}^{\infty} d_n x^n\big) = 1$.
 (b) Prove, by induction on n, that if $|c_n| \le Mr^n$ for all $n \ge 0$, then $|d_n| \le M(M+1)^{n-1} r^n$ for all $n \ge 1$.
 (c) Prove that if $f(0) \ne 0$ and if $f(x)$ is the sum of a convergent power series for $|x| < R$ for some $R > 0$, then $1/f(x)$ is the sum of a convergent power series for $|x| < \varepsilon$ for some $\varepsilon > 0$.

19. Suppose that $P(t)$ and $Q(t)$ are given near $t = 0$ by power series with positive radii of convergence. Take for granted that if $a(t)$ is given by a power series with a positive radius of convergence, then so is $e^{a(t)}$. Form the equation

$$t^2 y'' + tP(t)y' + Q(t)y = 0,$$

let s_1 and s_2 be the two roots of the indicial equation, and suppose that the differential equation has a solution given on some interval $(0, \varepsilon)$ by $f(t) = t^{s_1} \sum_{n=0}^{\infty} c_n t^n$ with $c_0 \ne 0$.
 (a) Using Problem 15a, prove that the differential equation has a linearly independent solution given on some interval $(0, \varepsilon')$ by

$$g(t) = cf(t)\log t + t^{s_2} \sum_{n=0}^{\infty} k_n t^n \qquad \text{with } k_0 \ne 0.$$

 (b) Prove that the coefficient c in $g(t)$ is $\ne 0$ if $s_1 = s_2$.
 (c) For Bessel's equation $t^2 y'' + ty' + (t^2 - p^2)y = 0$ with $p \ge 0$ an integer and with $s_1 = p$ and $s_2 = -p$, show that the coefficient c in $g(t)$ is $\ne 0$. Thus there is a solution of the form $J_p(t)\log t + t^{-p}$(power series) on some interval $(0, \varepsilon')$.

Problems 20–25 prove the **Cauchy–Peano Existence Theorem**, that a local solution in Theorem 4.1 to $y' = F(t, y)$ and $y(t_0) = y_0$ exists if F is continuous even if F does not satisfy a Lipschitz condition. The idea is to construct a sequence of polygonal approximations to solutions, check that they form an equicontinuous family, apply Ascoli's Theorem (Theorem 2.56) to extract a uniformly convergent subsequence, and then see that the limit of the subsequence is a solution. A member of the sequence of polygonal approximations depends on a number $\epsilon > 0$. With notation as in the statement of Theorem 4.1, the construction for $[t_0, t_0 + a']$ is as follows: Choose the δ of uniform continuity for F and ϵ on the set R. Fix a partition $t_0 < t_1 < \cdots < t_n = t_0 + a'$ of $[t_0, t_0 + a']$ with $\max_k\{t_k - t_{k-1}\} \le \min(\delta, \delta/M)$. Define $y(t)$, as a function of ϵ, for $t_{k-1} < t \le t_k$ inductively on k by $y(t_0) = y_0$ and

$$y(t) = y(t_{k-1}) + F(t_{k-1}, y(t_{k-1}))(t - t_{k-1}).$$

20. Check that the formula for $y(t)$ when $t_{k-1} < t \le t_k$ remains valid when $t = t_{k-1}$, and conclude that $y(t)$ is continuous. Then prove by induction on k that $|y(t) - y(t_0)| \le M(t - t_0) \le b$ for $t_{k-1} \le t \le t_k$, and deduce that $(t, y(t))$ is in R' for $t_0 \le t \le t_0 + a'$.

21. Prove that $|y(t) - y(t')| \le M|t - t'|$ if t and t' are both in $[t_0, t_0 + a']$.

22. The function $y'(t)$ is defined on $[t_0, t_0 + a']$ except at the points of the partition and is given by $y'(t) = F(t_{k-1}, y(t_{k-1}))$ if $t_{k-1} < t < t_k$. Prove that $y(t) = y_0 + \int_{t_0}^t y'(s)\,ds$ for $t_0 \le t \le t_0 + a'$ and that $|y'(s) - F(s, y(s))| \le \epsilon$ if $t_{k-1} < s < t_k$.

23. Writing $y(t) = y_0 + \int_{t_0}^t [F(s, y(s)) + [y'(s) - F(s, y(s))]]\,ds$ and using the result of the previous problem, prove for all t in $[t_0, t_0 + a']$ that

$$\Big|y(t) - \big(y_0 + \textstyle\int_{t_0}^t F(s, y(s))\,ds\big)\Big| \le \epsilon a'.$$

24. Let ϵ_n be a monotone decreasing sequence with limit 0, and let $y_n(t)$ be a function for t in $[t_0, t_0 + a']$ constructed as above for the number ϵ_n. Deduce from Problem 21 that $\{y_n(t)\}$ is uniformly bounded and uniformly equicontinuous for t in $[t_0, t_0 + a']$.

25. Apply Ascoli's Theorem to $\{y_n\}$, and let $y(t)$ be the uniform limit of a uniformly convergent subsequence of $\{y_n\}$. Prove that $y(t)$ is continuous, and use Problem 23 to prove that $y(t) = y_0 + \int_{t_0}^t F(s, y(s))\,ds$. What modifications are needed to the argument to handle $[t_0 - a', t_0]$?

Problems 26–28 use elementary complex analysis as in Appendix B to shed further light on results and problems in this chapter.

26. Let $u(x, y)$ be harmonic in the open unit disk. Bypassing Problems 10 and 11, write $u(x, y)$ as the real part of an analytic function $f(z)$, expand $f(z)$ in Taylor series about $z = 0$, take the real part of the expansion, and deduce the conclusion of Problem 12 that $u(x, y)$, when written in polar coordinates as $v(r, \theta)$, is of the form $\sum_{n=-\infty}^{\infty} c_n r^{|n|} e^{in\theta}$, the sum being absolutely convergent for all r with $0 \le r < 1$.

27. In the context of Problem 18, use the theory of analytic functions to deduce that if $f(z)$ has $f(0) \ne 0$ and is the sum of a convergent power series for $|x| < R$ for some $R > 0$, then $1/f(z)$ is the sum of a convergent power series for $|z| < \varepsilon$ for some $\varepsilon > 0$.

28. This problem derives an integral formula for the Bessel function $J_n(z)$ introduced near the end of Section 8. The function $e^{iz\sin\theta}$ is a continuous complex-valued function for (z, θ) in $\mathbb{C} \times \mathbb{R}$ that is analytic as a function of z for each fixed θ and is periodic in θ for each fixed z. The problem works with the Fourier coefficients of this function.

(a) Define $c_n(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{iz\sin\theta}e^{-in\theta}\,d\theta$. Why is this an entire function of z?

(b) Using the power series expansion of the exponential function, show that

$$e^{iz\sin\theta} = \sum_{p=0}^{\infty} \frac{1}{p!}\Big(\frac{z}{2}\Big)^p (e^{i\theta} - e^{-i\theta})^p,$$

and justify the interchange of limits that gives

$$c_n(z) = \sum_{p=0}^{\infty} \frac{1}{p!}\Big(\frac{z}{2}\Big)^p \Big(\frac{1}{2\pi}\int_{-\pi}^{\pi} (e^{i\theta} - e^{-i\theta})^p e^{-in\theta}\,d\theta\Big).$$

(c) Write $I_{n,p}$ for the expression $\frac{1}{2\pi}\int_{-\pi}^{\pi}(e^{i\theta} - e^{-i\theta})^p e^{-in\theta}\,d\theta$ in (b). Show that

$$I_{n,p} = \begin{cases} (-1)^k\binom{p}{k} & \text{if } p = n + 2k \text{ with } k \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

and simplify for $n \geq 0$ to obtain

$$c_n(z) = \Big(\frac{z}{2}\Big)^n \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\,(n+k)!}\Big(\frac{z}{2}\Big)^{2k}.$$

(d) Conclude that $c_n(z) = J_n(z)$ if $n \geq 0$ and that $c_{-n}(z) = (-1)^n J_n(z)$ if $n \geq 0$. In particular,

$$J_n(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{iz\sin\theta}e^{-in\theta}\,d\theta$$

for $n \geq 0$.

(e) Obtain the formula $e^{i\sin\theta} = J_0(z) + \sum_{n=1}^{\infty} J_n(z)(e^{in\theta} + (-1)^n e^{-in\theta})$.