PROJECT RAND

A SURVEY OF THE MATHEMATICAL THEORY OF TIME-LAG, RETARDED CONTROL, AND HEREDITARY PROCESSES

RICHARD BELLMAN
with the assistance of
JOHN M. DANSKIN, JR.

March 1, 1954

R-256

The views, conclusions and recommendations expressed herein do not necessarily reflect the official views or policies of the United States Air Force.

The RAND Corporation
1700 Main St, Santa Monica, California

Copyright 1954, The RAND Corporation
PREFACE

In recent years, the introduction of servomechanisms and other control devices has brought into prominence the physical and mathematical problems connected with time lags, retarded control, and hereditary processes in general. These problems, however, constitute only a small part of a vast class of physical phenomena which, for their study, require a knowledge not only of their present history, but also of all or part of their past history.

While classical examples are magnetic hysteresis and elastic fatigue, many further examples of hereditary influence arise in the applications of probability theory to economic, industrial, and military problems; in the general theory of fission processes, embracing the cascade theory of cosmic-ray showers and the biological process of mutation; in the investigation of learning processes and mental breakdowns; and in many other fields as well—too numerous to list in entirety.

Mathematically, this means a shift from the classical description of a physical system by means of a system of differential equations of the type

\[
\frac{dy_i}{dt} = F_i(y_1, y_2, \cdots, y_N), \quad i = 1, 2, \cdots, N,
\]

with initial conditions of the form

\[
y_i(0) = \epsilon_i,
\]

(1)

(2)

to functional equations of the form

\[
\frac{dy_i(t)}{dt} = F_i[y_1(t), y_2(t), \cdots, y_N(t), y_1(t-\tau), y_2(t-\tau), \cdots, y_N(t-\tau)],
\]

where \( \tau \) is the time lag, with initial conditions of the form

\[
y_i(t) = \phi_i(t), \quad 0 \leq t \leq \tau.
\]

These differential-difference equations are in turn only approximations to more realistic and complicated equations of the form

\[
\frac{dy_i(t)}{dt} = G_i[y_1(t), y_2(t), \cdots, y_N(t)]
\]

\[
\quad + \int_{-\infty}^{t} F_i[y_1(s), y_2(s), \cdots, y_N(s)] dK_i(s, t),
\]

\[
\quad i = 1, 2, \cdots, N,
\]

(3)

(4)

(5)
of which the simplest and most important representative is

\[ u(t) = f(t) + \int_0^t u(t - s) \, dG(s), \]

the "renewal" equation.

We may think of Eq. (1) as representative of a physical system in which there is instantaneous transmission of influence, so that only the present is required to determine the future. In Eq. (3), we are considering systems in which some effects are instantaneous, whereas others require a nonzero germination time. Equation (5), on the other hand, considers the system to be one in which the entire past history of the system plays a role in determining the future.

This is always the realistic situation, the others being merely approximations of greater or lesser validity. In many applications in which the quantity of interest varies slowly compared with the velocity of transmission of the underlying influences, it is permissible to simplify the mathematical formulation and to employ Eq. (1) in place of Eqs. (3) or (5).

As typical of a recent development in which the enormous increase in velocity—supersonic versus subsonic—requires Eq. (3) rather than Eq. (1) for any adequate analysis, let us consider the airplane. Any realistic analysis of automatic pilotless control, of the prediction of the course of an airplane from its position, or of the stability of course must take into account the inevitable time lag involved in control and computation, since this time is no longer negligible compared with the time required for the airplane to travel a nonnegligible distance.

In many cases, furthermore, it is impossible to understand the instability phenomena which occur without introducing the concept of time lags.

Much of the important research work in the theory and application of this general type of functional equation is quite recent, appearing in a variety of research journals and in several languages. In view of the growing value of this research, and the lack of any one systematic presentation of the results, we feel that a survey of the theory and application will serve a useful purpose at this time, when the applications are increasing in number and importance.

In order to make the survey as self-contained as possible, we have included a preliminary chapter (Chapter 1) on the Laplace transform, an indispensable tool in the study of linear functional equations of the differential, difference, or differential-difference type. In this chapter we present the most important results concerning the inversion and convolution of the Laplace transform, together with some brief remarks concerning numerical inversion.

The second chapter contains an exposition of the treatment of linear differential-difference equations with constant coefficients, using Laplace transform techniques.

In the third chapter we consider the renewal equation

\[ u(t) = f(t) + \int_0^t u(t - s) \, dG(s), \]

(7)
treating, for the sake of simplicity, only the case where the cumulative function has a
density, \( dG(t) = g(t) \, dt \). After proving various existence and uniqueness theorems, and deriving some elementary properties of the solution, the explicit solution obtained by means of the Laplace transform is exhibited. Following this, we present a number of typical situations which give rise to the renewal equation.

This sets the stage in the fourth chapter for the investigation of the asymptotic properties of the solution, usually the most important feature as far as applications are concerned. Various methods to determine this asymptotic behavior are presented, ranging from elementary approaches through the real Tauberian theorems of Hardy and Littlewood to the complex Tauberian result of Ikehara. We close the chapter with a discussion of the usual contour-shifting methods of complex variable theory.

The fifth chapter contains an introduction to the relatively uncharted territory of systems of renewal equations. The theory is not merely an extension of the one-dimensional theory, but contains genuinely new features of interest and difficulty. Here we come briefly into contact with the theory of positive operators.

The sixth chapter deals with a topic of fundamental importance in connection with the stability of a dynamic system—the location of the zeros of exponential polynomials of the form \( P(\gamma)e^{\gamma t} + Q(\gamma) = 0 \). A feature of this chapter is an exposition and application of the results of Pontryagin, which, although published in 1942, are still unfamiliar to many American engineers and mathematical physicists.

Following this, we turn to a discussion of the stability theory for differential-difference equations (Chapter 7). After a brief discussion of the corresponding classical theory of Poincaré and Liapounoff for differential equations, we present the analogous results for differential-difference equations. References are given to the further-detailed and extensive results of E. M. Wright.

The eighth chapter treats of various physical and mathematical aspects of the general control problem which leads to differential and differential-difference equations. Following the pioneering lead of Minorsky, we emphasize the fundamental importance of time lags and retardation in relation to stability theory. A number of applications are given.

In the final chapter we discuss some problems in mathematical economics arising from the macrodynamic model of Kalecki. These lead to differential-difference equations and to stability problems resoluble by the methods of Chapter 6.

In the Appendix a translation is given of a supplement to an interesting and valuable bibliographical article of Myshkis on differential-difference and allied functional equations.

The bibliography at the end of the work, given chronologically rather than alphabetically, is reasonably complete. It contains the majority of important papers concerned with the theory of differential-difference equations, and contiguous topics. Together with the references contained in these papers, particularly those of Bateman and Myshkis, this list furnishes a useful guide to the applications.

At some future time, we intend to transform this survey into a book which will include a number of topics that, for one reason and another, we have shunted aside here. Among those topics would be the theory of renewal equations involving Stieltjes integrals; a chapter on Tauberian techniques containing the results of Wiener and Pitt, as utilized in
the study of linear functional equations by Pitt and Wright; the work of Bochner on
linear difference equations with incommensurable spans; and finally the theory of
differential-difference equations with variable lags, as recently developed by Myshkis.

In addition to the increments of purely mathematical character described above, it
would be highly worth while to include an extensive treatment of applications, with
particular regard to the underlying physical hypotheses which give rise to the equations
of various types.

A number of associates have very kindly assisted us in the preparation of this survey.
We would like particularly to thank D. V. Widder, who made a number of valuable
suggestions concerning Chapter 1; T. E. Harris, who read portions of Chapter 3;
O. Gross, who read the manuscript through with his usual thoroughness and care, and, in
addition to catching a number of slips and misprints, suggested a number of helpful
stylistic changes; and I. Glicksberg, who, in addition to suggesting a number of important
simplifications, aided in the preparation of the Bibliography.

A number of interesting discussions with E. W. Paxson in connection with applications
of many varied types were instrumental in arousing our interest in a general study of the
theory and application of differential-difference equations.
SUMMARY

This report contains a summary of the mathematical techniques required to treat of physical phenomena involving time lags, retarded control, or hereditary effects. The functional equations which arise are no longer the differential equations of classical mathematical physics, but rather differential-difference equations, integrodifferential equations, and equations of even more complicated form.

The most important applications of the mathematical theory arise in connection with control problems, and the resulting stability investigations. Questions involving these advanced methods arise in the theory of guided missiles and pilotless aircraft. It is here that the tremendous velocities, which are now feasible, make the time lags created by the control mechanism of great significance. In many cases it is impossible even to understand the origin of various instability phenomena without taking into account time lags and retarded control. For this reason these ideas are becoming of increasing importance in the field of servomechanisms and automatic control.

Equations of this form also play an important role in mathematical economics where the analysis of interindustrial processes requires an awareness of the fact that some changes cannot occur instantaneously.

Other physical phenomena requiring these concepts occur in the theory of magnetism, in the theory of elasticity and plasticity, and throughout the theory of fission processes. In the field of biology these ideas are required to explain mutation and, in general, the growth of unicellular organisms. They are of particular importance in furnishing an understanding of the effects of radioactive exposure, and thus, indirectly, in the study of cancer. In the field of psychology, they are necessary for any treatment of learning theory and other long-term effects, such as mental breakdowns.

Problems of this type will occur wherever the future depends not only on the immediate present, but also on the past history of the system under consideration.
CONTENTS

Preface ................................................................. iii
Summary ............................................................... vii

Chapter 1
THE LAPLACE TRANSFORM

Section
1.1 Introduction ....................................................... 1
1.2 Existence and Convergence ....................................... 1
1.3 The Inversion Problem ........................................... 2
1.4 Contour Integration .............................................. 4
1.5 The Inverse Inversion Problem ................................. 6
1.7 The Convolution Theorem ....................................... 8
1.7 A Real Inversion Formula ..................................... 10

Chapter 2
THE LINEAR DIFFERENTIAL-DIFFERENCE EQUATION
WITH CONSTANT COEFFICIENTS

2.1 Introduction ...................................................... 15
2.2 Formal Aspects .................................................. 15
2.3 Rigorous Derivation ............................................. 17
2.4 A Real Representation of the Solution ....................... 18
2.5 Asymptotic Behavior of the Solution ....................... 21
2.6 Fourier-type Expansions ..................................... 23
2.7 Differential-difference Equations and Causality .......... 25

Chapter 3
THE RENEWAL EQUATION

3.1 Introduction ...................................................... 27
3.2 Existence and Uniqueness Theorems ............................ 27
3.3 Monotonicity and Bounded Variation ........................... 31
3.4 The Laplace Transform Solution ............................... 33
3.5 A Convolution Theorem ....................................... 34
3.6 Applications—I .................................................. 36
3.7 Applications—II ................................................ 37
3.8 Applications—III ............................................... 38
3.9 Applications—IV ............................................... 38
3.10 Applications—V ............................................... 39
CONTENTS

CHAPTER 4
ASYMPTOTIC BEHAVIOR OF THE SOLUTION OF THE RENEWAL EQUATION

SECTION
4.1 Introduction ................................................................. 43
4.2 The Result of Paley and Wiener ........................................ 44
4.3 The Tauberian Theorem of Hardy and Littlewood ................. 44
4.4 The Elementary Approach ............................................... 45
4.5 The Tauberian Theorem of Ikehara .................................... 48
4.6 Use of the Contour Integral Representation ......................... 49

CHAPTER 5
SYSTEMS OF RENEWAL EQUATIONS

5.1 Introduction ......................................................................... 51
5.2 The Perron Theorem for Positive Matrices ......................... 51
5.3 Expectation Matrices ........................................................ 54
5.4 Asymptotic Behavior ........................................................ 57
5.5 Applications ....................................................................... 58

CHAPTER 6
ON THE ZEROS OF EXPONENTIAL POLYNOMIALS

6.1 Introduction ......................................................................... 59
6.2 Summary of Pontryagin’s Results ........................................ 59

CHAPTER 7
THE STABILITY OF SOLUTIONS OF DIFFERENTIAL-DIFFERENCE EQUATIONS

7.1 Introduction ......................................................................... 73
7.2 The Analogue of the Poincaré-Liapounoff Theorem ................ 74
7.3 Linear Differential-difference Equations with Almost Constant Coefficients ......................................................... 75

CHAPTER 8
CONTROL PROBLEMS

8.1 Introduction ......................................................................... 77
8.2 Mathematical Preliminaries ................................................ 77
8.3 Derivation of the Hurwitz Criteria ....................................... 79
8.4 Nonlinear Instantaneous Control ......................................... 79
8.5 Discussion .......................................................................... 80
8.6 Some Examples .................................................................... 81
CONTENTS

CHAPTER 9
SOME ECONOMIC MODELS

SECTION
9.1 Introduction ................................................. 85
9.3 A Brief Survey of the Literature Related to Kalecki's Model .......... 88

APPENDIX: SUPPLEMENTARY BIBLIOGRAPHICAL MATERIALS TO THE ARTICLE,
"GENERAL THEORY OF DIFFERENTIAL EQUATIONS WITH A
RETARDED ARGUMENT" ........................................... 91

BIBLIOGRAPHY ..................................................... 99
CHAPTER 1

THE LAPLACE TRANSFORM

1.1. Introduction

In this chapter we shall discuss the fundamental properties of the Laplace transform, defined by means of the integral operator,

\[ F = L(f) = \int_{0}^{\infty} e^{-st} f(t) \, dt. \]  \hspace{1cm} (1.1)

Since we are interested only in applying this transform to the study of linear functional equations, we shall discuss only those aspects of the theory which are useful in this connection.

Three problems to which we require answers in connection with our subsequent applications are:

1. Given \( F = L(f) \), how does one determine \( L^{-1}(F) \)?
2. When is \( F(s) \) an \( L(f) \)?
3. Given \( L(f) \) and \( L(g) \), how does one determine \( L^{-1}[L(f)L(g)] \)? \hspace{1cm} (1.2)

Although the results we present are classical and may be found in the standard treatises of Doetsch [1] and Widder [4],\(^{*}\) we feel that a fairly self-contained account of what is required will materially aid our subsequent presentation and will be of some value to the reader who is primarily interested in the applications.

1.2. Existence and Convergence

The infinite integral appearing in Eq. (1.1) is defined as an improper Lebesgue integral

\[ F = \lim_{R \to \infty} \int_{0}^{R} e^{-st} f(t) \, dt, \]  \hspace{1cm} (1.3)

where \( f(t) \) is assumed to be integrable over any finite interval \([0, R]\). For our purposes there is no real need of the Lebesgue integral, and we could just as well dispense with it and use the old-fashioned Riemann integral, since most of the functions we shall encounter shall not only be continuous, but even analytic. If, perchance, any of our functions possess discontinuities, they will usually be jump singularities.

It is not difficult to show that the convergence of the integral for some \( s_0 = \sigma_0 + ir_0 \)

\(^{*}\)Bracketed numerals refer to the references at the end of each chapter.
entails its convergence for \( s = \sigma + it \) for any \( \sigma > \sigma_0 \). In most of our applications the integral will converge absolutely for \( \sigma \) sufficiently large.

It is also not difficult to see that if \( F(s) \) exists for \( s = \sigma_0 + it \), then it is analytic in the half-plane \( \text{Re}(s) > \sigma_0 \).

For a thorough discussion of the regions of convergence and absolute convergence of the Laplace transform, we refer to [1] and [4].

1.3. The Inversion Problem

Let us now discuss the problem of determining \( f \), given \( F \). We shall assume that

\[
F(s) = \int_0^\infty f(t_1) e^{st_1} dt_1,
\]

and that the integral converges absolutely for \( \text{Re}(s) > \sigma_0 \). Then, for \( \sigma > \sigma_0 \), we have

\[
F(\sigma + it) = \int_0^\infty f(t_1) e^{-(\sigma + it)t_1} dt_1.
\]

Multiplying both sides by \( e^{(\sigma + it)t} \) and integrating with respect to \( t \) between \(-T\) and \( T\), we obtain

\[
\int_{-T}^T e^{(\sigma + it)t} F(\sigma + it) dt = e^{\sigma u} \int_0^\infty f(t_1) e^{\sigma t_1} \left( \int_{-T}^T e^{iut_1 t} dt \right) dt_1.
\]

The change of orders of integration which yields Eq. (1.6) is justified by the absolute convergence of the double integral. Simplification of the right side yields

\[
\int_{-T}^T e^{(\sigma + it)t} F(\sigma + it) dt = 2e^{\sigma u} \int_0^\infty f(t_1) e^{\sigma t_1} \frac{\sin T(u - t_1)}{u - t_1} dt_1.
\]

The function \( \frac{\sin T(u - t_1)}{u - t_1} \) is called the Dirichlet kernel, since it was first encountered by Dirichlet in his investigation of the convergence of the Fourier series. Its graph has the shape of (1.8). As \( T \) gets larger, the peak at \( u \) becomes more and more
pronounced. It is consequently tempting to guess that as $T \to \infty$ the value of the integral becomes more and more dependent on the value of the function $f(t_1)e^{-at_1}$ at $t_1 = u$. If this guess is correct, we shall have a means of determining $f(t)$.

Let us show first that we can reduce the interval of integration from $[0, \infty]$ to an arbitrarily small interval about $u$ as $T \to \infty$. To do this we require the classical

Riemann-Lebesgue Lemma. If $\int_0^\infty |g(t)| \, dt < \infty$, then

$$\lim_{T \to \infty} \int_{-\infty}^\infty g(t) \begin{bmatrix} \sin Tt \\ \cos Tt \end{bmatrix} \, dt = 0.$$  

(1.9)

For a proof we refer to [2] or [3].

Employing this lemma, together with our hypothesis that $\int_0^\infty |f(t)| \, e^{-at} \, dt < \infty$, we see that, for $u > 0$,

$$\lim_{T \to \infty} \int_0^u f(t_1)e^{-at_1} \frac{\sin T(u - t_1)}{u - t_1} \, dt_1$$

$$= \lim_{T \to \infty} \int_{u-d}^{u+d} f(t_1)e^{-at_1} \frac{\sin T(u - t_1)}{u - t_1} \, dt_1,$$  

(1.10)

where $d$ is an arbitrarily small positive quantity.

Let us now assume that $f(t_1)$ is well enough behaved in a neighborhood of $u$ to possess a Taylor expansion,

$$f(t_1)e^{-at_1} = f(u)e^{-au} + g(u)(u-t_1) + O((u-t_1)^2),$$

(1.11)

valid for $|u-t_1| \leq d$.

Then,

$$\int_{u-d}^{u+d} f(t_1)e^{-at_1} \frac{\sin T(u-t_1)}{u-t_1} \, dt_1$$

$$= f(u)e^{-au} \int_{u-d}^{u+d} \sin T(u-t_1) \, dt_1$$

$$+ g(u) \int_{u-d}^{u+d} \sin T(u-t_1) \, dt_1 + O(d),$$

(1.12)

where the $O(d)$ estimate holds uniformly in $T$, since $|\sin T(u-t_1)| \leq 1$.

Since

$$\lim_{T \to \infty} \int_{u-d}^{u+d} \frac{\sin T(u-t_1)}{u-t_1} \, dt_1 = \int_0^\infty \frac{\sin t}{t} \, dt = \pi,$$

(1.13)

we obtain, for any fixed $d > 0$, the result

$$\lim_{T \to \infty} \int_0^\infty f(t_1)e^{-at_1} \frac{\sin T(u-t_1)}{u-t_1} \, dt_1 = \pi f(u)e^{-au} + O(d).$$

(1.14)
Since \( d \) is arbitrary, we obtain, under the assumption of Eq. (1.11), the result

\[
f(u) = \frac{1}{2\pi} \lim_{\tau \to \infty} \int_{\tau}^{\infty} e^{(a+it)u} F(a + it) \, dt,
\]

(1.15)

upon referring to Eq. (1.7).

We see then that we have an inversion formula, provided that \( f(t) \) is sufficiently smooth. It was shown by Dirichlet, essentially, that a condition as strong as Eq. (1.11) is not required, but that we merely need require that \( f(t) \) be of bounded variation in a neighborhood of \( t \). Under this assumption it can be shown (see [1], [3], and [4]) that

\[
\lim_{\tau \to \infty} \int_{u-\tau}^{u+\tau} f(t_1) \frac{\sin T(u - t_1)}{u - t_1} \, dt_1 = \frac{f(u - 0) + f(u + 0)}{2}.
\]

(1.16)

For further reference, let us state the following result concerning the inversion of the Laplace transform:

**Theorem 1.1.** Consider the integral equation

\[
F(t) = \int_{0}^{\infty} e^{-st} f(t_1) \, dt_1,
\]

(1.17)

where

(a) the integral converges absolutely for \( \text{Re}(s) > a \),

(b) \( f(t_1) \) is of bounded variation in the neighborhood of \( t \).

Then, for \( b > a \),

\[
\lim_{\tau \to \infty} \frac{1}{2\pi} \int_{-\tau}^{\tau} F(b + it_1) e^{(b+it_1)t} \, dt_1 = \frac{f(t - 0) + f(t + 0)}{2}, \quad t > 0,
\]

\[
= \frac{f(+0)}{2}, \quad t = 0.
\]

(1.19)

If \( f(t_1) \) is continuous at \( t_1 = t \), we obtain \( f(t) \). However, it may be shown by examples that continuity alone, without bounded variation or some other condition, is not sufficient.

### 1.4. Contour Integration

The most important application of Theorem 1.1 results from the fact that, under favorable circumstances, we may convert the left side of Eq. (1.19) into a contour integral and thus bring to bear the powerful machinery of the theory of functions of a complex variable.

We may write

\[
\frac{1}{2\pi} \int_{-\tau}^{\tau} F(b + it_1) e^{(b+it_1)t} \, dt_1 = \frac{1}{2\pi i} \int_{b-\tau}^{b+\tau} F(s) e^{st} \, ds,
\]

(1.20)

where the expression on the right is a contour integral taken along the line joining
THE LAPLACE TRANSFORM

$b - iT$ and $b + iT$. Let us introduce the notation

$$\int (b) F(s) e^{st} ds = \lim_{T \to \infty} \frac{1}{2\pi i} \int_{b-iT}^{b+iT} F(s) e^{st} ds.$$  \hspace{1cm} (1.21)

We then have the notationally simple formula

$$f(t) = \int (b) F(s) e^{st} ds,$$ \hspace{1cm} (1.22)

whenever $f$ is continuous as well as of bounded variation. That the solution is independent of $b$, for $b > a$, follows from the analyticity of the integrand for $Re(s) > a$, which is a consequence of our assumption concerning the absolute convergence of $\int_{-\infty}^{\infty} f(t) e^{-at} dt$. If $F(s)$ has an analytic continuation to the left, we may evaluate $f(t)$ explicitly or obtain its asymptotic behavior as $t \to \infty$. This technique is extremely powerful in various applications, and we shall make extensive use of it in later sections.

As a simplest possible example illustrating the technique, consider the case in which

$$F(s) = \int_{0}^{\infty} e^{-st} dt = \frac{1}{s}, \hspace{1cm} Re(s) > 0.$$ \hspace{1cm} (1.23)

If we were given the function $1/s$ and were required to find its Laplace inverse, we would have the problem of evaluating

$$\int (b) \frac{e^{st} ds}{s}, \hspace{1cm} b > 0, \hspace{1cm} t > 0.$$ \hspace{1cm} (1.24)

This may be done by using the classical techniques, i.e., by considering, for example, the contour of (1.25). The pole at $s = 0$ contributes a residue of 1, and it is easy to see

$$\begin{array}{c}
-\infty + iT \\
-\infty - iT \\
\infty - iT \\
\infty + iT
\end{array}$$

that the contributions from the sides of the contour marked by arrows tend to zero as $T \to \infty$.

It is clear that the integral is zero for $t < 0$, since we may shift the contour to $+\infty$ without encountering any poles.

In general, to evaluate an integral of the type given in Eq. (1.22), where $F(s)$ is
meromorphic over the whole plane, we push the line of integration to $-\infty$, obtaining a residue from every pole we overtake. This will yield an explicit evaluation of the form

$$f(t) = \sum_i a_i(t)e^{s_i t},$$

(1.26)

where the coefficients, $a_i(t)$, are polynomials of degree $k - 1$ if $s_i$ is a pole of multiplicity $k$.

Even if it is not possible, because of analytic difficulties, to push the contour to $-\infty$, we may still obtain valuable information by shifting the contour past the first few poles which occur.

Thus, for example, let us consider the integral

$$f(t) = \int_{(b)} \frac{e^{st}}{e^s - s - 1} ds, \quad b > 0, \quad t > 0.$$  

(1.27)

Using the easily demonstrated fact that $e^s - s - 1 = 0$ has no zeros with positive real part, and only one, $s = 0$, with real part $\geq 0$, we obtain

$$f(t) = -\frac{1}{2} + \int_{(b')} \frac{e^{st}}{e^s - s - 1} ds,$$

(1.28)

where $b'$ is some negative quantity. It is now not difficult to show that the integral is $O(e^{b't})$ as $t \to +\infty$. From this it follows that $-\frac{1}{2}$ is an excellent approximation to $f(t)$ for large $t$. Examples of this type will be discussed in more detail subsequently.

If $F(s)$ has more complicated types of singularities, such as branch points, the evaluation of the contribution from the singularity is more difficult. We shall not encounter any examples of this in our applications.

Let us insert one final word of warning. The extension of the contour to the left cannot be done without considering the terms obtained from the crossbars parallel to the real axis. This frequently requires a very accurate knowledge of the location of the singularities of $F(s)$, which in some cases is difficult to ascertain.

1.5. The Inverse Inversion Problem

In the previous section we considered the problem of determining $f$, given $F$. In later sections we shall be confronted with the problem of determining whether or not a given analytic function $F(s)$ is a Laplace transform of some function $f(t)$.

Since we are not interested in the most general case which can occur, but only in the classes of functions which arise in the course of our investigations, we shall restrict ourselves to proving

**Theorem 1.2.** If for some real number $a$, a function $F(s)$ satisfies

(a) $F(s)$ is analytic for $\text{Re}(s) \geq a$,

(b) $F(s) = c/e^s + o(1/|s|^p)$ as $|s| \to \infty$ with $\text{Re}(s) \geq a$,       

(1.29)
Then the integral

\[ f(t) = \int_{(b)}^b F(s) e^{st} \, ds, \quad b > a, \]  

exists for \( b > a \), and we have

\[ F(s) = \int_0^\infty e^{-st} f(t) \, dt \]  

for \( \text{Re}(s) > a \).

**Proof.** Writing \( F = c_1/e^s + g(s) \), we have

\[ f = c_1 \int_{(b)}^b e^{st} \, ds + \int_{(b)}^\infty g(t) e^{st} \, ds \]

\[ = c_1 + \int_{(b)}^\infty g(t) e^{st} \, ds \]  

for \( t > 0 \). Let us, without loss of generality, assume that \( b > 0 \). If \( t < 0 \), we already know that the first integral is zero. Similarly, the absolute convergence of \( \int_{(b)}^\infty g(t) e^{st} \, ds \) makes it clear that

\[ \int_{(b)}^\infty g(t) e^{st} \, ds \to 0 \]

as \( b \to +\infty \) if \( t < 0 \). Hence \( f = 0 \) for \( t < 0 \). *

The question now arises as to whether

\[ g(s) = \int_0^\infty [f(t) - c_1] e^{-st} \, dt \]  

for \( \text{Re}(s) \geq b \). The argument is very much as before. We have

\[ \int_0^s \left[ \int_0^\infty g(b + iu) e^{(b+iu)u} \, du \right] e^{-st} \, dt \]

\[ = \int_0^\infty g(b + iu) \left[ \int_0^\infty e^{(b+iu)u} \, du \right] dt, \]  

where the inversion of the orders of integration is justified by the absolute convergence.

---

*Here, and frequently in what follows, we omit any mention of the value at the boundary point \( t = 0 \), since this value is of no interest to us in the following pages. As might be expected, it is usually the average of the value at \( +0 \) and \( -0 \); see [1] and [4].
of all the integrals involved. Simplifying, we are led to investigate the existence and value of

$$\lim_{R \to \infty} \int_{-a}^{a} g(b + iu) \frac{e^{(b+iu-t)R} - 1}{(b + iu - t)} \, du.$$  \hfill (1.35)

Since \( F(t) \) is assumed to be analytic for \( Re(t) \geq a \), \( g(t) \) is certainly continuous and of bounded variation. Consequently, the representation in Eq. (1.31) is valid, upon referring to the argumentation preceding Theorem 1.1.

1.6. The Convolution Theorem

In connection with the practical and theoretical application of the Laplace transform, an invaluable aid is a table of known Laplace transforms and inverses. This table becomes more valuable if we add to it a rule for determining the Laplace inverse of \( FG \), given the inverses of \( F \) and \( G \).

If we proceed formally, we have

$$L^{-1}(FG) = \int_{(b)} F(s)G(s)e^{st} \, ds = \int_{(b)} F(s)e^{st} \left[ \int_{0}^{\infty} g(t_1)e^{-st_1} \, dt_1 \right] \, ds$$

$$= \int_{0}^{\infty} g(t_1) \left[ \int_{(b)} F(s)e^{s(t_1-t)} \, ds \right] \, dt_1.$$  \hfill (1.36)

Since

$$\int_{(b)} F(s)e^{s(t_1-t)} \, ds = f(t-t_1) \quad \text{for } t > t_1,$$

$$= 0 \quad \text{for } t < t_1,$$  \hfill (1.37)

we obtain from Eq. (1.36)

$$L^{-1}(FG) = \int_{0}^{t} g(t_1)f(t-t_1) \, dt_1 = \int_{0}^{t} f(t_1)g(t-t_1) \, dt_1.$$  \hfill (1.38)

In order to obtain this result rigorously without imposing too strict conditions upon \( F \) and \( G \), it is best to work backwards, now that we know the answer, and to consider the Laplace transform of the function

$$h(t) = \int_{0}^{t} f(t_1)g(t-t_1) \, dt_1,$$  \hfill (1.39)

which we shall call the convolution (in German, \textit{Faltung}) of \( f \) and \( g \).

We have

$$\int_{0}^{\infty} h(t)e^{-st} \, dt = \int_{0}^{\infty} e^{-st} \left[ \int_{0}^{t} f(t_1)g(t-t_1) \, dt_1 \right] \, dt.$$  \hfill (1.40)
THE LAPLACE TRANSFORM

Considering the shaded region, $S$,

\[ \int \int e^{-st} f(t_1) g(t - t_1) \mathrm{d}t \mathrm{d}t_1 = \int_0^R e^{-st} \left[ \int_0^t f(t_1) g(t - t_1) \mathrm{d}t_1 \right] \mathrm{d}t \]

\[ = \int_0^R f(t_1) \left[ \int_0^R e^{-st} g(t - t_1) \mathrm{d}t \right] \mathrm{d}t_1 \]

\[ = \int_0^R e^{-st} f(t_1) \left[ \int_{R-t_1}^R e^{-su} g(u) \mathrm{d}u \right] \mathrm{d}t_1. \]

(1.42)

As $R \to \infty$ we obtain, formally, $L(f)L(g)$. Let us now prove

**THEOREM 1.3.** If

(a) \[ \int_0^\infty e^{-st} \vert f(t_1) \vert \mathrm{d}t_1 < \infty, \] and

(b) \[ \int_0^\infty e^{-s(i+1)t} g(t_1) \mathrm{d}t_1 \] converges,

then

\[ \int_0^\infty b(t) e^{-st} \mathrm{d}t = \left[ \int_0^\infty e^{-st} f(t) \mathrm{d}t \right] \left[ \int_0^\infty e^{-st} g(t) \mathrm{d}t \right] \]

(1.44)

for $s = a + ib$, and generally for $\Re(s) > a$, where $b$ is the convolution of $f$ and $g$.

**PROOF.** We have, from Eq. (1.42),

\[ \int_0^R b(t) e^{-st} \mathrm{d}t = \int_0^R e^{-st} f(t_1) \left[ \int_{R-t_1}^R e^{-su} g(u) \mathrm{d}u \right] \mathrm{d}t_1 \]

\[ = \int_0^R e^{-st} f(t_1) \left[ \int_{R-t_1}^\infty e^{-su} g(u) \mathrm{d}u \right] \mathrm{d}t_1 \]

\[ - \int_0^R e^{-st} f(t_1) \left[ \int_0^{R-t_1} e^{-su} g(u) \mathrm{d}u \right] \mathrm{d}t_1. \]

(1.45)

To obtain an estimate for the second integral, we break the range of integration up into
[0, R/2] and [R/2, R]. Since \( \int_{0}^{\infty} e^{-nu} g(u) \, du \) converges, we have

\[
\begin{align*}
(a) & \quad \left| \int_{0}^{\infty} e^{-nu} g(u) \, du \right| \leq \epsilon \quad \text{for } N \geq N_{0}(\epsilon), \text{ and} \\
(b) & \quad \left| \int_{N}^{\infty} e^{-nu} g(u) \, du \right| \leq c_{1} \quad \text{for all } N \geq 0. 
\end{align*}
\]

Therefore,

\[
\begin{align*}
& \quad \left| \int_{0}^{\infty} e^{-st} f(t) \left[ \int_{0}^{\infty} e^{-nu} g(u) \, du \right] \, dt_{1} \right| \\
& \quad \leq \left[ \int_{0}^{\infty} e^{-st} \, dt_{1} \right] \epsilon \\
& \quad \leq \epsilon \int_{0}^{\infty} e^{-st} |f(t_{1})| \, dt_{1} = c_{1} \epsilon
\end{align*}
\]

and

\[
\begin{align*}
& \quad \left| \int_{b_{0}}^{R} e^{-st} f(t) \left[ \int_{b_{0}}^{\infty} e^{-nu} g(u) \, du \right] \, dt_{1} \right| \leq c_{1} \int_{b_{0}}^{R} e^{-st} |f(t_{1})| \, dt_{1} \leq \epsilon c_{1}
\end{align*}
\]

for \( R \geq N_{0}(\epsilon). \)

From these inequalities it follows that \( \lim_{R \to \infty} \int_{0}^{R} b(t) e^{-st} \, dt = L(f) L(g). \)

From the above theorem it can be shown that Eq. (1.44) holds whenever all three integrals exist. The proof is merely an application of Abel summability for integrals (see [3]). In most of our applications \( L(f) \) and \( L(g) \) will be absolutely convergent, obviating the need to go into finer details.

As a simple application, consider the problem of finding the Laplace inverse of \( F(t)/t \), given \( L^{-1}(F) = f \). Since \( L^{-1}(1/t) = 1 \), we have

\[
L^{-1} \frac{F}{t} = \int_{0}^{t} f(t - t_{1}) \, dt_{1} = \int_{0}^{t} f(t_{1}) \, dt_{1},
\]

a result which is readily verified.

1.7. A Real Inversion Formula

In Sec. 1.4 we derived a complex inversion formula for the Laplace transform. Despite its theoretical elegance, this formula is of no value if \( F(s) \) is an empirical function known only for real values of \( s \). Although there are many different techniques for computing the inverse of a function of this type, depending on interpolation techniques, expansion into orthogonal series by means of Laguerre polynomials, and so on, it must be admitted that there is no one uniformly successful method.
The method we describe below has the merit of requiring only real values of \( F(s) \), but has the demerit of requiring arbitrarily high derivatives. Consequently, it is unsuit for the inversion of empirical functions. For functions of simple analytical form which escape a table of inverses, it may be useful. We shall present the result, together with its heuristic basis, and a numerical application, referring the interested reader to [1] and [4] for a proof and for related results, where similar formulae involving differences rather than derivatives are given.

**Theorem 1.4 (Post-Widder Inversion Formula).** If

\[
F(s) = \int_0^\infty e^{-st} f(t) \, dt, \quad Re(s) > 0,
\]

then whenever \( f(t) \) is continuous and of bounded variation

\[
f(t) = \lim_{k \to \infty} \frac{(-1)^k}{k!} F^{(k)} \left( \frac{1}{s} \right) \left( e^{i\pi t} \right)^{k+1}.
\]

To obtain Eq. (1.51) heuristically, we let \( \phi(x) \) be the Fourier transform of \( f(t) \),

\[
\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-ixt} \, dt.
\]

Then, since

\[
\frac{1}{s - ix} = \int_{-\infty}^{\infty} e^{-st} e^{ixt} \, dt, \quad Re(s) > 0,
\]

we obtain, using the Parseval-Plancherel theorem (see Titchmarsh [3]),

\[
\int_{-\infty}^{\infty} e^{-st} f(t) \, dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi(x) \, dx.
\]

In order to obtain \( f(t) \), we want an operation which converts \( 1/(s - ix) \) into \( e^{ixt} \), since we know that

\[
f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi(x) e^{ixt} \, dx.
\]

Since we have

\[
\lim_{k \to \infty} \left( 1 - \frac{ix}{k} \right)^k = e^{-ixt},
\]

we can obtain an operation of this type using repeated differentiation. We have

\[
F^{(k)}(t) = \frac{(-1)^k}{k!} \int_{-\infty}^{\infty} \frac{\phi(x)}{(s - ix)^{k+1}} \, dx.
\]
Hence
\[ \frac{(-1)^k}{k!} \left( \frac{k}{i} \right)^{k+1} F^{(k)} \left( \frac{k}{i} \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\phi(x)}{1 - \frac{ix}{k}} e^{ix} \, dx. \] (1.58)

It is now not difficult to show that the formal limit actually exists if, for example,
\[ \int_{-\infty}^{\infty} e^{-\epsilon t} |f(t)| \, dt < \infty \] for some \( \epsilon > 0 \). The customary proof, as in Widder, proceeds along different principles from the above.

As a numerical illustration, consider \( F(t) = e^{-\sqrt{t}} \), whose Laplace inverse is \( f(t) = e^{-\sqrt{t}/2} \sqrt{\pi} \Gamma^k \). Set
\[ u_k \left( \frac{1}{t} \right) = \frac{(-1)^k}{k!} F^{(k)} \left( \frac{1}{i} \right) \left( \frac{k}{i} \right)^{k+1} \] (1.59)

Table 1 shows that even if the convergence of \( u_k(1) \) to \( u(1) \) is not very rapid, one quickly obtains an excellent qualitative picture of the function.

<table>
<thead>
<tr>
<th>( t )</th>
<th>( u_1 )</th>
<th>( u_2 )</th>
<th>( u_3 )</th>
<th>( u(1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>.5108</td>
<td>.4191</td>
<td>.3612</td>
<td>.1700</td>
</tr>
<tr>
<td>0.10</td>
<td>.6593</td>
<td>.6989</td>
<td>.7075</td>
<td>.7322</td>
</tr>
<tr>
<td>0.15</td>
<td>.6508</td>
<td>.7347</td>
<td>.7753</td>
<td>.9172</td>
</tr>
<tr>
<td>0.20</td>
<td>.5975</td>
<td>.6964</td>
<td>.7455</td>
<td>.9036</td>
</tr>
<tr>
<td>0.25</td>
<td>.5413</td>
<td>.6429</td>
<td>.6883</td>
<td>.8302</td>
</tr>
<tr>
<td>0.30</td>
<td>.4902</td>
<td>.5829</td>
<td>.6270</td>
<td>.7461</td>
</tr>
<tr>
<td>0.40</td>
<td>.4066</td>
<td>.4834</td>
<td>.5179</td>
<td>.5969</td>
</tr>
<tr>
<td>0.50</td>
<td>.3458</td>
<td>.4060</td>
<td>.4322</td>
<td>.4839</td>
</tr>
<tr>
<td>0.60</td>
<td>.2959</td>
<td>.3463</td>
<td>.3661</td>
<td>.4001</td>
</tr>
<tr>
<td>0.70</td>
<td>.2584</td>
<td>.2996</td>
<td>.3147</td>
<td>.3370</td>
</tr>
<tr>
<td>0.80</td>
<td>.2284</td>
<td>.2624</td>
<td>.2740</td>
<td>.2884</td>
</tr>
<tr>
<td>0.90</td>
<td>.2041</td>
<td>.2323</td>
<td>.2412</td>
<td>.2503</td>
</tr>
<tr>
<td>1.00</td>
<td>.1839</td>
<td>.2075</td>
<td>.2144</td>
<td>.2197</td>
</tr>
<tr>
<td>1.50</td>
<td>.1203</td>
<td>.1307</td>
<td>.1324</td>
<td>.1300</td>
</tr>
<tr>
<td>2.00</td>
<td>.0872</td>
<td>.0920</td>
<td>.0919</td>
<td>.0880</td>
</tr>
<tr>
<td>2.50</td>
<td>.0672</td>
<td>.0693</td>
<td>.0686</td>
<td>.0646</td>
</tr>
<tr>
<td>3.00</td>
<td>.0540</td>
<td>.0546</td>
<td>.0536</td>
<td>.0499</td>
</tr>
<tr>
<td>3.50</td>
<td>.0447</td>
<td>.0445</td>
<td>.0435</td>
<td>.0401</td>
</tr>
<tr>
<td>4.00</td>
<td>.0379</td>
<td>.0372</td>
<td>.0361</td>
<td>.0331</td>
</tr>
</tbody>
</table>

A discussion of other techniques particularly suited to functions given numerically will be found in Widder [5].

*These results were computed by O. Gross.
REFERENCES

CHAPTER 2
THE LINEAR DIFFERENTIAL-DIFFERENCE EQUATION
WITH CONSTANT COEFFICIENTS

2.1. Introduction
In this chapter we shall begin our study of the application of the Laplace transform
to the problem of the solution of linear functional equations by considering the linear
differential-difference equation with constant coefficients,

$$\sum_{k=0}^{N} a_k u^{[k]}(t - d_k) = g(t), \quad t > d_N,$$

(2.1)

where $0 < d_1 < d_2 < \cdots < d_N$ and where $u(t)$ is prescribed in some initial interval

$$u(t) = f(t), \quad 0 \leq t \leq d_N.$$  

(2.2)

The method is sufficiently illustrated by considering in detail the first-order equation

$$\frac{du(t)}{dt} = au(t) + bu(t - 1) + g(t), \quad t > 1,$$

$$u(t) = f(t), \quad 0 \leq t \leq 1.$$  

(2.3)

After presenting two explicit expressions for the solution of Eq. (2.3)—one in terms of contour integrals and one in terms of real integrals—we shall discuss the asymptotic behavior of the solution and then turn to a discussion of various types of Fourier expansions associated with these equations.

2.2. Formal Aspects
As usual in the theory of linear functional equations, the solution is composed of two
parts, one owing its presence to the initial function $f(t)$ and the other being generated
by the forcing term $g(t)$.
Thus we may set $u = v + w$, where

$$\begin{align*}
v'(t) &= av(t) + bv(t - 1) + g(t), \quad t > 1, \\
v(t) &= 0, \quad 0 \leq t \leq 1, \\
w'(t) &= aw(t) + bw(t - 1), \\
w(t) &= f(t), \quad 0 \leq t \leq 1.
\end{align*}$$

(2.4)
There is no advantage in using this decomposition here, since we shall employ a uniform technique to solve both types of equations.

The change of variable, \( u = e^{st}v \), converts Eq. (2.3) into

\[
v'(t) = be^{s(t-1)} + e^{st}g(t), \quad t > 1, \\
v(t) = e^{st}f(t), \quad 0 \leq t \leq 1,
\]

which is a slight simplification. We shall consider, therefore, the simpler equation

\[
u'(t) = cu(t-1) + g(t), \quad t > 1, \\
u(t) = f(t), \quad 0 \leq t \leq 1.
\]

To begin with, we shall proceed formally, and then we shall discuss the modifications required to make our procedure rigorous. From Eq. (2.6) we obtain, upon multiplying by \( e^{-st} \) and integrating between 1 and \( \infty \) with respect to \( t \),

\[
\int_1^\infty u'(t)e^{-st} \, dt = c \int_1^\infty u(t-1)e^{-st} \, dt + \int_1^\infty e^{-st}g(t) \, dt \\
= ce^s \int_0^\infty u(t)e^{-st} \, dt + \int_1^\infty e^{-st}g(t) \, dt.
\]

Upon integration by parts, the left-hand side becomes

\[
-u(1)e^{-s} + \int_1^\infty u(t)e^{-st} \, dt.
\]

Using the fact that \( u(t) \) is given in the interval \([0, 1]\), we obtain

\[
\int_0^\infty u(t)e^{-st} \, dt = \int_0^1 f(t)e^{-st} \, dt + \int_1^\infty u(t)e^{-st} \, dt.
\]

Combining these relations, Eq. (2.7) yields

\[
\left[ \int_1^\infty u(t)e^{-st} \, dt \right] (s - ce^s) = f(1)e^{-s} + ce^s \int_0^1 f(t)e^{-st} \, dt \\
+ \int_1^\infty e^{-st}g(t) \, dt,
\]

an equation which determines \( \int_1^\infty u(t)e^{-st} \, dt \) in terms of known quantities.

Employing the inversion formula for the Laplace transform, we obtain

\[
u(t) = \int \frac{e^{st}f(1)e^{-st} + ce^s \int_0^1 f(t)e^{-st} \, dt + \int_1^\infty e^{-st}g(t) \, dt}{s - ce^s} \, dt.
\]
for \( t > 1 \), where the line of integration is to the right of all zeros of \( s - c e^{-a} \). We shall discuss the possibility of this in a moment.

If we attempt to use the expression given in Eq. (2.11) in order to determine \( u(t) \) in \([0, 1]\), we obtain zero, owing to the fact that we obtained this result from the expression for \( \int_0^t u(t) e^{-at} \, dt \). Subsequently, in Sec. 2.6, we shall obtain an expression which is valid for \( 0 \leq t \leq 1 \).

Let us observe that we may write

\[
\begin{align*}
\frac{u(t)}{s - c e^{-a}} &= \int_0^t e^{st} \, f(t) e^{-st} \, dt + \int_0^t e^{st} \, g(t) \, dt,
\end{align*}
\]

where the line of integration is to the right of all zeros of \( s - c e^{-a} \), and

\[
\begin{align*}
\frac{u(t)}{s - c e^{-a}} &= \int_0^t e^{st} \, f(t) e^{-st} \, dt + \int_0^t e^{st} \, g(t) \, dt,
\end{align*}
\]

a form which very clearly shows the division of influence between the initial and the forcing functions.

At first sight it may seem odd that the value of \( f(t) \) at the single point \( t = 1 \) should play such an important role. If, however, we shrink the time lag to zero, the integral disappears and the term \( f(1) \) is replaced by \( f(0) \). The relation then reduces to the familiar one of defining a solution of a linear differential equation in terms of the initial value.

2.3. Rigorous Derivation

It is clear that Eq. (2.6) has a unique solution which may be determined by recurrence from interval to interval by means of the relation

\[
\begin{align*}
\frac{u(t)}{s - c e^{-a}} &= \int_0^t e^{st} \, f(t) e^{-st} \, dt + \int_0^t e^{st} \, g(t) \, dt, \quad t \geq 1,
\end{align*}
\]

assuming, as we may, that \( f(t) \) is continuous, or merely integrable, in \([0, 1]\).

In order to employ the Laplace transform, we must impose some exponential bound upon \( g(t) \). Let us assume that for some \( b \geq 0 \)

\[
|g(t)| \leq ab^t, \quad t \geq 1.
\]

Under this assumption we shall show that \( u \) satisfies an inequality of the form \( |u| \leq de^{bt} \). Assuming that this inequality holds in \([0, N]\), we have in \([N, N + 1]\)

\[
|u(t)| \leq |f(1)| + \epsilon \int_1^t de^{b(t-1)} \, dt + a \int_1^t e^{bt} \, dt.
\]
Choosing $k \geq b$, we have

$$|u(t)| \leq |f(1)| + \left| e \cdot \frac{d e^{-k} + a}{k} e^{b t} \right| \leq d e^{b t} \tag{2.16}$$

for $t \geq 1$, if $k$ is sufficiently large.

It follows that $\int_{1}^{\infty} u(t) e^{-s t} d t$ converges absolutely and represents an analytic function for $Re(s) > k$, and, consequently, that $s - c e^{-s} \neq 0$ for $Re(s) > k$.

We have thus established rigorously

**Theorem 2.1.** If

$$u'(t) = cu(t - 1) + g(t), \quad t \geq 1,*$$

$$u(t) = f(t), \quad 0 \leq t \leq 1, \tag{2.17}$$

and

(a) \hspace{1cm} |g(t)| \leq a e^{b t}, \quad t \geq 1,

(b) \hspace{1cm} f(t) \text{ is integrable in } [0, 1], \tag{2.18}

the solution of Eq. (2.17) is given by

$$u(t) = \int_{(b, t)} f(1) e^{-s} + c e^{-s} \int_{(b, t)} f(t) e^{-s t} d t + \int_{(b, t)}^{\infty} e^{-s t} g(t) \frac{d t}{s - c e^{-s}} e^{s t} d s \tag{2.19}$$

for $t > 1$, provided that $b_1$ is sufficiently large.

### 2.4. A Real Representation of the Solution

Having obtained the solution in the form of a contour integral, it is natural to ask whether or not this solution can be represented as a real integral. As we shall see, the contour integral representation is exceedingly useful in determining the asymptotic behavior of the solution. On the other hand, it has its drawbacks for other purposes, particularly for stability theory.

In our search for a real representation we may be guided by the corresponding results in the theory of differential equations. The first-order linear equation

$$u'(t) = a u(t) + g(t), \quad u(0) = c, \tag{2.20}$$

under the same assumption concerning $g(t)$, has the solution

$$u(t) = \int_{(b, t)} c + \int_{(b, t)}^{\infty} g(t) e^{-s t} d t \frac{d t}{s - a} e^{s t} d s, \quad b > a, \tag{2.21}$$

which is obtained by using the above procedure.

---

*The derivative here is to be interpreted as a right-hand derivative.
On the other hand, the method of the integrating factor—which must be replaced by the Lagrange variation-of-parameters technique in the general case of nth order linear inhomogeneous equations, or else it must be reinterpreted in terms of vector-matrix notation (see Ref. [2] of Chapter 7)—yields the simple expression

$$ u(t) = e^{st} + \int_{0}^{t} e^{s(t-t_1)} g(t_1) \, dt_1. \quad (2.22) $$

This formula shows very clearly what we know to be true, namely, that $u(t)$ is only influenced by values of $g(t_1)$ for $0 \leq t_1 \leq t$, whereas Eq. (2.21) does not show this immediately.

To obtain Eq. (2.22) from Eq. (2.21), we write

$$ u(t) = \int_{b}^{t} \frac{e^{st}}{s-a} \, ds + \int_{b}^{t} \frac{\int_{a}^{s} g(t_1)e^{s-t_1} \, dt_1}{s-a} \, e^{st} \, ds $$

$$ = e^{st} + \int_{b}^{t} \frac{\int_{a}^{s} g(t_1)e^{s-t_1} \, dt_1}{s-a} \, e^{st} \, ds. \quad (2.23) $$

Since $1/(s-a)$ is the transform of $e^{st}$, the convolution theorem tells us that

$$ \left[ \int_{a}^{s} g(t_1)e^{s-t_1} \, dt_1 \right]/(s-a) \text{ is the transform of } \int_{a}^{s} e^{s(t-t_1)} g(t_1) \, dt_1. $$

This shows the equality of Eqs. (2.22) and (2.23).

For many purposes, Eq. (2.22) is a more useful representation, as it is, for example, in a discussion of the stability of solutions of nonlinear equations (see Chapter 8).

Let us now apply the same treatment to Eq. (2.19). According to Theorem 1.2, we know that $1/(s-c e^{s})$ is the Laplace transform of the function $K(t)$, defined by

$$ K(t) = \int_{b}^{t} \frac{e^{st}}{s-c e^{s}} \, ds, \quad (2.24) $$

where $b$ is sufficiently large to ensure that the real parts of the roots of $s-c e^{s} = 0$ are to the left of the line $s = b + iu, -\infty < u < \infty$. (See the discussion in Sec. 2.5 on this point.)

This function has the following properties:

(a) \hspace{1cm} K(t) = 0, \hspace{1cm} t < 0,

(b) \hspace{1cm} K'(t) = c K(t - 1), \hspace{1cm} t > 1,

(c) \hspace{1cm} K(t) = 1, \hspace{1cm} 0 < t < 1. \quad (2.25) $$

That Eq. (2.25a) holds we observe immediately by shifting the contour to $+\infty$. To establish Eq. (2.25b) requires a little more effort, since direct differentiation under the sign of integration is not legitimate. To circumvent this difficulty, we employ the following useful device.
We have, for \( t > 0 \),

\[
K(t) = \int_0^t \frac{e^{st}}{s} \, ds + \int_0^t e^{st} \left[ \frac{1}{s - ce^{-s}} - \frac{1}{s} \right] \, ds
\]

\[
= 1 + \int_0^t \frac{e^{st}}{s} \frac{ce^{-s}}{s - ce^{-s}} \, ds
\]

\[
= 1 + c \int_0^t \frac{e^{s(t-1)}}{s(t - ce^{-s})} \, ds.
\]

(2.26)

It is not difficult to show that it is now legitimate, for \( t > 1 \), to differentiate under the sign of integration, so that we obtain

\[
K'(t) = c \int_0^t \frac{e^{s(t-1)}}{s - ce^{-s}} \, ds = cK(t - 1).
\]

(2.27)

The representation in Eq. (2.26) also shows that \( K(t) = 1 \) for \( 0 < t < 1 \).

Turning to Eq. (2.19), above, we see that we have demonstrated, with the aid of the convolution theorem,

**Theorem 2.2.** The solution of

\[
u'(t) = cu(t - 1) + g(t), \quad t > 1, \\
u(t) = f(t), \quad 0 \leq t \leq 1,
\]

is given by

\[
u(t) = f(1)K(t - 1) + c \int_0^t K(t - t_1 - 1)f(t_1) \, dt_1 + \int_t^1 K(t - t_1)g(t_1) \, dt_1,
\]

(2.29)

for \( t > 1 \), where

\[
K(t) = \int_0^t \frac{e^{st}}{s - ce^{-s}} \, ds,
\]

(2.30)

and \( b \) is sufficiently large.

Note that we no longer require any bound on \( g(t) \) near infinity, since Eq. (2.29) may be verified to be a solution by direct substitution.

Precisely the same techniques may be employed to obtain the solution of the general linear equation with constant coefficients given in Eq. (2.1). In the still more general case in which we have systems of linear differential-difference equations, it is convenient, and almost essential if we wish to understand the formalism, to use vector-matrix notation. With the aid of this notation, the procedure is the same for all dimensions.

If we consider equations of the form

\[
u'(t) = cu(t + 1) + g(t),
\]
in which the present depends on the future, the situation is much more complex, since the associated exponential polynomial, \( s - ce^s \), has roots with arbitrarily large positive real parts (see Theorem 6.1).

If we employ an expression such as Eq. (2.19) or Eq. (2.24), we see that the value of the integral will depend on \( b \), no matter how large \( b \) is. Problems of this type are of little or no practical interest.

### 2.5. Asymptotic Behavior of the Solution

As we see from the representation for the solution afforded by Theorem 2.2, the asymptotic behavior of \( u(t) \) as \( t \to \infty \) depends strongly on \( K(t) \), whose asymptotic behavior, in turn, depends on the location of the roots of \( s - ce^s = 0 \). The general theory of the location of the zeros of transcendental functions of this type will be discussed in greater detail in Chapter 6. Here we shall consider a very simple case in order to illustrate the importance of this discussion. Let us assume that \( c > 0 \). Then it is easy to demonstrate

**Lemma 2.1.** If \( c > 0 \), the root of \( s - ce^s \) with largest real part is positive.

**Proof.** The equation \( s e^s - c = 0 \) has a positive root, \( r \), since \( c > 0 \). If \( s \) is another root, we have \( |se^s| = c \), or

\[
c = |s| e^{Re(s)} \geq Re(s) e^{Re(s)}.
\]

From this it follows that \( Re(s) < r \), unless \( s = r \). Furthermore, this inequality holds uniformly; i.e., \( Re(s) \leq r - d \), where \( d > 0 \). This follows from the fact that if \( s = u + iv \) is a root, then \( u \to -\infty \) as \( v \to \pm \infty \), as we see from the equation.

Consequently, for \( |v| \leq v_0 \), there are no roots of the form \( u + iv \) with \( r - u \leq d \) for sufficiently small \( d \). The rectangle formed by \( r \pm iv_0 \) and \( u \pm iv_0 \) for any \( u < r \), contains only a finite number of roots of \( s - ce^s = 0 \), and hence one with the real part nearest to \( r \); let \( d \) be this minimum distance between real parts.

On the line \( r - d/2 + iv \), \( -v_0 \leq v \leq v_0 \), \( |s - ce^s| \) has a nonzero minimum, \( m \), since \( |s - ce^s| \geq v - ce^{(r-d/2)} \) when \( v \) is large, and \( |s - ce^s| \neq 0 \) on the line when \( v \) is small.

Having disposed of these preliminaries, let us return to the question of the asymptotic behavior of

\[
K(t) = \int_{b}^{t} \frac{e^{st}}{s - ce^s} \, ds, \quad b > r.
\]

Shifting the line of integration past the line \( r + iv \), we obtain a residue, \( e^{rt}/(1 + ce^{r}) \), from the simple pole at \( s = r \). Taking the new line of integration to be \( r - d/2 + iv \), we have

\[
K(t) = \frac{e^{rt}}{1 + ce^{r}} + \int_{r-d/2}^{t} \frac{e^{st}}{s - ce^s} \, ds.
\]
To estimate the second integral, we use the device employed in Eq. (2.26), by writing

\[ \int_{(r-d/2)}^{r} \frac{e^{it}}{s - ce^{-t}} \, dt = \int_{(r-d/2)}^{r} \frac{e^{it}}{s} \, dt + \int_{(r-d/2)}^{r} \frac{ce^{it(1-\epsilon)}}{s(s-ce^{-t})} \, dt. \]  

(2.34)

The first integral is 0, or 1, depending on the value of \( r - d/2 \), while the second integral is \( O[e^{t(r-d/2)}] \), since

\[ \int_{(r-d/2)}^{r} \frac{dt}{s(s-ce^{-t})} \]  

(2.35)

is convergent.

Continuing this process, we see that we obtain

**Theorem 2.3.** For \( t > 0 \), we have

\[ K(t) = \sum_{\nu=1}^{\infty} \frac{e^{
u t}}{1 + ce^{-\nu s}}, \]  

(2.36)

where \( \{\nu_i\} \) are the roots of \( s - ce^{-s} = 0 \) arranged in decreasing order of real parts.

Turning to Eq. (2.28), above, we see that this representation suffices to describe the asymptotic behavior of \( \int_{t_1}^{t} K(t - t_1 - 1)f(t_1) \, dt_1 \). In order to derive that of \( \int_{t_1}^{t} K(t - t_1)g(t_1) \, dt_1 \), it is necessary to make some assumption concerning the asymptotic behavior of \( g(t) \).

Let us now show that the asymptotic behavior of the roots of \( s - ce^{-s} = 0 \) may be readily ascertained. Let \( s = \nu + iv \). The equations for \( \nu \) and \( v \) are

(a) \[ \nu = ce^{-\nu} \cos v, \]  

(b) \[ v = -ce^{-\nu} \sin v. \]  

(2.37)

It is clear to begin with that the roots, as the zeros of an entire function, can have no finite point of accumulation. From Eqs. (2.37a) and (2.37b), we see that the real and imaginary parts have no finite points of accumulation. Hence, if \( \nu_n = \nu_n + iv_n \), as \( n \to \infty \), both \( |\nu_n| \) and \( |v_n| \) increase without bound. It is further clear from Eq. (2.37a) that \( \nu_n \to -\infty \), and thus that \( \cos v_n \to 0 \). Let us now take \( n \) large and determine the form of \( \nu_n \) and \( v_n \); for convenience let us drop the subscripts.

Since \( \cos v \) is close to zero, we must have

\[ v = \frac{\pi}{2} + 2N\pi + o(1), \]  

(2.38)

for \( |N| \) large. Since the roots occur in conjugate pairs, it is sufficient to consider positive \( N \). Using Eq. (2.37b), this yields

\[ e^{-\nu} = \frac{v}{c \sin v} \sim \frac{v}{c}, \]  

(2.39)
or

\[ -u \approx \log \left( \frac{\pi}{2} + \frac{2N\pi}{c} + o(1) \right) \approx \log \left( \frac{2N\pi}{c} \right) + o(1). \]  \hspace{1cm} (2.40)

By combining, we have, as a first approximation to \( s = u + iv \),

\[ s = \log \left( \frac{2N\pi}{c} \right) + \left( \frac{\pi}{2} + 2N\pi \right) + o(1). \]  \hspace{1cm} (2.41)

It is easy to see that this procedure may be continued to yield arbitrarily accurate approximations to \( s \), in the form of asymptotic series.

2.6. Fourier-type Expansions

In the previous sections we have concentrated on finding an analytic expression for the solution of the differential-difference equation which would be valid for \( t > 1 \). In doing this we observed that we obtained a series for \( K(t) \), given in Eq. (2.36), dependent on the roots of \( s - ce^{-s} = 0 \), which was vaguely reminiscent of the Fourier series. Let us now show that these series are actually of this type by using the previous technique to obtain a contour integral representation for \( u(t) \) in [0, 1], which is to say, a representation for \( j(t) \).

Let us consider the homogeneous equation

\[
\begin{align*}
   u'(t) &= cu(t - 1), \quad t > 1, \\
   u(t) &= f(t), \quad 0 \leq t \leq 1,
\end{align*}
\]  \hspace{1cm} (2.42)

where we shall assume that \( f(t) \) is continuous and of bounded variation. We obtain, as before,

\[ -u(1)e^{-s} + s \int_{1}^{\infty} u(t)e^{st} \, dt = ce^{-s} \int_{0}^{\infty} u(t)e^{st} \, dt. \]  \hspace{1cm} (2.43)

Let us now vary our procedure and solve for \( \int_{0}^{\infty} u(t)e^{st} \, dt \) instead of for \( \int_{1}^{\infty} u(t)e^{st} \, dt \), obtaining in this way

\[
\int_{0}^{\infty} u(t)e^{st} \, dt = \frac{f(1)e^{-s} + s \int_{1}^{\infty} f(t) e^{st} \, dt}{s - ce^{-s}}.
\]  \hspace{1cm} (2.44)

Employing the inversion formula, we obtain for \( t > 0, b > r \),

\[
\begin{align*}
   u(t) &= \int_{0}^{1} \frac{f(1)e^{-s\tau}}{s - ce^{-s}} \, d\tau + \int_{1}^{\infty} \left[ \int_{1}^{\infty} f(t) e^{-st} \, dt \right] e^{st} \, ds \\
   &= \int_{0}^{1} \left[ \int_{1}^{\infty} f(t) e^{st} \, dt \right] e^{st} \, ds,
\end{align*}
\]  \hspace{1cm} (2.45)
if we restrict \( t \) to be between 0 and 1, since, for \( t < 1 \),

\[
\int_{(0)}^{t} \frac{e^{s(t-1)}}{s - ce^{s}} \, ds = 0,
\]

(2.46)
as was pointed out in Eq. (2.25a).

Shifting the line of integration to \( s = \rho + iv \), where \( \rho \) is a large negative quantity, we obtain

\[
u(t) = \sum_{n=1}^{\infty} \frac{\gamma e^{\rho t}}{1 + \rho e^{-\rho}} \int_{0}^{1} f(t_1) e^{-\rho t_1} \, dt_1 \\
+ \int_{(\rho)} \frac{s}{s - ce^{s}} \left[ \int_{0}^{1} f(t_1) e^{-\rho t_1}, dt_1 \right] e^{st} \, ds.
\]

(2.47)

In order to justify pushing \( \rho \) to \(-\infty\) and eliminating the contour integral, we must show that the contribution of the contour integral is negligible for large negative \( \rho \). This is not difficult to do under the assumption that \( f(t) \) is of bounded variation, which permits integration by parts in \( \int_{0}^{1} f(t_1) e^{-\rho t_1}, dt_1 \). We shall not go into the details here, since we shall not use any of these results in what follows. Rigorous results are given in [1].

To obtain the ordinary Fourier series expansion of a periodic function, consider the difference equation

\[
u(t + 1) - \nu(t) = 0, \quad t > 1, \\
\nu(t) = f(t), \quad 0 \leq t \leq 1.
\]

(2.48)
Employing the Laplace transform technique, we obtain,

\[
u(t) = \int_{(b)} \left[ \int_{0}^{t} f(t_1) e^{-\rho t_1}, dt_1 \right] e^{st} \, ds, \quad b > 0,
\]

(2.49)
for \( t > 0 \), under the assumption that \( f(t) \) is continuous and of bounded variation in \([0, 1]\). Since the zeros of \( e^{s} - 1 \) are at \( s = 0, \pm 2\pi i, \cdots, \pm 2n\pi i, \cdots \), we obtain, formally, the ordinary Fourier series expansion

\[
f(t) = \sum_{n=-\infty}^{\infty} \left[ \int_{0}^{1} f(t_1) e^{2\pi i t_1}, dt_1 \right] e^{2\pi i t}
\]

(2.50)
for \( 0 < t < 1 \). If \( f(t) \) is merely of bounded variation, we have as the sum \( [f(t + 0) + f(t - 0)]/2 \).

In a similar fashion, each differential-difference equation gives rise to a Fourier-type expansion. Conversely, if we wish to expand a function \( f(t) \) in the form

\[
f(t) = \sum_{n=-\infty}^{\infty} a_n e^{\rho n t},
\]

(2.51)
where the \( \{ \lambda_n \} \) are the roots of an exponential polynomial \( P(t, e^{i\lambda}) = 0 \), we may obtain the expansion by considering the related differential-difference equation.

As an example, consider the problem of determining the expansion

\[
f(t) = \sum_{n=-\infty}^{\infty} a_n e^{i\lambda_n t}, \quad -1 \leq t \leq 1,
\]

where \( \lambda_n \) runs over the roots of \( \cos t + at \sin t = 0 \). By writing

\[
\cos t + at \sin t = \frac{e^{it} + e^{-it}}{2} + \frac{at}{2i} (e^{it} - e^{-it}),
\]

and then \( it = s \), we see that the relevant exponential polynomial is that obtained from

\[
(e^{s} + e^{-s}) - as(e^{s} - e^{-s}) = 0,
\]

or, since \( e^{s} \neq 0 \),

\[
e^{2s}(1 - at) + (1 + at) = 0.
\]

The corresponding differential-difference equation is

\[
u(t + 2) - au(t + 2) + u(t) + au'(t) = 0, \quad t > 2,
\]

\[
u(t) = f(it), \quad 0 \leq t \leq 2,
\]

since Eq. (2.55) is obtained from Eq. (2.56), if we attempt a particular solution of the form \( e^{as} \).

### 2.7. Differential-difference Equations and Causality

The preceding results concerning the asymptotic behavior of solutions of differential-difference equations lead to some observations concerning causality and mathematical formulations which may perhaps be useful in constructing mathematical models of the physical phenomena.

The usual mathematical formulation is by means of a system of differential equations, ordinary or partial, having the property that the solution is determined for all subsequent time by its state at some particular time instant. In classical mechanics the state of a system is determined by numerical values, while in nonclassical mechanics the state of a system is described by a probability distribution. In both cases, the systems are causal in the above sense.

There are, however, many simple physical phenomena—such as magnetic hysteresis and various kinds of elastic behavior, and various models of economic and psychological behavior—in which it is necessary to know not only the present state, but also how it was attained, if one wishes to predict the future. The second case can always be reduced to the first case at the expense of increasing the number of parameters which describe the system. In some cases, this may mean replacing functions by functionals.
Simple types of functional equations describing physical systems with hereditary effects are the differential-difference equations discussed above,

$$u'(t) = cu(t - 1),$$  \hspace{1cm} (2.57)

and the renewal equation to be discussed,

$$u(t) = f(t) + \int_{0}^{t} u(t - s) g(s) \, ds.$$  \hspace{1cm} (2.58)

It has recently been suggested that causality may have to be abandoned in the discussion of physical effects involving very small times or distances. It is probably true that the real meaning of this suggestion is that many more dependent variables are required to specify the state of a system as the time interval gets shorter and shorter.

Let us now point out that a theory based on various types of functional equations, such as Eq. (2.57), and in particular one based on differential-difference equations has precisely the properties required to bridge the gap between theories valid for large time intervals and theories valid for small time intervals.

Consider Eq. (2.57), which requires for the complete determination of $u(t)$ the knowledge of $u(t)$ in some interval of length one. For large $t$, we know that $u(t)$ has the form $c_{1}e^{rt}$, where $r$ is the positive root of $xe^{r} = c$, plus terms of much smaller order. Consequently, for large $t$, $u(t)$ acts like the solution of

$$u'(t) = ru(t), \quad u(0) = c_{1},$$  \hspace{1cm} (2.59)

where the solution is specified completely by the value at one time instant.

As $t$ decreases toward 1, more and more of the past history of $u(t)$ is required. We see then that a formulation in terms of differential-difference equations yields a description which is causal in the classical sense for large $t$, and becomes less and less causal as the time interval gets smaller and smaller.

References


(Note: The article by Pringsheim in Enzyklopädie der mathematischen Wissenschaften will also be of interest, particularly in connection with Sec. 2.5. The method used there is essentially due to Cauchy.)
CHAPTER 3

THE RENEWAL EQUATION

3.1. Introduction

In this chapter we shall discuss the elementary properties of the solution of the linear functional equation

\[ u(t) = f(t) + \int_0^t u(t-s) \phi(s) \, ds. \]  \hspace{1cm} (3.1)

In some applications, the more general form of Eq. (3.1) occurs as

\[ u(t) = f(t) + \int_0^t u(t-s) \, dG(s). \]  \hspace{1cm} (3.2)

The discussion of this equation will be quite similar to that presented for Eq. (3.1), with the difference that it will require a knowledge of the Stieltjes integral, and, consequently, in certain points will require a more delicate handling.

In the majority of applications, \( G(t) \) either possesses a density function, i.e., \( dG(t) = \phi(t) \, dt \), or is a step function with a finite number of steps, in which case Eq. (3.2) reduces to a difference equation. Consequently, we feel that in this introductory discussion it is wise to by-pass any discussion of Eq. (3.2) in its most general form.

We shall begin with some existence and uniqueness theorems, and then we shall obtain various special properties of \( u \), such as monotonicity and bounded variation, under certain simple assumptions concerning \( \phi \) and \( f \).

Following these basic results, which are obtained by elementary arguments, we shall indicate how the Laplace transform may be used to derive an explicit analytic representation for the solution of Eq. (3.1). Guided by a heuristic application of the Laplace transform, we shall then discuss the connection between the solution of Eq. (3.1) and the solution of the simpler equation

\[ v(t) = 1 + \int_0^t v(t-s) \phi(s) \, ds. \]  \hspace{1cm} (3.3)

Finally, we shall close the chapter with a brief description of some questions in probability theory and mathematical physics which give rise to renewal equations.

3.2. Existence and Uniqueness Theorems

Let us now consider the functional equation
\[ u(t) = f(t) + \int_0^t u(t-s)\phi(s) \, ds, \] (3.4)

mentioned above. We shall state and prove two existence and uniqueness theorems which cover most of the equations of this type and which arise in applications. The integral will be taken to be the Lebesgue integral, although the unsophisticated reader may, if he desires, consider it to be a Riemann integral, without any harm done.

Although it might seem that the convolution calls for immediate application of the Laplace transform, the elementary analysis that we will present first is actually necessary to make our subsequent application of the transform technique rigorous.

The first result is

**Theorem 3.1.** If there exist a \( \epsilon_1 \geq 0 \) and a \( t_0 > 0 \) such that

(a) \[ |f(t)| \leq \epsilon_1 \text{ in } [0, t_0], \]

(b) \[ \int_0^{t_0} |\phi(s)| \, ds < \infty, \]

there is a unique bounded solution to Eq. (3.4) for \( 0 \leq t \leq t_0 \).

**Proof.** We employ the method of successive approximations. Define

\[ u_0 = f, \]

\[ u_{n+1}(t) = f(t) + \int_0^t u_n(t-s)\phi(s) \, ds. \] (3.6)

Let \([0, t_1]\) be an interval such that \( \int_0^{t_1} |\phi(s)| \, ds \leq b < 1 \), and assume first that \( t_1 \leq t_0 \). If \( t_1 \geq t_0 \), the Liouville-Neumann solution obtained by straightforward iteration is valid in the interval \([0, t_0]\). If \( t_1 < t_0 \), we proceed as follows. In \([0, t_1]\) we have, setting \( v_n = \sup_{[0,t]} |u_n| \), \( 0 \leq t \leq t_1 \), the inequality

\[ |u_{n+1}| \leq \epsilon_1 + v_n \int_0^{t_1} |\phi(s)| \, ds \leq \epsilon_1 + bv_n. \] (3.7)

Hence, if \( a_{n+1} = \epsilon_1 + ba_n, a_0 = \epsilon_1 \), we have \( |v_{n+1}| \leq a_{n+1} \) in \([0, t_1]\). It is easy to see that the sequence \( \{a_n\} \) is monotone increasing and uniformly bounded by \( a = \epsilon_1/(1-b) \), under our assumption that \( 0 < b < 1 \). It follows then that each integral in Eq. (3.6) exists and that the sequence \( \{u_n\} \) is uniformly bounded in \([0, t_1]\). To establish convergence, we write

\[ u_{n+1}(t) - u_n(t) = \int_0^t [u_n(t-s) - u_{n-1}(t-s)]\phi(s) \, ds, \] (3.8)

and obtain, for \( n \geq 1 \),

*There may exist unbounded solutions. The problem of determining the weakest conditions to impose upon \( f \) and \( \phi \) which will ensure uniqueness of solution does not seem to have been discussed.*
THE RENEWAL EQUATION

\[ w_{n+1} = \sup_{0 \leq s \leq t_1} |u_{n+1} - u_n| \leq \left( \sup_{0 \leq s \leq t_1} |u_n - u_{n-1}| \right) \int_0^t |\phi(s)| \, ds \leq b w_n. \]

(3.9)

This shows that the series \( \sum_{n=1}^{\infty} (u_{n+1} - u_n) \) is uniformly convergent in \([0, t_1]\), by comparison with the geometric series \( \sum_{n=1}^{\infty} b^n \). Hence the sequence \( \{u_n\} \) converges to a function \( u(t) \), which is bounded. Employing the Lebesgue convergence theorem, we may pass to the limit in Eq. (3.6) and establish the fact that \( u(t) \) is a solution to Eq. (3.4).

Having determined a solution over \([0, t_1]\), we now proceed to obtain a solution over the interval \([t_1, 2t_1]\) as follows. Define, for \( t_1 \leq t \leq 2t_1 \),

\[ u_0 = f, \]

\[ u_{n+1}(t) = f(t) + \int_0^{t-t_1} u_n(t-s) \phi(s) \, ds + \int_{t-t_1}^t u(t-s) \phi(s) \, ds, \]

where \( u(t) \) is the function obtained above for \( 0 \leq t \leq t_1 \). Hence,

\[ u_{n+1}(t) = f(t) + \int_0^{t-t_1} u_n(t-s) \phi(s) \, ds, \]

where

\[ f(t) = f(t) + \int_{t-t_1}^t u(t-s) \phi(s) \, ds. \]

(3.10)

(3.11)

This is a set of recurrence relations of precisely the same form as that given above. Consequently, the sequence converges for \( t_1 \leq t \leq 2t_1 \) to a solution of

\[ v(t) = f(t) + \int_0^{t-t_1} v(t-s) \phi(s) \, ds + \int_{t-t_1}^t u(t-s) \phi(s) \, ds. \]

(3.12)

If we now consider \( u(t) \) and \( v(t) \) as defining one function, \( u(t) \), over \([0, 2t_1]\), we have a solution over \([0, 2t_1]\). Continuing in this way, we obtain a solution over \([0, 3t_1]\), and so on, until we have covered the interval \([0, t_0]\).

To establish uniqueness of the solution over \([0, t_0]\), we first establish uniqueness over the interval \([0, t_1]\) and then over \([t_1, 2t_1]\), and so on. For example, let \( v(t) \) be another bounded solution of Eq. (3.4) in \([0, t_0]\). Then, in \([0, t_1]\),

\[ u_{n+1}(t) - v(t) = \int_0^t [u_n(t-s) - v(t-s)] \phi(s) \, ds, \]

whence

\[ |u_{n+1}(t) - v(t)| \leq \left( \sup_{0 \leq s \leq t} |u_n(t-s) - v(t-s)| \right) \int_0^t |\phi(s)| \, ds, \]

(3.13)

(3.14)
and consequently,
\[
\sup_{0 \leq t \leq t_1} |u_{n-1}(t) - v(t)| \leq b \sup_{0 \leq t \leq t_1} |u_n(t) - v(t)| \leq b^k \sup_{0 \leq t \leq t_1} |u_n(t) - v(t)|.
\]
(3.15)

From this it follows that \(\sup_{0 \leq t \leq t_1} |u(t) - v(t)| = 0\). Having established the identity of \(u\) and \(v\) in \([0, t_1]\), we proceed similarly in \([t_1, 2t_1]\), and so on.

Interchanging the assumptions in Sec. 3.5, above, our second result is

**Theorem 3.2.** If for some \(c_1 \geq 0\) and \(t_0 > 0\), we have

(a) \[|\phi(t)| \leq c_1 \text{ in } [0, t_0],\]
(b) \[\int_0^{t_1} |f(t_1)| \, dt_1 < \infty,\]

there is a unique solution to Eq. (3.1) which is absolutely integrable in \([0, t_0]\).

**Proof.** We employ the same successive approximants as used above, and consider now the interval \([0, t_1]\), where \(t_1\) is chosen so that \(c_1 t_1 \leq b < 1\). Then if

\[\int_0^{t_1} |u_k(t)| \, dt \leq a_k \quad \text{for } k = 0, 1, 2, \ldots, n,\]

we have

\[
|u_{n+1}| \leq |f| + \int_0^{t_1} |u_k(t - s)| \, |\phi(s)| \, ds
\leq |f| + c_1 \int_0^{t_1} |u_n(s)| \, ds
\leq |f| + c_1 a_n,
\]
(3.17)

which shows that \(u_{n+1}\) is absolutely integrable in \([0, t_1]\). Furthermore,

\[
\int_0^{t_1} |u_{n+1}| \, dt \leq \int_0^{t_1} |f| \, dt + c_1 a_n.
\]
(3.18)

Hence, if we set

\[a_{n+1} = \int_0^{t_1} |f| \, dt + c_1 a_n,\]
\[a_0 = \int_0^{t_1} |f| \, dt,\]

we have

\[
\int_0^{t_1} |u_{n+1}| \, dt \leq a_{n+1} \leq \int_0^{t_1} |f| \, dt / \left(1 - c_1 t_1\right).
\]

To establish convergence, we write
THE RENEWAL EQUATION

\[ u_1(t) - u_0(t) = \int_0^t u_0(t - s) \phi(s) \, ds, \]

\[ u_{n+1}(t) - u_n(t) = \int_0^t [u_n(t - s) - u_{n-1}(t - s)] \phi(s) \, ds, \quad n = 1, 2, \ldots. \]

(3.19)

We then have

\[ |u_1(t) - u_0(t)| \leq \epsilon_1 \int_0^t |f| \, dt, \]

\[ |u_2(t) - u_1(t)| \leq \epsilon_1^2 \int_0^t \left[ \int_0^t |f| \, ds \right] \, dt \leq \epsilon_1^2 \int_0^t (t - t_1) |f| \, dt, \]

and, inductively,

\[ |u_{n+1}(t) - u_n(t)| \leq \frac{\epsilon_1^{n+1}}{n!} \int_0^t (t - t_1)^n |f| \, dt \leq \frac{\epsilon_1^{n+1}}{n!} \int_0^t |f| \, dt. \]

(3.20)

Hence the series \( \sum_{n=0}^\infty (u_{n+1} - u_n) \) converges uniformly in \([0, t_1]\), and thus \( u_n(t) \) converges uniformly to \( u(t) \), a solution of Eq. (3.4). The extension to the full interval and the uniqueness proof proceed as given above.

3.3. Monotonicity and Bounded Variation

Having established the existence and uniqueness of the solutions, let us now discuss some further properties.

**Theorem 3.3.** Under the hypotheses of either Theorem 3.1 or Theorem 3.2, we have

(a) \( u(t) \) is continuous if \( f(t) \) is continuous;

(b) \( u(t) \) is monotone increasing if \( f(t) \) is monotone increasing, \( \phi(t) \geq 0 \), and \( f(0) \geq 0 \);

(c) \( u(t) \) is of bounded variation if \( f(t) \) is of bounded variation.

(3.22)

**Proof.** Let us consider the assertion in (a) first. Since in both cases we have proved that \( u_n(t) \) converges uniformly to \( u(t) \), it follows that \( u(t) \) is continuous whenever \( u_n(t) \) is. This will be so if \( f(t) \) is continuous.

Similarly, \( u(t) \) will be monotone increasing if \( u_n(t) \) is monotone increasing for each \( n \). Since \( \int_0^t f(t - s) \phi(s) \, ds \) is monotone increasing whenever \( f \) possesses this property and whenever \( \phi(t) \) and \( f(0) \) are both nonnegative, we see that (b) is valid.

To prove the third statement simply, for a particular case of importance, we use the fact that a function of bounded variation may be written as the difference of two monotone-increasing functions. It follows from the linearity of the equation, therefore,
that \( u \) will be of bounded variation if \( f \) is of bounded variation and, in addition, \( \phi \geq 0 \).

To derive the result in general, where \( \phi \) is not necessarily nonnegative, we may use an inductive argument, based on the successive approximants. Referring to Eq. (3.6), we have

\[
 u_{n+1}(t) = f(t) + \int_0^t u_n(t - \tau) \phi(\tau) \, d\tau. \tag{3.23}
\]

This yields, for any two quantities \( t_1 \) and \( t_2 \), with \( t_2 > t_1 > 0 \),

\[
 u_{n+1}(t_2) - u_{n+1}(t_1) = f(t_2) - f(t_1) + \int_{t_1}^{t_2} [u_n(t_2 - \tau) - u_n(t_1 - \tau)] \phi(\tau) \, d\tau 
 + \int_{t_1}^{t_2} u_n(t_2 - \tau) \phi(\tau) \, d\tau. \tag{3.24}
\]

Taking, for the sake of illustration, the case in which \( f(t) \) is bounded, we have \( |u_n| \leq c_1 \), whence

\[
 |u_{n+1}(t_2) - u_{n+1}(t_1)| \leq |f(t_2) - f(t_1)| + c_1 \int_{t_1}^{t_2} |\phi(\tau)| \, d\tau 
 + \int_{t_1}^{t_2} |u_n(t_2 - \tau) - u_n(t_1 - \tau)||\phi(\tau)| \, d\tau. \tag{3.25}
\]

Considering the points \( 0 < t_1 < t_2 < \ldots < t_N \) we see, by the addition of the inequalities corresponding to Eq. (3.25), that

\[
 \sum_{k=1}^{N-1} |u_{n+1}(t_{k+1}) - u_{n+1}(t_k)| \leq \sum_{k=1}^{N-1} |f(t_{k+1}) - f(t_k)| 
 + c_1 \int_0^{t_N} |\phi(\tau)| \, d\tau 
 + \int_0^{t_N} \left[ \sum_{k=1}^{N-1} |u_n(t_{k+1} - \tau)| - |u_n(t_k - \tau)| \right] |\phi(\tau)| \, d\tau, \tag{3.26}
\]

where \( u_n(t) \) is to be interpreted as 0 for \( t < 0 \).

From this it follows readily that the variation of \( u_{n+1} \) over \([0, \tau]\), defined by

\[
 V(u_{n+1}) = \sup \left[ \sum_{k=1}^{N-1} |u_{n+1}(t_{k+1}) - u_{n+1}(t_k)| \right], \tag{3.27}
\]

where the supremum is taken first over all partition of the interval \([0, \tau]\) into \( N \)
parts, and then over $N = 1, 2, \cdots$, satisfies the inequality

$$V(u_{n+1}) \leq V(f) + \epsilon_1 \int_0^\tau |\phi| \, ds + \int_0^\tau V(u_n) |\phi(s)| \, ds. \quad (3.28)$$

Choosing $\tau$ so that $\epsilon_1 \int_0^\tau |\phi| \, ds < 1$, we see that $V(u_n)$ is bounded as before. This establishes the bounded variation over $[0, \tau]$. To obtain it over the entire interval of existence, we use the technique employed for existence and uniqueness theorems in Sec. 3.2.

Let us note in passing that if $f$ and $\phi$ are both uniformly bounded for all $t$, the solution obtained by direct iteration,

$$u(t) = f(t) + \int_0^t f(t-s) \phi(s) \, ds + \cdots,$$

will converge for all $t \geq 0$.

If $f$ and $\phi$ are analytic in some neighborhood of $t = 0$, the power series expansion for $u$ in a neighborhood of $t = 0$ may be obtained easily by computing the derivatives $u^{(k)}(0)$ recurrently, using the equation.

### 3.4. The Laplace Transform Solution

In this section we shall obtain an explicit analytic expression for the Laplace transform of $u$—the solution of the renewal equation—under certain assumptions concerning $f$ and $\phi$.

We first require a preliminary result which permits us to estimate $u$ by means of bounds upon $f$ and $\phi$.

**Lemma 1.** If, for $t \geq 0$, and some $a$,

(a) $|f(t)| \leq \epsilon_1 e^{at}$,

(b) $\int_0^\infty e^{-at} |\phi(t)| \, dt = \epsilon_2 < 1$,

we have

$$|u| \leq \frac{\epsilon_1 e^{at}}{1 - \epsilon_2}. \quad (3.30)$$

**Proof.** We have

$$|u| \leq \epsilon_1 e^{at} + \int_0^t |u(t-s)| |\phi(s)| \, ds, \quad (3.31)$$

or

$$|ue^{-at}| \leq \epsilon_1 + \int_0^t |u(t-s)e^{-a(t-s)}| |e^{-as}| |\phi(s)| \, ds. \quad (3.32)$$
TIMELAG, RETARDED CONTROL, AND HEREDITARY PROCESSES

Setting \( v = |ue^{-at}| \), this may be written

\[
v \leq c_1 + \int_0^t e^{-as} |v(t - s)\phi(s)| \, ds.
\]  \hspace{1cm} (3.33)

Let

\[
M(t) = \max_{0 \leq t \leq t_1} |v(t)|.
\]  \hspace{1cm} (3.34)

Then

\[
M(t) \leq c_1 + M(t) \int_0^t e^{-as} |\phi(s)| \, ds = c_1 + c_e M(t),
\]  \hspace{1cm} (3.35)

whence \( M(t) \leq c_1/(1 - c_e) \), the desired inequality.

Using the above lemma and the convolution theorem, we readily establish

THEOREM 3.4. If for some \( a \) we have

(a) \[ |f(t)| \leq c_1 e^{at} \text{ for } t \geq 0, \]

(b) \[ \int_0^\infty e^{at} |\phi(t)| \, dt < 1, \] \hspace{1cm} (3.36)

the Laplace transform, \( L(u) \), of the solution of

\[
u(t) = f(t) + \int_0^t u(t - s)\phi(s) \, ds
\]  \hspace{1cm} (3.37)

is given by

\[
L(u) = \frac{L(f)}{1 - L(\phi)},
\]  \hspace{1cm} (3.38)

for Re(\( a \)) > 0.

At every point \( t \), where \( u \) is continuous and of bounded variation in some interval containing \( t \), we have

\[
u = \int_{(b)} \frac{L(f)}{1 - L(\phi)} e^{at} \, ds,
\]  \hspace{1cm} (3.39)

where \( b > a \).

In the previous section we gave simple conditions upon \( f \) and \( \phi \) which permitted us to conclude that \( u \) was continuous and of bounded variation.

3.5. A Convolution Theorem

The purpose of this section is to derive a formula connecting the solutions of
THE RENEWAL EQUATION

\[ u(t) = 1 + \int_0^t u(t - s) \phi(s) \, ds \quad (3.40) \]

and

\[ v(t) = f(t) + \int_0^t v(t - s) \phi(s) \, ds. \quad (3.41) \]

This result is of service occasionally in connection with the study of the asymptotic behavior of the solutions. In order to derive the formula, let us use the Laplace transform in a heuristic fashion. We have

\[ L(u) = \frac{1}{s(1 - L(\phi))}, \]

\[ L(v) = \frac{L(f)}{1 - L(\phi)}, \quad (3.42) \]

whence

\[ \frac{L(v)}{L(u)} = sL(f) = \int_0^\infty e^{-st} f(t) \, dt. \quad (3.43) \]

From the convolution theorem it follows that

\[ v = f(0)u + \int_0^t u(t - s)f'(s) \, ds. \quad (3.44) \]

It is this formula which we wish to establish rigorously, under appropriate assumptions concerning \( f \) and \( \phi \).

**Theorem 3.5.** If

(a) \( f'(t) \) exists for \( 0 \leq t \leq t_0 \), \( \int_0^{t_0} |f'(t)| \, dt < \infty \), and

(b) \( \int_0^{t_0} |\phi(s)| |ds| < \infty \), \quad (3.45)

then a solution to Eq. (3.41) is given by Eq. (3.44) for \( 0 \leq t \leq t_0 \).

**Proof.** We have, using Eq. (3.44),

\[ \int_0^t v(t - s) \phi(s) \, ds = f(0) \int_0^t u(t - s) \phi(s) \, ds \]

\[ + \int_0^t \left[ \int_0^{t-s} u(t - s - t_1) f'(t_1) \, dt_1 \right] \phi(s) \, ds. \quad (3.46) \]

Interchanging the orders of integration, a legitimate operation because of the absolute convergence of the double integral, we obtain for the second term on the right in Eq. (3.46),
\[ \int_0^t \left[ \int_0^{t-s} u(t-s,i) \phi(s) \, ds \right] f'(s_i) \, ds_i = \int_0^t \left[ \int_0^{t-s} u(t-s,i) \, ds \right] f'(s_i) \, ds_i, \]

(3.47)

using Eq. (3.40). Combining the results, we obtain

\[ \int_0^t \: (i - s) \phi(s) \, ds = f(0) \int_0^t u(t-s) \phi(s) \, ds + \int_0^t [u(t-s) - 1] f'(s) \, ds \]

\[ = f(0) u(t) - f(0) + \int_0^t u(t-s) f'(s) \, ds - \int_0^t f'(s) \, ds \]

\[ = f(0) u(t) - f(t) + \int_0^t u(t-s) f'(s) \, ds \]

\[ = v(t) - f(t), \]

(3.48)

which shows that \( v \) satisfies Eq. (3.41).

An integration by parts in Eq. (3.44) yields

\[ v = u(0) f(t) + \int_0^t f(t-s) u'(s) \, ds, \]

(3.49)

provided that \( u'(s) \) exists. Since this formula has a meaning, even if \( f(t) \) is not differentiable, it is reasonable to suspect that either Eq. (3.49) yields the solution of Eq. (3.41) under suitable conditions upon \( u(t) \). We shall not discuss this question in further detail here, since it is more properly a part of the theory of renewal equations when Stieltjes, rather than Lebesgue, integrals are employed. An expression for \( v \) of wider validity would be

\[ v = u(0) f(t) + \int_0^t f(t-s) \, du(s). \]

3.6. Applications—I

In this section and in the five following sections, we shall consider some simple examples of industrial replacement and population-growth problems whose mathematical formulation leads to renewal equations.

Our first example is the following: Suppose that we have a shipment of light bulbs all possessing a common life-length distribution, \( F(x) = \int_0^x \, dF(t) \). These bulbs are to be used in one lamp, with a new bulb replacing an old one whenever it burns out. The problem is to determine the expected number of bulbs required to keep the lamp in service for a time interval of length \( t \).

Let us show that this expected number, which we denote by \( u(t) \), satisfies the functional equation

\[ u(t) = 1 + \int_0^t u(t-s) \, dF(s). \]

(3.50)
THE RENEWAL EQUATION

The probability that the original bulb remains burning for at least a time \( t \) is

\[ u(t) = 1 - \int_0^t dF(x) + \int_0^t [1 + u(t - x)] dF(x) \]

\[ = 1 + \int_0^t u(t - x) dF(x). \]  
(3.51)

3.7. Applications—II

Let us now consider an example of historical interest [6]. Suppose that we have a

Let us further assume that as soon as one individual dies, he is replaced by an individual

age zero, who has the same survival probability. The effect of this is to keep the
total number in the group constant.

With this policy in effect, let \( Nf(t) \) be the expected number dying between \( s \) and

\( s + ds \), which is to say the number of replacements required between \( s \) and \( s + ds \). It is

required that the relation between \( p(t) \) and \( f(t) \) be found.

The expected number of charter members surviving to time \( t \) is \( Np(t) \), while the

expected number of survivors at time \( t \) of new members is

\[ N \int_0^t f(s)p(t - s) \, ds. \]  
(3.53)

Since the membership is constant, this yields the equation

\[ Np(t) + N \int_0^t f(s)p(t - s) \, ds = N, \]  
(3.54)

or

\[ p(t) + \int_0^t f(s)p(t - s) \, ds = 1. \]  
(3.55)

By differentiating, and using the fact that \( p(0) = 1 \), the resultant equation is

\[ f(t) = -p'(t) - \int_0^t f(s)p'(t - s) \, ds. \]  
(3.56)

The problem is equivalent to one in the industrial replacement of an equipment
comprising $N$ original units installed at time 0 and maintained at constant size of $N$
by means of the replacement of faulty units by new ones.

3.8. Applications—Ill

A prolific source of renewal-type equations, with their application to neutron cascades
and cosmic-ray showers, population growth, biological mutation, and many similar
classes of generation phenomena, is the general theory of branching processes. For a
discussion of these problems and further references, let us cite [1], [2], [4], and [5].

Let us discuss here a simple representative example: A particle existing at time 0 is
assumed to have a life-length whose cumulative probability distribution is given by a
function $G(t)$. At the end of its life it is transformed into $n$ similar particles with
probability $q_n$, $n \geq 0$. These new particles are taken to have the same life-length distri-
bution and transformation probabilities as the original one, and the process now continues.

Under the hypothesis that the life-length distribution and the transformation probabilities
for each particle are independent of its time of birth and the number of other particles
existing at the time, the problem is to determine the distribution of the number of
particles existing at time $t$, which we call $Z(t)$.

Setting $b(t) = \sum_{n=0}^{\infty} q_n G_n$, where $q_n$ is as given above, which is the probability
that a particle is transformed into $n$ similar particles, our starting point is the fact that the
generating function

$$F(t, \lambda) = E e^{\lambda t}$$

satisfies the nonlinear integral equation

$$F(t, \lambda) = \int_0^t \left[ b[F(s, \lambda - \gamma)] dG(\gamma) + \lambda[1 - G(t)] \right].$$

(3.58)

The above expression represents the expected value of $s^2$ at time $t$ for a fixed $s$, where
$Z = Z(t)$ is the random variable equal to the number of particles existing at time $t$, 
mentioned above.

Let us denote the expected value of $Z(t)$ by $v(t)$. Then differentiation of Eq. (3.58)
with respect to $\lambda$, and setting $\lambda$ equal to 1, yields for $v$ the equation

$$v(t) = b'(1) \int_0^t v(t - \gamma) dG(\gamma) + 1 - G(t),$$

(3.59)

a renewal equation. For further details, see [2] and [5].

3.9. Applications—Iv

As a simple example of a physical problem leading to an equation similar to the
renewal equation, let us consider the equation

$$m \frac{d^2 u}{dt^2} + \sigma^2 u(t) + \int_0^t K(t - \gamma) \frac{du(\gamma)}{ds} ds = q(t),$$

(3.60)
which arises from the equation of motion of a system with one degree of freedom possessing hereditary damping. Here \( u \) is the generalized coordinate, \( m \) an inertia coefficient, \( a^2 \) an elastic constant, \( K \) the heredity function, and \( q(t) \) the generalized external force.

The stress-strain relationship underlying Eq. (3.60) is of the form

\[
\sigma = E\varepsilon(t) + \int_0^t \phi(t - \tau) \frac{d\varepsilon(\tau)}{ds} d\tau.
\]  

(3.61)

If

\[
K(t) = \sum_{i=1}^n a_ie^{-a_i t},
\]

(3.62)

the equation reduces to a linear differential equation upon repeated differentiation.

Integrodifferential equations of the type appearing in Eq. (3.60) were first studied by V. Volterra [9]. The most recent discussion, together with other references, will be found in E. Volterra [8].

### 3.10. Applications—V

If we consider a single species, of size \( N(t) \) at time \( t \), the simplest hypothesis is that the change in the size of the species is proportional to the size of the species. This leads to the differential equation

\[
\frac{dN}{dt} = a_i N(t).
\]

(3.63)

This is not a bad approximation for small \( N \). For large \( N \), however, it neglects the feeding problem.* To compensate for the fact that a large colony cannot feed itself as well as a small one, it is customary to include a term \(-a_2 N^2(t)\), where \( a_2 \) is a small constant, obtaining

\[
\frac{dN}{dt} = a_i N(t) - a_2 N^2(t).
\]

(3.64)

This equation, commonly called the "logistics" equation, describes various types of population growth in an amazingly accurate way, with proper fitting of the coefficients.

In considering the growth of two coexisting species, the second of which lives on the first, similar arguments yield the equations

\[
\frac{dN_1}{dt} = a_{11} N_1(t) - a_{12} N_1(t) N_2(t),
\]

\[
\frac{dN_2}{dt} = -a_{21} N_2(t) + a_{22} N_1(t) N_2(t),
\]

(3.65)

where \( a_{ij} > 0 \).

*This is not to imply that we take a stand one way or the other on the Malthusian dictum.
The models on which the above equations are based neglect any stochastic effects and assume purely deterministic behavior. The relationship between these two opposing schools of thought was first investigated by Feller \[4\], where a detailed discussion may be found.

The stochastic aspects can be of considerable importance in the discussion of animal populations, such as rats, mice, blackbirds, and so on, and possibly of lesser importance in the discussion of biologic populations of paramecia, etc., because of the greater size. However, size alone is not a determining factor. The stability of equilibrium states must also be investigated.

We are, at the moment, interested not in a debate between stochastic versus deterministic models, but in the validity of the assumption of simultaneity of effects inherent in Eq. (3.65). Let us consider a more realistic model from this point of view. Let us assume that in the devouring species the distribution of ages remains constant, and denote by \(\psi(y)\) the ratio of the number between ages \(y\) and \(y + dy\) to the total number.

The number of the second species having at time \(t\) an age of at least \(z\) is then

\[
N_2(t) \int_{-\infty}^{z} \psi(y) \, dy = N_2(t)f(z).
\]  

(3.66)

Among the \(N_2(t)\) individuals existing at time \(t\), there are therefore \(N_2(t)f(t - s)\) which were in existence at time \(s\) before. On the basis of the approximations which lead to Eq. (3.65), the number of individuals of the first kind devoured by members of the second species which existed at both \(s\) and \(t\) is, in the interval \([s, s + ds]\),

\[
a_1 f(t - s)N_1(s)N_2(t) \, ds.
\]  

(3.67)

Let us assume that this created a growth of

\[
\psi(t - s) \, ds \cdot a_1 f(t - s)N_1(s)N_2(t) \, ds
\]  

(3.68)

individuals of the second kind in the interval \([t, t + dt]\), where \(\psi\) is a nonnegative function.

Assuming that these growths are independent, we have as a total effect in \([t, t + dt]\),

\[
N_2(t) \int_{-\infty}^{t} a_1 \psi(t - s)f(t - s)N_1(s) \, ds.
\]  

(3.69)

In place of Eq. (3.65) we have then, as a further approximation,

\[
\frac{dN_1(t)}{dt} = a_1 N_1(t) - a_{12} N_1(t) N_2(t),
\]  

\[
\frac{dN_2(t)}{dt} = N_2(t) \left[ -a_{21} + a_1 \int_{-\infty}^{t} \psi(t - s)f(t - s)N_1(s) \, ds \right].
\]  

(3.70)

Since these equations are nonlinear, any explicit solution is not to be expected. How-
ever, an investigation of the stability of equilibrium solutions,

\[ N_2 = \frac{a_{12}}{a_{12}}, \]

\[ N_1 = a_{12} \int_{b}^{\infty} \frac{\psi(u)f(u)}{d_1} du, \]

(3.71)

will yield linear equations of a renewal type. For further discussion, see [9].

REFERENCES


CHAPTER 4

ASYMPTOTIC BEHAVIOR OF THE SOLUTION OF THE RENEWAL EQUATION

4.1. Introduction

Since, in applications, it is usually the asymptotic behavior of the solution of the renewal equation which is most important, in this chapter we shall present a number of the more common techniques which are used to determine the asymptotic character of the solution of the equation

\[ u(t) = f(t) + \int_0^t u(t-s)\phi(s) \, ds. \]  \hspace{1cm} (4.1)

First we shall cite an important theorem due to Paley and Wiener, referring the reader to their book for further details. Turning from this general result, we shall consider the case of most importance in applications, i.e., the case in which \( \phi(t) \geq 0 \). Here a Tauberian theorem of Hardy and Littlewood occupies a central position. Following this, we shall present an elementary attack on the two cases in which \( \phi(t) \geq 0 \) and

\[ \int_0^\infty \phi(s) \, ds = 1 \quad \text{or} \quad \int_0^\infty \phi(s) \, ds < 1. \]  \hspace{1cm} (4.2)

For the case in which neither condition in Eq. (4.2) is satisfied, we shall turn to an important theorem of Ikehara.

Finally, we shall show how to use the Laplace transform to obtain not only the leading term in the asymptotic expansion, but also the lower-order terms, under suitable assumptions concerning \( \phi \) and \( f \). Other techniques, particularly one of the Tauberian type due to Haar, will be found in Feller [1], and some, which are more computational in nature but nonetheless important, will be found in Lotka (Ref. [6] of Chapter 3), where a number of interesting graphs appear.

If \( f(t) \) is a function possessing an absolutely integrable derivative over \([0, \infty]\), the convolution theorem (Theorem 3.5) may be used to obtain the asymptotic behavior of the solution of Eq. (4.1), in terms of that of the solution of

\[ v(t) = 1 + \int_0^t v(t-s)\phi(s) \, ds, \]  \hspace{1cm} (4.3)

about which much can be obtained with ease. To simplify our presentation, we shall consider only Eq. (4.3).

The problem of asymptotic behavior of more complicated functional equations may
be treated by using some extensions of the fundamental Tauberian theorem of Wiener (see Wright [6]), but we shall not discuss this here.

4.2. The Result of Paley and Wiener

The following result is a consequence of the general Tauberian theorem of Wiener (see [2] and [5]).

**Theorem 4.1.** If

\[ u(t) = f(t) + \int_{0}^{t} u(t-s)\phi(s) \, ds, \]  

(4.4)

where

(a) \[ \lim_{t \to \infty} f(t) = 1, \]

(b) \[ f(t) \text{ is bounded over every finite interval}, \]

(c) \[ \int_{0}^{\infty} |\phi(t)| \, ds < \infty, \]

(d) \[ \int_{0}^{\infty} \phi(t)e^{ws} \, ds \neq 1 \quad \text{for} \quad \Re(w) \geq 0, \]  

(4.5)

then, as \( t \to \infty \), we have

\[ u(t) \sim \frac{1}{1 - \int_{0}^{\infty} \phi(s) \, ds}. \]  

(4.6)

It is sometimes possible, by a judicious substitution, to transform a problem not satisfying (4.5) into one satisfying the prescribed conditions.

4.3. The Tauberian Theorem of Hardy and Littlewood

For the important case in which \( \phi \geq 0 \), the following result of Hardy and Littlewood is extremely useful in furnishing a guide to the asymptotic behavior of the solution, and occasionally in furnishing the precise leading term.

**Theorem 4.2.** If \( u(t) \geq 0 \) and

\[ \int_{0}^{\infty} u(t)e^{-st} \, dt \sim \frac{c_{1}}{sk}, \quad c_{1} > 0, \quad k \geq 0, \]  

(4.7)

as \( s \to +0 \), then

\[ \int_{0}^{T} u(t) \, dt \sim \frac{c_{1}T^{k}}{1(k+1)} \]  

(4.8)

as \( T \to \infty \).

Let us apply this result to the determination of the asymptotic character of the
solution of

\[ u(t) = 1 + \int_0^t u(t - s) \phi(s) \, ds, \quad (4.9) \]

assuming that

\[ \int_0^\infty \phi \, dt = 1, \quad m_1 = \int_0^\infty t \phi \, dt < \infty. \quad (4.10) \]

The Laplace transform of \( u \), \( L(u) \), is given by

\[ L(u) = \frac{1}{s[1 - L(\phi)]}. \quad (4.11) \]

By virtue of Eq. (4.10), we see that as \( s \to +0 \) we have

\[ L(u) \propto \frac{1}{m_1 s^2}. \quad (4.12) \]

By using Theorem 4.2, this yields

\[ \int_0^T u \, dt \propto \frac{T^2}{2 m_1}. \quad (4.13) \]

In order to obtain the asymptotic behavior of \( u \) itself, we differentiate Eq. (4.9), obtaining

\[ u'(t) = \phi(t) + \int_0^t u'(t - s) \phi(s) \, ds; \quad (4.14) \]

and by applying Theorem 4.2 again, we obtain

\[ L(u') = \frac{L(\phi)}{1 - L(\phi)} \propto \frac{1}{m_1 s}. \quad (4.15) \]

Hence,

\[ u(T) - u(0) = \int_0^T u' \, dt \propto \frac{T}{m_1}. \quad (4.16) \]

4.4. The Elementary Approach

Consider Eq. (4.9), in which we assume only

\[ \phi(t) \geq 0, \quad \int_0^\infty \phi(s) \, ds < 1. \quad (4.17) \]

Since \( u(t) \) is monotone increasing and bounded, as a consequence of Eq. (4.17), we see
that \( u(\infty) = \lim_{t \to \infty} u(t) \) exists. Hence,

\[
u(\infty) = 1 + \lim_{t \to \infty} \int_0^t u(t-s) \phi(s) \, ds = 1 + u(\infty) \int_0^\infty \phi(s) \, ds, \tag{4.18}\]

whence

\[
u(\infty) = \frac{1}{1 - \int_0^\infty \phi(s) \, ds}. \tag{4.19}\]

As another example of the use of elementary methods, let us demonstrate by an alternative technique the result obtained by using the Hardy and Littlewood theorem, namely,

**Theorem 4.3.** If

(a) \( \phi(t) \geq 0 \),

(b) \( \int_0^\infty \phi(s) \, ds = 1 \), \quad \mu_1 = \int_0^\infty s \phi(s) \, ds < \infty, \tag{4.20}\)

the solution of

\[
u(t) = 1 + \int_0^t u(t-s) \phi(s) \, ds \tag{4.21}\]

that

\[
u \sim \frac{1}{\mu_1}, \tag{4.22}\]

**Proof.** In Eq. (4.21), set \( \dot{h} = 1, m_1 \), and \( \dot{\nu} = b(t + 1) \). Then,

\[
\begin{align*}
\dot{\nu} &= b(t + 1) - 1 \int_0^t (t-s) \phi(s) \, ds - b \int_0^\infty (t-s) \phi(s) \, ds \\
&= b(t + 1) - 1 \int_0^t \phi(s) \, ds - b \int_0^\infty \phi(s) \, ds - \int_0^t s \phi(s) \, ds,
\end{align*} \tag{4.23}\]

or

\[
\begin{align*}
\int_0^\infty \phi(s) \, ds &= \int_0^\infty \phi(s) \, ds + \int_0^\infty \phi(s) \, ds \\
&= \int_0^\infty \phi(s) \, ds + \int_0^\infty \phi(s) \, ds,
\end{align*} \tag{4.24}\]
we see that Eq. (4.25) may be written in the form

\[ \nu(t) = f(t) + \int_0^t \nu(t - s) \phi(s) \, ds, \]

where \( f(t) \to 0 \) as \( t \to \infty \).

We now wish to show that \( \nu(t) = O(t) \) as \( t \to \infty \). To do this, let us prove that

\[ |\nu| \leq a + et \quad \text{as} \quad t \to \infty, \]

where \( e \) is any preassigned positive constant and \( a = a(e) \).

Consider the solution to Eq. (4.25), as obtained by the method of successive approximations

\[ \nu_0 = f, \]

\[ \nu_{n+1} = f + \int_0^t \nu_n(t - s) \phi(s) \, ds. \]  

(4.26)

Let us now choose \( t_0 \) with the condition that \( |f| \leq e \) for \( t \geq t_0 \) and \( t_0 \geq 1 \). Let \( a_0 = \max |f| \quad \text{in} \quad [0, t_0] \) if this maximum is nonzero; otherwise, it is equal to 1. Then clearly, for all \( t \geq 0 \), we have \( |\nu_0| \leq a_0 + et_0 \). Using this bound in \( \nu_1 \), as given by Eq. (4.26), we obtain in \([0, t_0]\)

\[ |\nu_1| \leq \int_0^t [a_0 + e(t - s)] \phi(s) \, ds \]

\[ \leq a_0 + a_0 \int_0^t \phi(s) \, ds + et_0 \int_0^t \phi(s) \, ds - e \int_0^t s \phi(s) \, ds \]

\[ \leq a_0 + a_0 \int_0^t \phi(s) \, ds + et_0, \]  

(4.27)

since \( \int_0^a \phi(s) \, ds = 1 \) and \( \phi \geq 0 \). For \( t \geq t_0 \), we obtain

\[ |\nu_1| \leq e + \int_0^t [a_0 + e(t - s)] \phi(s) \, ds \]

\[ \leq e + a_0 \int_0^t \phi(s) \, ds + et \int_0^t \phi(s) \, ds \]

\[ \leq e + a_0 + et. \]  

(4.28)

Let us define \( a_1 = a_0 + a_0 \int_0^t \phi(s) \, ds \). If \( e \) is small enough and \( t_0 \) is large enough, we have \( a_1 \geq a_0 + e \). We see then that \( |\nu_1| \leq a_1 + et_0 \) for \( t \geq 0 \).

All the requirements for an inductive proof are now at hand. If we have

\[ |\nu_n| \leq a_n + et \quad \text{for} \quad t \geq 0, \]

the same argument as above yields \( |\nu_{n+1}| \leq a_{n+1} + et \), where

\[ a_{n+1} = a_0 + a_n \int_0^t \phi(s) \, ds. \]  

(4.29)

If \( \phi \) is not identically zero for \( t \geq t_0 \), the conditions \( \phi \geq 0 \) and \( \int_0^a \phi \, ds = 1 \)
yield \( \int_t^a \phi(s) \, ds < 1 \), and thence

\[
a_n < a_\infty = \frac{a_0}{1 - \int_0^{t_0} \phi(s) \, ds}.
\]

(4.30)

If \( \phi \) is identically zero for \( t \geq t_0 \), for some \( t_0 \), there is no difficulty in obtaining

the asymptotic behavior of \( u \) by other means, since

\[
1 - \int_0^\infty \phi e^{-st} \, ds = 1 - \int_0^{t_0} \phi e^{-st} \, dt
\]

(4.31)

is now an entire function. Hence we may with impunity assume that \( \int_0^{t_0} \phi(s) \, ds < 1 \)

for any fixed finite \( t_0 \).

Since \( |s_n| < a_n + \varepsilon t < a_\infty + \varepsilon t \) for all \( n \) and for \( t \geq 0 \), it follows that the solution enjoys the same property, which means that \( v(t) = 0(t) \) as \( t \to \infty \), since \( \varepsilon \) is arbitrary.

This completes the proof of Theorem 4.3.

4.5. The Tauberian Theorem of Ikehara

The case in which \( \int_0^\infty \phi \, ds = \infty \) requires some new techniques. Let us assume that

\( \int_0^\infty \phi e^{-st} \, ds \)

has a finite abscissa of convergence, and hence that there is a positive number \( a \) such that \( \int_0^\infty \phi(t)e^{-st} \, dt = 1 \). Assume also that \( \int_0^\infty \phi(t)e^{-st} \, ds \) is convergent.

Then, considering the equation

\[
u(t) = \int_0^t u(t-s) \phi(s) \, ds,
\]

(4.32)

we see that the change of variable \( u(t) = e^{st}v(t) \) converts Eq. (4.32) into

\[
v(t) = e^{-st} + \int_0^t v(t-s)e^{-st}\phi(s) \, ds.
\]

(4.33)

Applying the Laplace transform, we have

\[
L(v) = \frac{L(e^{-st})}{1 - L(e^{-st}\phi)} \underset{s \to 0}{\sim} \frac{1}{s\int_0^{\infty} te^{-st}\phi \, ds}.
\]

(4.34)

as \( s \to 0 \). Applying the Tauberian theorem of Hardy and Littlewood (Theorem 4.2),

we obtain

\[
\int_0^T v \, dt \underset{s \to 0}{\sim} \frac{T}{s\int_0^{\infty} te^{-st}\phi \, ds}.
\]

(4.35)

To obtain information concerning \( v \) itself, rather than its average, we require the following deeper result of Ikehara (see \([5]\)):

**Theorem 4.4.** If \( u(t) \) is a nonnegative, nondecreasing function in \( 0 \leq t < \infty \),
such that the integral

$$f(s) = \int_0^{\infty} e^{st} \phi(t) \, dt, \quad s = \sigma + it,$$

converges for $\sigma > 1$, and if, for some constant $A$ and for some function $g(t)$,

$$\lim_{\sigma \to \infty} f(s) - \frac{A}{s - 1} = g(t)$$

(4.36)

uniformly in every interval $[-a \leq \tau \leq a]$, then

$$\lim_{t \to \infty} u(t)e^{-t} = A.$$  

(4.37)

Using this theorem we may prove

**Theorem 4.5.** If, for some $a > 0$ and some $b > 0$,

$$u = 1 + \int_0^t u(t - s)\phi(s) \, ds,$$  

(4.38)

where

(a) $\phi(t) \geq 0$, for $t \geq 0$,

(b) $\int_0^{\infty} e^{-at}\phi(t) \, dt = 1$, for $a > 0$,

(c) $\int_0^{\infty} \int_0^{\infty} e^{-at}\phi(t) \, dt < \infty$, for $b > 0$,

(4.39)

then

$$ue^{-at} \leq \frac{1}{\int_0^{\infty} te^{-at}\phi \, dt}$$

(4.40)

at $t \to \infty$.

### 4.6. Use of the Contour Integral Representation

The principal advantage of the Tauberian technique resides in the fact that one requires only a knowledge of the behavior of $\int_0^{\infty} e^{-at}\phi(t) \, dt$ for $Re(s) \geq a$. On the other hand, one obtains only a weak result of the form $u(t) \sim ct$ or $u(t) \sim ce^{at}$.

If we assume that more is known concerning the analytic behavior of $\int_0^{\infty} e^{-at}\phi(t) \, dt$, we can obtain not only the principal term in the asymptotic expression, but also an error term by using the contour integral representation for the solution of Eq. (4.9),

$$u = \int_{(b)} \frac{e^{st}}{s[1 - L(\phi)]} \, ds, \quad b > a.$$  

(4.41)

(See Theorem 2.3.)
If, for example, \( \phi \geq 0 \), but not identically zero, then

\[
1 - \int_{0}^{\infty} e^{st}\phi(t) \, dt = 0
\]

has only one real root, which we have taken to be \( a \). Shifting the contour to the line

\[ c + i\epsilon, \quad -\infty < \epsilon < \infty, \quad c < a, \]

we obtain a residue of \( e^{at}/ma \), where \( m = \int_{0}^{\infty} e^{-at}\phi(t) \, dt \), plus an error term, \( O(e^{\epsilon t}) \), provided that we can find some way of estimating

\[
\int_{(c)} \frac{e^{st}}{[1 - L(\phi)]} \, ds = \int_{(c)} \frac{e^{st}}{s} \, ds + \int_{(c)} \frac{e^{st}L(\phi)}{[1 - L(\phi)]} \, ds. \tag{4.42}
\]

The first term is merely a constant. If \( L(\phi)/s \) is absolutely integrable, and if

\[ |1 - L(\phi)| \geq d > 0 \]

on the contour, we see that the second term is \( O(e^{\epsilon t}) \).

There are now many types of conditions upon \( \phi \) which will permit us to shift the contour of integration. The simplest are those that assume that \( \phi(t) \) has a derivative, so that integration by parts is permissible in \( L(\phi) \). For a detailed discussion, we refer to Feller [1] and Täcklind [3].

REFERENCES


For some additional results concerning the asymptotic behavior of solutions, see also R. Bellman and T. E. Harris in the References of Chapter 3.
CHAPTER 5
SYSTEMS OF RENEWAL EQUATIONS

5.1. Introduction

Although the renewal equation

\[ u(t) = f(t) + \int_0^t u(t - s) \, dG(s) \]  \hspace{1cm} (5.1)

is by now a quite familiar sight in papers on probability theory and its applications, very little has been done in connection with systems of renewal equations of the form

\[ u_i(t) = f_i(t) + \int_0^t \sum_{j=1}^N u_j(t - s) \, dG_{ij}(s), \quad i = 1, 2, \cdots, N. \]  \hspace{1cm} (5.2)

Despite the fact that many of the theorems concerning existence and uniqueness extend with great ease to the N-dimensional case, as far as more specific properties such as asymptotic behavior and so on are concerned, the transition from the one-dimensional case to the N-dimensional case is not routine. New methods are definitely required.

We shall limit ourselves here to a treatment of one important question which arises in N-dimensional branching processes. This will provide us with an opportunity to display a number of useful techniques which may be utilized in further research. The fundamental tool will be a variational approach due to Bohnenblust, which will be applied first to prove a classical fundamental theorem of Perron, and then to derive an extension of this result, which is required for the renewal equation.

5.2. The Perron Theorem for Positive Matrices

We shall call a matrix, \( A \), positive if all of its elements are positive, and we shall write \( A > 0 \). The notation \( A > B \) will be used to denote \( A - B > 0 \). Similarly, a vector \( x \) will be called positive if all of its components are positive, and we shall write \( x > 0 \). The notations \( A \geq 0, x \geq 0 \) will denote nonnegativity of the respective elements and components.

The classic theorem of Perron, which has come to occupy a pivotal position in the modern theory of probability and its applications, is

**Theorem 5.1.** If \( A \) is a positive matrix, there is a unique characteristic root, \( \lambda_\text{max}(A) \), of \( A \) which has greatest absolute value. This root is positive and simple, and its associated characteristic vector may be taken to be positive.

Although there are many proofs of this result, of quite diverse origin, one of the most important proofs, from the point of view of understanding and characterizing

51
\( \lambda_M(A) \), is that which is based on the following result of Bohnenblust:

**Theorem 5.2.** Let \( A \) be a positive matrix and let \( \lambda_M(A) \) be the characteristic root of \( A \) of largest absolute value. Let \( S(\lambda) \) be the set of nonnegative \( \lambda \) for which there exist nonnegative vectors \( x \) such that \( Ax \geq \lambda x \). Let \( T(\lambda) \) be the set of positive \( \lambda \) for which there exist positive vectors \( x \) such that \( Ax \leq \lambda x \). Then

\[
\lambda_M = \max_{\lambda \in S(\lambda)}, \quad \lambda \leq T(\lambda).
\]

**Proof of Theorem 5.2.** Let us normalize the vectors we are considering, so that \( \| x \| = \sum_{i=1}^{n} x_i = 1 \), and set \( \| A \| = \sum_{i,j} a_{ij} \). This normalization has the advantage of automatically excluding the trivial vector. If \( \lambda x \leq A x \), we obtain

\[
\lambda \| x \| \leq \| A x \| \leq \| A \| \| x \|, \quad \text{or} \quad 0 \leq \lambda \leq \| A \|,
\]

which means that \( S(\lambda) \) is a bounded set. It is easy to see that it contains points in addition to \( \lambda = 0 \). Let \( \lambda_0 = \sup \lambda \) for \( \lambda \in S(\lambda) \), let \( \{ \lambda_1 \} \) be a sequence of \( \lambda \)'s in \( S(\lambda) \) converging to \( \lambda_0 \), and let \( \{ x^{(1)} \} \) be an associated sequence of vectors such that \( \lambda_0 x^{(1)} \leq A x^{(1)} \). Since \( \| x^{(1)} \| = 1 \), we may choose a subsequence of the \( x^{(1)} \), say \( \{ x^{(2)} \} \), which converges to \( x^{(0)} \), a nonnegative, nontrivial vector. Since \( \lambda_0 x^{(0)} \leq A x^{(0)} \), it follows that \( \lambda_0 \in S(\lambda) \), which means that the supremum is actually a maximum.

Let us now demonstrate that the inequality is actually an equality with \( \lambda_0 x^{(0)} = A x^{(0)} \). The proof is by contradiction. Let us suppose, without loss of generality, that

\[
\sum_{j=1}^{N} a_{ij} x_i^{(0)} - \lambda_0 x_i^{(0)} = d_i > 0,
\]

and

\[
\sum_{j=1}^{N} a_{kj} x_j^{(0)} - \lambda_0 x_k^{(0)} \geq 0, \quad k = 2, \cdots, N.
\]

If we consider the vector

\[
y = x^{(0)} + \begin{bmatrix} d_1 \\ 2\lambda_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},
\]

it is clear that \( Ay > \lambda_0 y \). This, however, contradicts the maximum property of \( \lambda_0 \). Hence, there must be equality. This means that \( \lambda_0 \) is a characteristic root of \( A \), since \( x^{(0)} \) is nontrivial; furthermore, it shows that \( x^{(0)} \) is actually positive.

To show that \( \lambda_0 = \lambda_M \), assume that there exists some characteristic root \( \lambda \) of \( A \) for which \( |\lambda| \geq \lambda_0 \). Let \( z \) be an associated characteristic vector. Then \( Az = \lambda z \) leads to \( |\lambda||z| \leq |A||z| \), where \( |z| \) denotes the vector whose components are the absolute values of the components of \( z \). It follows from this last inequality that \( |\lambda| \leq \lambda_0 \).
If $|\lambda| = \lambda_0$, it follows, as above, that the inequality $|A| |z| \leq A |z|$ is actually an equality. This implies that $|Az| = A |z|$, whence $z = c_i w$, where $w > 0$ and $c_i$ is a complex number. Consequently, $Az = \lambda z$ is equivalent to $Aw = \lambda w$, which means that $\lambda$ is real and positive, and thus equal to $\lambda_0$.

To verify that the second method of defining $\lambda_M$ is valid, we may proceed directly as above, or we may employ the adjoint of $A, A'$. The second method is worth discussing, since it may be generalized to apply to more general situations involving more complicated linear operators.

Since the characteristic roots of $A$ and $A'$ are the same, we have $\lambda_M(A) = \lambda_M(A')$.

Let $(x, y)$ denote the vector inner product

$$ (x, y) = \sum_{k=1}^{N} x_k y_k. $$

Then,

$$ (Ax, y) = (x, A'y). $$

If $Ay \leq \lambda y$, for some $y > 0$, we have, for $z \geq 0$,

$$ \lambda(z, y) \geq (z, Ay) = (A'z, y). $$

Let $z$ be a characteristic vector of $A$ associated with $\lambda_M$. Then,

$$ \lambda(z, y) \geq (A'z, y) = (\lambda_M z, y) = \lambda_M(z, y). $$

Since $(z, y) > 0$, we obtain $\lambda \geq \lambda_M$. This completes the proof of the minimum property.

The proof that $\lambda_M(A)$ is a simple root is a bit more complex. From the variational definition of $\lambda_M$, it follows that $\lambda_M(A) \geq \lambda_M(B)$ if $A \geq B$, with strict inequality if $A > B$. Furthermore, it is true that if $A_N$ is the positive $N \times N$ matrix $(a_{ij})$, and if $A_{N-1}$ is any $(N-1) \times (N-1)$ matrix obtained by striking out an $i$th row and a $j$th column, then

$$ \lambda_N = \lambda_M(A_N) > \lambda_M(A_{N-1}) = \lambda_{N-1}. $$

The proof is by contradiction. Assume, for the moment, that $\lambda_N \leq \lambda_{N-1}$, and take, again without loss of generality, $A_{N-1} = (a_{ij}), i, j = 1, 2, \cdots, N - 1$. We have then the equations

$$ \sum_{j=1}^{N-1} a_{ij} y_j = \lambda_{N-1} y_i, \quad i = 1, 2, \cdots, N - 1, \quad y_i > 0, $$

and

$$ \sum_{j=1}^{N} a_{ij} x_j = \lambda_N x_i, \quad i = 1, 2, \cdots, N, \quad x_i > 0. $$

Using the first $N - 1$ equation in Eq. (5.10), we obtain...
\[ \sum_{j=1}^{n-1} a_{i,j}x_j = \lambda x_i - a_{i,n}x_n = x_i \frac{\lambda - a_{i,n}x_n}{x_i} < \lambda_{n-1}x_i, \quad (5.12) \]

which contradicts the minimum property of \( \lambda_{n-1} \).

We now apply this result to show that \( \lambda(W)(A) \) is a simple root of \( f(\lambda) = |A - \lambda I| = 0 \). To do this, we show that \( f'(\lambda_W) \neq 0 \). Using the rule for differentiating a determinant, we obtain readily

\[ f'(\lambda) = - |A_1 - \lambda I| - |A_2 - \lambda I| \cdots - |A_n - \lambda I|, \quad (5.13) \]

where by \( A_k \) we denote the matrix obtained from \( A \) by striking out the \( k \)th row and column. Since \( \lambda(W)(A) > \lambda(W)(A_k) \) for each \( k \), and each \( |A_k - \lambda I| \) is a polynomial in \( \lambda \) with the same leading term, each \( |A_k - \lambda I| \) has the same sign at \( \lambda = \lambda(W)(A) \). Hence \( f'(\lambda_W) \neq 0 \).

The condition \( A > 0 \) may be relaxed somewhat. However, as consideration of the unit matrix shows, it cannot be altogether weakened to \( A \geq 0 \). Since the characteristic roots of \( A^k \) are the \( k \)th powers of the characteristic roots of \( A \), various conditions can be found from the fact that \( A^k > 0 \) for some \( k \) is sufficient.

To show that, apart from scalar multiples, there is only one characteristic vector corresponding to \( \lambda_W \), assume that we have a second vector, \( z \), not necessarily positive, such that \( Az = \lambda_W z \). Then \( x^0 + \varepsilon z \), for all \( \varepsilon \), is a characteristic vector of \( A \). Varying \( \varepsilon \) about 0, we reach a first value of \( \varepsilon \) for which one or several components of \( x^0 + \varepsilon z \) are zero, with the remaining components positive, provided that \( x \) and \( z \) are linearly independent. This, however, contradicts \( A(x^0 + \varepsilon z) = \lambda_W(x^0 + \varepsilon z) \) if \( A \) is strictly positive.

Combining this result with known results concerning the reduction of a matrix to canonical form, we can give an alternate proof of the simplicity of \( \lambda_W \). If \( \lambda_W \) is not a simple root, the Jordan canonical form for \( A \) shows that there exists a vector, \( y \), with the property that

\[ (A - \lambda_W I)^k y = 0, \quad (A - \lambda_W I)^{k-1} y \neq 0, \quad (5.14) \]

for some \( k \geq 2 \). From this we see that \( (A - \lambda_W I)^{k-1} y \) is a characteristic vector of \( A \) associated with \( \lambda_W \), and hence a scalar multiple of \( x^0 \). By suitable choice of \( y \), we can take the multiple to be 1. Thus, \( x^0 = (A - \lambda_W I)^{k-1} y \). Now, let \( z = (A - \lambda_W I)^{k-2} y \). Then,

\[ Az = \lambda_W z + x^0 > \lambda_W z, \quad (5.15) \]

since \( x^0 > 0 \). Hence, \( A |z| > \lambda_W |z| \), which contradicts the maximum property of \( \lambda \).

**5.3. Expectation Matrices**

If we return to Eq. (5.2) of Sec. 5.1, and take the Laplace transform, proceeding formally, we obtain

\[ L(\mu) = [I - L(dG_{ij})]^{-1}L(f), \quad (5.16) \]
where \( u \) and \( f \) are the vectors whose components are \( u_i \) and \( f_i \), respectively, and where the matrix \( L(dG_{ij}) \) is defined as

\[
L(dG_{ij}) = \left[ \int_0^\infty e^{-st} dG_{ij}(t) \right].
\]  

(5.17)

From this it is clear that the asymptotic behavior of \( u \) will depend on the location of the roots of the transcendental determinantal equation

\[
|I - L(dG_{ij})| = 0,
\]  

(5.18)

which is to say, on the values of \( s \) which render \( I - L(dG_{ij}) \) singular.

In order to avoid extraneous details of secondary importance, we shall assume that

\[
dG_{ij}(t) = \phi_{ij}(t) dt.
\]

The following result was conjectured by T. E. Harris and the author and proved by Bohnenblust, using the variational characterization above.

**Theorem 5.3.** Let

(a) \( \phi_{ij} > 0 \), \( j = 1, 2, \ldots, N \),

(b) \( \int_0^\infty \phi_{ii} dt > 1 \) for some \( i \),

(c) \( \int_0^\infty \phi_{ij} e^{-at} dt < \infty \) for some \( a > 0 \), \( i, j = 1, 2, \ldots, N \).

(5.19)

There is a positive vector \( x \) and a positive number \( s_0 \) for which

\[
\left( \int_0^\infty e^{-s t} \phi_{ij} ds \right) x = x.
\]

(5.20)

Furthermore, \( s_0 \) is the root of \( |I - L(\phi)| = 0 \) with greatest real part, and it is a simple root.

**Proof.** Consider the matrix \( B(s) = \left( \int_0^\infty e^{-st} \phi_{ij} ds \right) \). As \( s \) increases from 0 to \( \infty \), \( B(s) \) decreases steadily; \( B(s_1) < B(s_2) \) if \( s_1 > s_2 \). At \( s = 0 \), by virtue of our assumption that \( \int_0^\infty \phi_{ii} dt > 1 \), the largest characteristic root of \( B(s) \) must be greater than 1, whereas as \( s \to \infty \), \( \lambda_\infty[B(s)] \to 0 \). Therefore, there is exactly one value of \( s \) for which \( \lambda_\infty[B(s)] = 1 \), since \( \lambda_\infty[B(s)] \) is monotone and continuous in \( s \), and the associated characteristic vector, \( x \), is positive.

It remains to establish the extremum property of \( s_0 \). Assume that there exists a root of

\[
|I - L(\phi)| = 0, \quad s = s_0 + \epsilon r,
\]

for which \( \sigma_0 \geq s_0 \). Let \( y \) be the associated vector such that

\[
B(s)y = y.
\]

(5.21)

Then,

\[
|y| = |B(s)y| \leq |B(s)| |y| \leq B(s_0) |y|.
\]

(5.22)
If \( \sigma_0 > \tau_0 \), we have an immediate contradiction of the extremum property, since 
\( B(\sigma_0) < B(\tau_0) \) implies that \( \lambda_{\text{m}}[B(\sigma_0)] < 1 \), which is contradicted by Eq. (5.19b).
If \( \tau = \tau_0 \), we must have equality, which implies, as before, that \( \tau = \sigma_0 \) and that \( y \) is a 
constant multiple of \( x \).

Although the result concerning the simplicity of \( \tau_0 \) seems to be true for systems of 
arbitrary order, we shall restrict ourselves to \( N = 2 \), where the proof is quite simple. The general proof appears to require more complicated tools.

We require the result that \( \phi_{ij} \geq \psi_{ij} \) imply that the root of \( |I - L(\psi)| = 0 \) of 
largest real part, which we know to be positive, under the assumption that \( \int_0^\infty \phi_{ii} dt > 1 \), 
for some \( i \), is not smaller than the corresponding root of \( |I - L(\psi)| = 0 \). If we suppose 
to the contrary, we have

\[
L[\psi(\sigma_0)]x = x, 
\]

with \( x > 0 \) for \( \sigma_0 > \tau_0 \), where \( \tau_0 \) is the value of \( \tau \) for which \( L[\phi(i)] \) has root of largest 
absolute value equal to 1. However, Eq. (5.23) implies

\[
L[\phi(\tau_0)]x \geq L[\psi(\tau_0)x] \geq L[\psi(\sigma_0)x] = x, 
\]

which, in turn, will imply from the variational property that \( L[\phi(\tau_0)] \) has a character-
istic root larger than 1 in absolute value.

In particular, we conclude that the positive roots of

\[
\int_0^\infty e^{-st}\psi_{11} dt - 1 = 0, \quad \int_0^\infty e^{-st}\psi_{22} dt - 1 = 0, 
\]

must be less than the positive root of largest real part \( \tau_0 \) of

\[
f(\tau) = \left| \begin{array}{cc} \int_0^\infty e^{-st}\psi_{11} dt - 1 & \int_0^\infty e^{-st}\psi_{12} dt \\ \int_0^\infty e^{-st}\psi_{21} dt & \int_0^\infty e^{-st}\psi_{22} dt - 1 \end{array} \right| = 0. 
\]

Let us use this fact to show that \( \tau_0 \) is a simple root of \( f(\tau) \). We have

\[
f'(\tau) = \left( - \int_0^\infty te^{-st}\psi_{11} dt \right) \left( \int_0^\infty e^{-st}\psi_{22} dt - 1 \right) \\
+ \left( - \int_0^\infty te^{-st}\psi_{22} dt \right) \left( \int_0^\infty e^{-st}\psi_{11} dt - 1 \right) \\
+ \left( \int_0^\infty te^{-st}\psi_{12} dt \right) \left( \int_0^\infty e^{-st}\psi_{21} dt \right) \\
+ \left( \int_0^\infty e^{-st}\psi_{12} dt \right) \left( \int_0^\infty te^{-st}\psi_{21} dt \right) \\
= (+) + (+) + (+) + (+) > 0, 
\]

at \( \tau = \tau_0 \).
To obtain the result for $N = 3$, we consider the equation

$$f(\lambda) = \begin{vmatrix}
\int_0^\infty e^{-\lambda t}\phi_{11} \, dt - 1 & \int_0^\infty e^{-\lambda t}\phi_{12} \, dt & \int_0^\infty e^{-\lambda t}\phi_{13} \, dt \\
\int_0^\infty e^{-\lambda t}\phi_{21} \, dt & \int_0^\infty e^{-\lambda t}\phi_{22} \, dt - 1 & \int_0^\infty e^{-\lambda t}\phi_{23} \, dt \\
\int_0^\infty e^{-\lambda t}\phi_{31} \, dt & \int_0^\infty e^{-\lambda t}\phi_{32} \, dt & \int_0^\infty e^{-\lambda t}\phi_{33} \, dt - 1
\end{vmatrix} = 0.$$  

(5.28)

If

$$\phi_{12} = a_{12}\phi_{11}, \quad \phi_{13} = a_{13}\phi_{11}, \quad \phi_{21} = a_{21}\phi_{22}, \quad \phi_{23} = a_{23}\phi_{22}, \quad \phi_{31} = a_{31}\phi_{33}, \quad \phi_{32} = a_{32}\phi_{33},$$

(5.29)

where $a_{ij} \geq 0$, this may be written

$$\begin{vmatrix}
1 - \frac{1}{\int_0^\infty e^{-\lambda t}\phi_{11} \, dt} & a_{12} & a_{13} \\
1 - \frac{1}{\int_0^\infty e^{-\lambda t}\phi_{22} \, dt} & a_{21} & a_{23} \\
1 - \frac{1}{\int_0^\infty e^{-\lambda t}\phi_{33} \, dt} & a_{31} & a_{32}
\end{vmatrix} = 0.$$  

(5.30)

The proof that the root with largest real part is simple may now be obtained by using the analogue of Eq. (5.13). This method, however, does not seem to extend to the general case where the $\phi_{ij}$ are unrelated.

5.4. Asymptotic Behavior

To determine asymptotic behavior, it is now necessary to make further assumptions concerning $f_1$ and $\phi_{ij}$ which will permit either a shift of the contour of integration past the point $s = s_0$ or the use of Tauberian theorems. In any case, the continuation is quite similar to the one-dimensional procedure.

5.5. Applications

At the moment, the only applications of $N$-dimensional systems of renewal equations appear to be those occurring in the theory of $N$-dimensional branching processes. Those interested may refer to the excellent expository paper of T. E. Harris, mentioned in the References of Chapter 3, where further references may be found. The principal sources are the papers of Hawkins and Ulam [3], and Everett and Ulam [1].
REFERENCES


CHAPTER 6
ON THE ZEROS OF EXPONENTIAL POLYNOMIALS

6.1. Introduction

In the preceding chapters we have studied the analytic nature of the solutions of two
particular classes of linear functional equations—the differential-difference equation and
the renewal equation. In the chapters to follow we shall concentrate on differential-
difference equations, focusing our attention on the subject of stability. We shall also
discuss, in some detail, the way in which these equations arise in applications, with
particular reference to control problems and to the field of mathematical economics.

In our study of the asymptotic behavior of the solutions of differential-difference equa-
tions, we observed that the growth of the solutions, as $t$ becomes arbitrarily large, was
determined by the location in the complex plane of the zeros of certain simple classes of
entire functions. These have the form

$$
\sum_{k=0}^{N} a_k t^k e^{-\lambda_k t},
$$

and are usually called "exponential polynomials." They possess a number of interesting
properties which have been investigated by, among others, Ritt [9] and Lax [6].

The problem of determining the nature of the zeros of functions of this class is one
that has received a great deal of attention owing to the frequency with which it enters into
many different fields. Any adequate treatment of the problem would itself require a
separate survey. We refer the interested reader to the expository article by Langer [5]
and to the excellent treatise of Meiman and Chebotarev [7].

In considering the question of the stability of solutions of differential-difference equa-
tions, it is sufficient to know whether or not all the zeros lie to the left of the imaginary
axis. For the case of ordinary polynomials, the problem was first resolved by Hurwitz
(see [7] in the References to Chapter 8). His results will be presented in "Control
Problems," Chapter 8. The fundamental results for exponential polynomials were first
presented by Pontryagin [8]. We can do no better in summarizing his results than to
present the English summary appended to his original paper.

Following this, we shall apply these results to a number of equations which occur in
the theory of control and in mathematical economics.

An elegant technique, which we shall not discuss here, is due to Collatz [1]. We
refer the reader to his paper for details.

6.2. Summary of Pontryagin's Results

The following is taken verbatim from the English summary of Pontryagin's paper,
which was cited above, with the exception of minor grammatical changes.

"Let \( b(z, t) \) be a polynomial, with real or complex coefficients, depending on the variables \( z \) and \( t \). In this paper we shall give the necessary and sufficient conditions for the negativity of the real parts of the zeros of the function \( H(z) = b(z, e^t) \).

"Let \( r \) and \( s \) be the degrees of the polynomial \( b(z, t) \) with respect to \( z \) and \( t \). We call the term of the polynomial \( b(z, t) \) containing the product \( z^r t^s \) the principal term. Evidently, not every polynomial has a principal term.

"**Theorem 6.1.** If the polynomial \( b(z, t) \) does not contain a principal term, then the function \( H(z) \) possesses an infinity of zeros with arbitrarily large positive real parts.

"Proof. The solution of the equation \( H(z) = 0 \) is sought in the form \( z = \alpha \ln 2k\pi + 2\kappa i + \ln \theta + \xi \), where \( \alpha \) is a positive rational number, \( \theta \neq 0 \) is a complex number depending on the form of the polynomial \( b(z, t) \), \( k \) is a sufficiently large positive number, and \( \xi \) is a sufficiently small variable. It follows that if \( b(z, t) \) does not possess a principal term, there will exist solutions of this kind with \( \xi \) tending to zero as \( k \to \infty \).

"If there exists a principal term of the polynomial \( b(z, t) \), then the character of the zeros of \( H(z) \) is determined by the behavior of \( H(z) \) on the imaginary axis. If \( H(iy) = F(y) + ig(y) \), where \( F(y) \) and \( G(y) \) are real functions of the real variable \( y \), then it is easily seen that \( F(y) = f(y, \cos y, \sin y) \) and \( G(y) = g(y, \cos y, \sin y) \), where \( f(y, u, v) \) and \( g(y, u, v) \) are polynomials with real coefficients. Our problem can be reduced to the investigation of the zeros of the function \( F(z) = f(z, \cos z, \sin z) \), where \( f(z, u, v) \) is a polynomial with real coefficients. \( f(z, u, v) \) can be represented in the form

\[
f(z, u, v) = \sum_{m, n} z^m u^n f_m(n, u, v),
\]

where \( f_m(n, u, v) \) is a homogeneous (with respect to the variables \( u, v \)) polynomial of degree \( n \). As we are going to use the substitution \( u = \cos z, v = \sin z \), we assume, without loss of generality, that no \( f_m(n, u, v) \) is divisible by \( u^2 + v^2 \) or that \( f_m(n, 1, i) \neq 0 \) for every polynomial \( f_m(n, u, v) \) in the sum in Eq. (6.2). Let \( r \) be the degree of \( f(z, u, v) \) with respect to \( z \), and let \( s \) be its degree with respect to both \( u \) and \( v \). We shall call \( z^r f_r(z, u, v) \) the principal term of \( f(z, u, v) \). Evidently, not every polynomial contains a principal term.

"**Theorem 6.2.** If the polynomial \( f(z, u, v) \) does not contain the principal term, then there exists an infinity of zeros of the function \( F(z) \) with arbitrarily large imaginary parts.

"The proof is based on Theorem 6.1.

"If the polynomial \( f(z, u, v) \) contains a principal term, then we select the terms containing \( z^r \), and we write \( f(z, u, v) = z^r f_r(z, u, v) \). The function \( \psi_1(e + iy) \) does not vanish for any real \( y \). Choose \( \epsilon \) so that \( \psi_1(e + iy) \neq 0 \) for all \( y \).

"**Theorem 6.3.** The function \( F(z) \) possesses exactly \( 4k\pi + r \) zeros in the region \(-2k\pi + \epsilon \leq x \leq 2k\pi + \epsilon \) if \( k \) is sufficiently large. Thus, the necessary and sufficient condition that all the zeros of the function \( F(z) \) be real is that there exist exactly
THE ZEROS OF EXPONENTIAL POLYNOMIALS

2ks + r real zeros of the function \( F(x) \) in the interval \(-2k\pi + \epsilon \leq x \leq 2k\pi + \epsilon\) if \( k \) is sufficiently large.

**Proof.** Consider the rectangle \( P_{ba} \) in the z-plane determined by the inequalities

\[-2k\pi + \epsilon \leq x \leq 2k\pi + \epsilon, \quad -b \leq y \leq b.\]

It is easy to see that \( F(z) = z^n e^{\text{i}n}(z)[1 + \delta(z)] \), where \( \delta(z) \) is arbitrarily small on the boundary of \( P_{ab} \) when \( k \) and \( b \) are sufficiently large. According to the theorem on the logarithmic residue (Rouché’s theorem), the functions \( F(z) \) and \( z^n e^{\text{i}n}(z) \) have the same number of zeros in \( P_{ab} \), \( k \) and \( b \) being sufficiently large. For the latter function this number is evidently equal to \( 4ks + r \).

Consider again the function \( H(z) \) under the assumption that \( b(z, t) \) has a principal term. Selecting the terms with \( z^t \), we write \( b(z, t) \) in the form \( b(z, t) = z^t \chi(t) + \cdots \), where \( \chi(t) \) is a polynomial of degree \( s \) with respect to \( t \). \( X(t) = \chi(t) e^t \) is a periodic function with the period \( 2\pi \), which possesses only a finite number of zeros in the region \( 0 \leq \gamma < 2\pi \). Consequently, for almost every real \( \epsilon \), \( X(t)(x + i\epsilon) \neq 0 \) for all real \( x \). Suppose that \( \epsilon \) is chosen in such a way that these inequalities hold.

**Theorem 6.4.** Let \( b(z, t) \) be a polynomial with a principal term and such that \( H(z) \) does not vanish anywhere on the imaginary axis. Denote by \( N_k \) the number of zeros of the function \( H(z) \) in the region \(-2k\pi + \epsilon \leq y \leq 2k\pi + \epsilon, \quad x > 0 \) (\( z = x + iy \)).

Denote further by \( V_k \) the angle drawn by the vector \( w = H(z) \) around the origin when \( \gamma \) ranges through the interval \(-2k\pi + \epsilon \leq \gamma \leq 2k\pi + \epsilon \). It appears that \( V_k = 2\pi (2ks - N_k + r/2) + \delta_k \), where \( \delta_k \) tends to zero as \( k \to \infty \).

**Proof.** Consider a rectangle \( P_{ba} \) in the z-plane determined by \( 0 \leq x \leq a, \quad -2k\pi + \epsilon \leq y \leq 2k\pi + \epsilon \). As is known, the number of zeros of the function \( H(z) \) in \( P_{ab} \) is equal to the number of complete turns of the vector \( w = H(z) \) when \( z \) passes along the boundary of \( P_{ab} \) in the positive sense (counterclockwise). As is easy to establish, \( H(z) = z^n e^{\text{i}n}(z)[1 + \delta(z)] \), where \( \delta(z) \) tends to zero when both \( k \) and \( n \) become infinite, if \( z \) belongs to the boundary of \( P_{ab} \), except for the side \( x = 0 \). In view of this relation, we can prove that the angle drawn by the vector \( w \), when \( z \) passes along the lower, the right, and the upper sides of the rectangle \( P_{ba} \), is equal to \( 4nks + \pi\epsilon + \delta_k \).

The assertion of the theorem follows immediately from this evaluation.

**Theorem 6.5.** Let \( b(z, t) \) be a polynomial with the principal term and let \( H(z) = F(y) + G(y) \), where \( F(y) \) and \( G(y) \) take on real values whenever \( y \) is real.

If all the zeros of \( H(z) \) have negative real parts, the vector \( w = H(z) \) circulates around the origin in the positive direction with a positive velocity when \( \gamma \) varies from \(-\infty \) to \( \infty \), which can be expressed in the form \( G(y) F'(y) - F'(y) G(y) > 0 \). Furthermore, the vector \( w \) draws the angle \( 4nks + \pi\epsilon + \delta_k \), with \( \delta_k \) tending to zero as \( k \to \infty \), when \( \gamma \) ranges through the interval \(-2k\pi \leq y \leq 2k\pi \). Conversely, if the vector \( w \) draws the angle \( 4nks + \pi\epsilon + \delta_k \), when \( \gamma \) ranges through the interval \(-2k\pi \leq y \leq 2k\pi \), then \( w \) circulates in the positive direction with a positive velocity, and all the zeros of the function \( F(z) \) have negative real parts. In the second assertion of this theorem, \( H(z) \) is supposed to have no purely imaginary zeros.

**Theorem 6.5** will be proved simultaneously with Theorem 6.6. The formulation of
the latter will be preceded by the following remark concerning terminology: Let $p(y)$ and $q(y)$ be real functions of the real variable $y$; we shall say that the zeros of these functions alternate if all their zeros are simple; that between every two zeros of one of these functions there exists at least one zero of the other; and finally that $p(y)$ and $q(y)$ have no common zeros.

**Theorem 6.6.** Let $b(z, t)$ be a polynomial with a principal term and let $H(iy) = F(iy) + KG(iy)$. If all the zeros of the function $H(z)$ have negative real parts, then all the zeros of $F(iy)$ and $G(iy)$ are real, alternate, and

$$G'(i)F(i) - G(i)F'(i) > 0. \quad (6.3)$$

Further, each of the following conditions is sufficient in order that all the zeros of the function $H(z)$ have negative real parts:

(a) all zeros of $F(iy)$ and $G(iy)$ alternate and are real, and Eq. (6.3) holds for at least one value of $y$;

(b) all zeros, $\gamma_0$, of $F(iy)$ are real, and Eq. (6.3) holds for every zero $\gamma_0$, i.e., $F'(\gamma_0)G(\gamma_0) < 0$;

(c) all zeros, $\gamma_0$, of $G(iy)$ are real, and Eq. (6.3) holds for every $\gamma_0$, i.e., $G'(\gamma_0)F(\gamma_0) > 0. \quad (6.4)$

**Proof of Theorems 6.5 and 6.6.** If all the zeros of the function $H(z)$ have negative real parts, then, by Theorem 6.4, $V_b = 4k\pi + \pi r + \delta_b$. Thus the curve drawn by the variable $w = H(w)$, when $y$ ranges through the interval $-2\pi \leq y \leq 2\pi$, has at least $4k + r$ intersection points, with each straight line passing through the origin. If $\lambda w' + \mu w'' = 0$ ($w = w' + iw''$) is the equation of such a straight line, then the function $\lambda F(y) + \mu G(y)$ has at least $4k + r$ different zeros in the interval $-2\pi \leq y \leq 2\pi$. According to Theorem 6.3, $\lambda F(y) + \mu G(y)$ cannot possess more zeros in this interval. Consequently, all zeros of $\lambda F(y) + \mu G(y)$ are real and simple. This implies that the vector $w$ circulates in the positive sense when $y$ increases. Consequently, the zeros of $F(y)$ and $G(y)$ alternate. Thus the first parts of the theorems are proved.

If $V_b = 4k\pi + \pi r + \delta_b$, then, according to Theorem 6.4, all zeros of the function $H(z)$ have negative real parts; this is the second assertion of Theorem 6.5. Further, if one of the conditions of Theorem 6.6 is fulfilled, then by geometrical considerations and by Theorem 6.3, we have $V_b = 4k\pi + \pi r + \delta_b$. This completes the proof of Theorem 6.6."

**6.2.1. Example 1.** The equation $H(z) = 0$, where

$$H(z) = pe^z + q - ze^z. \quad (6.5)$$

has been the subject of investigations by Frisch and Holme [2], Kalecki [4], and Hayes [3]. Hayes was the first to give a complete solution to the question: When are all the real parts of the roots of Eq. (6.5) negative? The papers of Frisch and Holme and of
Kalecki contain only special cases of Hayes’ result. The method used by Hayes was independent of Pontryagin’s results and is interesting in itself. We shall, however, use Theorem 6.5 to derive Hayes’ result, since it illustrates the power of the process very nicely.

**Theorem 6.7 (Hayes).** The roots of \( pe^z + q - ze^z = 0 \), where \( p \) and \( q \) are real, have negative real parts if, and only if,

\[
\begin{align*}
(a) & \quad p < 1, \\
(b) & \quad p < -q < \sqrt{a_1^2 + p^2},
\end{align*}
\]

where \( a_1 \) is the root of \( a = p \tan a \) such that \( 0 < a < \pi \). If \( p = 0 \), we take \( a_1 = \pi/2 \).

Figure 1 illustrates the region of the \((p, q)\)-plane in which all the roots of Eq. (6.5) have negative real parts.

**Proof.** The Condition Is Necessary. In what follows, \( a \) will be a real variable. We begin by writing

\[
H(ia) = F(a) + iG(a),
\]

where

\[
\begin{cases}
F(a) = e^a \sin a + a \cos a, \\
G(a) = e^a \cos a.
\end{cases}
\]

With \( \alpha = a + bi \), we have

\[
F(\alpha) = e^\alpha \sin \alpha + \alpha \cos \alpha,
\]

\[
G(\alpha) = e^\alpha \cos \alpha.
\]

Now, let \( v_1 = -q < -p \leq v_2 \) be the points such that

\[
\begin{align*}
\text{(a)} & \quad p < 1, \\
\text{(b)} & \quad p < -v_1 < \sqrt{v_1^2 + p^2},
\end{align*}
\]

where \( v_1 \) is the root of \( a = p \tan a \) such that \( 0 < a < \pi \). If \( p = 0 \), we take \( v_1 = \pi/2 \).

Figure 1 illustrates the region of the \((p, q)\)-plane in which all the roots of Eq. (6.5) have negative real parts.
where
\[
F(a) = p \cos a + a \sin a + q,
\]
\[
G(a) = p \sin a - a \cos a.
\] (6.8)

From Theorem 6.5, we see that \(G(a) = 0\) must have all its roots real and simple. It is easy to see, employing Theorem 6.3, that we must then have \(p < 1\). In fact, according to Theorem 6.3, there must be precisely \(4k + 1\) roots of \(G(a) = 0\) on \((-2k\pi, 2k\pi)\) if \(k\) is sufficiently large (here, since \(g_*(a) = \cos a\), we may take \(\epsilon = 0\)). We write \(G(a) = 0\) in the form
\[
\tan a = \frac{a}{p}.
\] (6.9)

First observe that \(p \neq 1\); otherwise Eq. (6.9) would have a triple root at the origin, which would contradict the inequality of Eq. (6.3). Now, for any \(p\), and for any integer \(n\) neither 0 nor \(-1\), there is on the interior of the interval \([n\pi, (n+1)\pi]\) exactly one real root of Eq. (6.9). Now suppose that \(p > 1\). Then, except for the root at the origin, there would be no root of Eq. (6.9) on either \((-\pi, 0)\) or \((0, \pi)\). Then, regardless of how large we should choose the integer \(k\), we would find on the interval \([-2k\pi, 2k\pi]\) only \(4k - 1\) real roots. This is two less than we should have; accordingly, \(p < 1\).

There is, then, aside from the root at \(a = 0\), in each interval \([n\pi, (n+1)\pi]\) \((n\) is any integer\) one root of \(G(a) = 0\). Beginning with the root on \(0 < a < \pi\), we label the positive roots \(a_1, a_2, \ldots\), and put \(a_0 = 0\). \(a_1\) is the root on \(-\pi < a < 0\), and so on, for \(a_2, a_3, \ldots\).

At any root of \(G(a)\) we must have, according to Theorem 6.5,
\[
F(a)G'(a) > 0.
\] (6.10)

Applying this at \(a = a_0 = 0\), we obtain \((p + q)(p - 1) > 0\), which by virtue of \(p < 1\) requires that \(p + q < 0\).

At any odd-labeled root, positive or negative, we have
\[
F(a)G'(a) = \frac{1}{\sqrt{a^2 + p^2}} \left( \sqrt{a^2 + p^2} + q \right) (a^2 + p^2 - p).
\] (6.11)

At any even-labeled given root, positive or negative, we have
\[
F(a)G'(a) = \frac{1}{\sqrt{a^2 + p^2}} \left( \sqrt{a^2 + p^2} - q \right) (a^2 + p^2 - p).
\] (6.12)

We first observe that at any nonzero root of \(G(a) = 0\) we have
\[
a^2 + p^2 - p = \frac{a}{\sin^2 a} \left( a - \frac{1}{2} \sin 2a \right).
\] (6.13)

This quantity is obviously always positive at such a root.
We now use \( F(a)G'(a) > 0 \) at \( a = a_i \). From Eq. (6.11) we see that we must have
\[
\sqrt{a_i^4 + p^2} + q > 0. \tag{6.14}
\]
On combining this with the fact that \( p + q < 0 \), we obtain the result desired. Therefore, the condition is necessary.

The Condition Is Sufficient. We assume that \( p < 1 \) and that
\[
p < -q < \sqrt{a_i^4 + p^2}. \tag{6.15}
\]
From Theorem 6.5, employing the condition \( p < 1 \), we see that all the roots of \( G(a) = 0 \) are real. First, then, we observe that
\[
F(a_n)G'(a_n) = (p + q)(p - 1) > 0. \tag{6.16}
\]
Now, since \( a_{-n} = a_n \) for \( n = 1, 2, \ldots \), we need only check the sign of \( F(a)G'(a) \) for \( a_n \), with \( n \) positive. Also, from Eq. (6.15), \( \sqrt{a_i^4 + p^2} - p \) is positive at any root; hence we need only check the value of \( \sqrt{a_i^4 + p^2} + q \) at odd positive roots and the value of \( \sqrt{a_i^4 + p^2} - q \) at even positive roots. Both quantities are increasing with \( a \); hence it suffices to check them both at \( a = a_i \). That the first quantity is positive at \( a = a_i \) follows from the hypothesis of Eq. (6.15). Again, from Eq. (6.15), since \( -q > p \), certainly \( q < \sqrt{-a_i^4 + p^2} \), as we see by considering the two possibilities that \( p > 0, p < 0 \).
This completes the proof of sufficiency.

6.2.2. Example 2. The equation \( H(z) = 0 \), where
\[
H(z) = (z^2 + pz + q)e^{\alpha z}, \tag{6.17}
\]
where \( p \) and \( q \) are real, \( p > 0, q \geq 0, \alpha \) is real and \( \neq 0 \), and \( n \) is a positive integer or zero, appears in many applications (see Ansoff and Krumpansl, 1948, in the Bibliography). Again we ask the question: Under what circumstances do all the roots of \( H(z) = 0 \) lie to the left of the real axis?

We may dispose immediately of the case \( n > 2 \). For if \( r \neq 0 \), then \( H(z) \) has no leading term and therefore has, by Theorem 6.1, roots with arbitrarily large positive real parts. Thus we are left with the cases \( n = 0, n = 1, \) and \( n = 2 \) to analyze. These we shall treat separately, giving the detailed calculation only for \( n = 0 \).

The Case \( n = 0 \). For this case we have
\[
H(z) = (z^2 + pz + q)e^{\alpha z}. \tag{6.18}
\]
On writing, for real \( a \), \( H(ia) = F(a) + iG(a) \), we obtain
\[
F(a) = (q - a^2) \cos a - pa \sin a + r, \quad G(a) = (q - a^2) \sin a + pa \cos a. \tag{6.19}
\]
We shall begin by showing that all the roots of \( G(a) = 0 \) are real. As the leading term of \( G(a) \) is \( a^2 \sin a \), we may take \( \epsilon \) in Theorem 6.3 to be \( \pi/2 \). Referring to Theorem 6.5 again, we see that a necessary and sufficient condition that all the roots of
\[ G(a) = 0 \] be real is that, for sufficiently large \( k \), there are \( 4k + 2 \) real roots of \( G(a) = 0 \) on \([ -2k\pi + \pi/2, 2k\pi + \pi/2 ]\). First observe that the origin is a root. We write \( a_0 = 0 \). Now let us write the equation \( G(a) = 0 \) in the form
\[
\cot a = \frac{a^2 - q}{ap}.
\] (6.20)

Let us count the positive roots of Eq. (6.20) on \((0, 2k\pi + \pi/2]\). Observe first that the function represented by the right-hand side of Eq. (6.20) is concave for \( a > 0 \) and increases steadily and continuously from \(-\infty\) to \( +\infty \) as \( a \) increases from \( 0 \) to \( \infty \). We take \( k \) large enough so that \((a^2 - q)/ap\) is positive when \( a = 2(k - 1)\pi \). The cotangent curve has two branches on each interval \((0, 2\pi), (2\pi, 4\pi), \ldots\). Hence there are exactly two roots of Eq. (6.20) on each of those intervals, and, accordingly, \( 2k \) roots on \((0, 2k\pi]\).

It is easy to see that there is an additional root on \((2k\pi, 2k\pi + \pi/2]\). Hence there are \( 2k + 1 \) positive roots on \((0, 2k\pi + \pi/2]\). Now observe that the largest root on \((0, 2k\pi]\) is in fact on \((0, 2k\pi - \pi/2]\), the left side of Eq. (6.20) being negative and the right side being positive on \((2k\pi + \pi/2, 2k\pi]\). It follows that Eq. (6.20) has \( 2k \) roots on the open interval \((0, 2k\pi - \pi/2]\). Since each member of Eq. (6.20) is odd, we see that on \([-2k\pi + \pi/2, 0]\) there are \( 2k \) roots. Thus Eq. (6.20) has on \([-2k\pi + \pi/2, 2k\pi + \pi/2]\) exactly \( 2k + 1 + 2k + 1 = 4k + 2 \) roots. Hence, by Theorem 6.3, all of the roots of Eq. (6.20) are real, as we set out to prove.

For convenience, we shall label the roots. We have already put \( a_0 = 0 \). Counting to the right, we shall label the positive roots \( a_1, a_2, \ldots \); we shall then label the negative roots by putting \( a_{-1} = -a_i, i = 1, 2, \ldots \).

On turning to Theorem 6.5, it will easily be seen that a necessary and sufficient condition that all the zeros of Eq. (6.18) have their real parts negative is that \( F(a)G'(a) > 0 \) at all the nonnegative roots, \( a_0, a_1, \ldots \), of Eq. (6.20).

Observe first that
\[
F(0)G'(0) = (q + r)(q + p) \tag{6.21}
\]
A simple computation yields
\[
F(a) = r - \frac{\sin a}{ap} \left[ (a^2 - q)^2 + a^2 p^2 \right] \tag{6.22}
\]
and
\[
G'(a) = -\frac{\sin a}{ap} \left[ (q - a^2)^2 + a^2 (p^2 + p) + p q \right] \tag{6.23}
\]
for any nonzero root of \( G(a) = 0 \). Since the expression in brackets in Eq. (6.23) is obviously positive, we see that the sign of \( F(a)G'(a) \) for \( a \neq 0 \) is the same as that of
\[
L(a) = \frac{\sin^2 a}{a^2 p^2} \left[ (a^2 - q)^2 + a^2 p^2 \right] - \frac{r \sin a}{ap} \tag{6.24}
\]
Making use of Eq. (6.23), we simplify this to

$$L(a) = 1 - \frac{r \sin a}{a^p}$$  \hspace{1cm} (6.25)

at any root \(a\) of \(G(a) = 0\) other than \(a = 0\).

Let us first consider the case \(r \geq 0\). Then, evidently, \(F(0)G'(0) > 0\). Now, at an even-labeled root \((a_2, a_4, \text{ etc.}), \sin a < 0\), and so \(L(a)\) is positive. Let us turn to the odd-labeled positive roots. Using Eq. (6.20), we obtain

$$\frac{r \sin a}{a^p} = \frac{r}{a^2 p^2 + (a^2 - q)^2}. \hspace{1cm} (6.26)$$

We examine the quantity inside the square root. It may be rewritten as \(a^4 + (p^2 - 2q) a^2 + q^2\). Suppose first that \(p^2 \geq 2q\). Then this quantity is increasing in \(a\), and so the quantity \((r \sin a)/a^p\) is decreasing as \(a\) runs through the positive odd roots of \(G(a) = 0\). Hence, if

$$L(a_i) = 1 - \frac{r \sin a_i}{a_i^p} > 0, \hspace{1cm} (6.27)$$

then

$$L(a_i) = 1 - \frac{r \sin a_i}{a_i^p} > 0, \hspace{1cm} (6.28)$$

for any odd positive integer \(i\) and thus for any positive integer \(i\), even or odd.

If \(p^2 < 2q\), let \(a_{io}\) be the odd root closest to \(\sqrt[4]{q - p^2}/2\). Then \(a_{io}\) maximizes \(L(a_i)\) for odd positive \(i\), as can be seen from differential calculus. Recalling that

$$\text{sgn} \ L(a) = \text{sgn} \ F(a)G'(a), \hspace{1cm} (6.29)$$

we thus have our result: A necessary and sufficient condition that all the roots of Eq. (6.18) lie to the left of the imaginary axis is that \(p^2 \geq 2q\) and \(L(a_i) > 0\), or that \(p^2 < 2q\) and \(L(a_{io}) > 0\). A similar argument works when \(r < 0\). Thus we have proved the following theorem.

**Theorem 6.8.** Let \(H(z) = (z^2 + pz + q)e^z + r\), where \(p\) is real and positive, \(q\) is real and nonnegative, and \(r\) is real. Denote by \(a_k (k \geq 0)\) the real root of the equation

$$\cot a = \frac{a^2 - q}{p^2 a}, \hspace{1cm} (6.30)$$

which lies on the interval \((k\pi, (k + 1)\pi)\). We define the number \(io\) as follows:

(a) if \(r \geq 0\) and \(p^2 \geq 2q\), \(io = 1\);

(b) if \(r \geq 0\) and \(p^2 < 2q\), \(io\) is the odd \(k\) for which \(a_k\) lies closest to \(\sqrt[4]{q - p^2}\);

(c) if \(r < 0\) and \(p^2 \geq 2q\), \(io = 2\);
(d) if \( r < 0 \) and \( p^2 < 2q \), \( Io \) is the even \( k \) for which \( a_k \) lies closest to \( \sqrt{q} - p^2 \).

\[
(6.31)
\]

Then a necessary and sufficient condition that all the roots \( H(z) = 0 \) lie to the left of the imaginary axis is that

(a) \( r \geq 0 \) and \( \frac{r \sin a_{io}}{pa_{io}} < 1 \), or

(b) \( -q < r < 0 \) and \( \frac{r \sin a_{io}}{pa_{io}} < 1 \).

\[
(6.32)
\]

This completes the analysis for \( n = 0 \).

The cases \( n = 1 \) and \( n = 2 \). The analysis for cases \( n = 1 \) and \( n = 2 \) is analogous to that for \( n = 0 \). Accordingly, we shall not give the proofs of the following theorems:

**Theorem 6.9.** Let \( H(z) = (z^2 + pz + q)e^z + rz^2 \), where \( p \) is real and positive, \( q \) is real and nonnegative, and \( r \) is real. Denote by \( a_k \) \( (k \geq 0) \) the sole root of the equation

\[
\tan a = \frac{q - a^2}{pa},
\]

which lies on the interval \( (k\pi - \pi/2, k\pi + \pi/2) \). We define the number \( Io \) as follows:

(a) if \( r \geq 0 \), \( Io \) is the odd \( k \) for which \( a_k \) lies closest to \( \sqrt{q} \);

(b) if \( r < 0 \), \( Io \) is the even \( k \) for which \( a_k \) lies closest to \( \sqrt{q} \).

\[
(6.34)
\]

Then a necessary and sufficient condition that all the roots of \( H(z) = 0 \) lie to the left of the imaginary axis is that

\[
1 + \frac{r}{p} \cos a_{io} > 0.
\]

\[
(6.35)
\]

**Theorem 6.10.** Let \( H(z) = (z^2 + pz + q)e^z + rz^2 \), where \( p \) is real and positive, \( q \) is real and nonnegative, and \( r \) is real. Denote by \( a_k \) \( (k \geq 0) \) the sole positive root of the equation

\[
\cot a = \frac{a^2 - q}{ap},
\]

which lies on the interval \( (k\pi, (k + 1)\pi) \). For the case \( p^2 - 2q < 0 \), we define the number \( Io \) as follows:

(a) if \( r \geq 0 \), \( Io \) is the even \( k \) for which \( a_k \) lies closest to \( q\sqrt{2/2q} = p^2 \);

(b) if \( r < 0 \), \( Io \) is the odd \( k \) for which \( a_k \) lies closest to \( q\sqrt{2/2q} = p^2 \).

Then a necessary and sufficient condition that all the roots of \( H(z) = 0 \) lie to the left
of the imaginary axis is that

(a) \[ p^2 - 2q \geq 0 \text{ and } -1 \leq r \leq 1, \text{ or} \]

(b) \[ p^2 - 2q < 0 \text{ and } 1 + \frac{rd_{10} \sin a_{10}}{p} > 0. \]  \hfill (6.38)

6.2.3. Example 3. This section is devoted to the equation

\[ z^2 e^z + pz + q = 0. \]  \hfill (6.39)

As the details of the study of Eq. (6.39) differ from those involved in the equations of Sec. 6.2.2, and as we are able to study Eq. (6.39) for all real values of \( p \) and \( q \), we shall give those details here. A preliminary heuristic study of Eq. (6.39) will be found in Minorsky [11] in the References to Chapter 8.

**Theorem 6.11.** Let \( H(z) = z^2 e^z + pz + q \), where \( p \) and \( q \) are real. Denote by \( a_2 \) the root of the equation (if there is one)

\[ \sin a = \frac{p}{a}, \]

which lies on the open interval \((0, \pi/2)\). A necessary and sufficient condition that all the roots of \( H(z) = 0 \) lie to the left of the imaginary axis is that

(a) \[ 0 < p < \frac{\pi}{2}, \]

(b) \[ 0 < q < a_2^2 \cos a. \]  \hfill (6.40)

**Proof.** We write, as usual, \( H(ia) = F(a) + iG(a) \), where \( a \) is real. Then

\[ F(a) = q - a^2 \cos a, \]

\[ G(a) = pa - a^2 \sin a. \]  \hfill (6.41)

At a root of \( G(a) = 0 \), other than \( a = 0 \),

\[ G'(a) = -p - a^2 \cos a. \]  \hfill (6.42)

The Condition Is Necessary. Let us begin by showing that \( p \) and \( q \) are both positive. According to Theorem 6.5, we must have \( F(a)G'(a) > 0 \) at every root of \( G(a) = 0 \). In particular, therefore, \( F(0)G'(0) > 0 \); hence \( pq > 0 \), so that \( p \) and \( q \) have the same sign. Suppose that they are both negative. Let \( k \) be any positive integer. On \([-2k\pi, 2k\pi]\), the equation \( F(a) = 0 \) has at most \( 4k \) roots, as can be seen by writing it in the form \( \cos a = \frac{q}{a^2} \). However, according to Theorem 6.3, it has on that interval, with \( k \) large enough, \( 4k + 2 \) roots. Therefore, not all the roots of \( F(a) = 0 \) are real. This, however, contradicts Theorem 6.5; hence \( p \) and \( q \) are both positive.

Let us next show that \( p \geq \pi/2 \). Suppose that \( p \geq \pi/2 \). Then, if \( a \) is on the open interval \((0, \pi/2)\), we have \( G(a) = a^2(p/a - \sin a) > 0 \). Hence \( G(a) = 0 \) has no root on \([0, \pi/2]\) other than 0. As it must, according to Theorem 6.3, have two roots other
than 0 on \([0, 2\pi]\), and therefore on \([0, \pi]\), these roots must lie on \([\pi/2, \pi]\). Since the roots of \(F(a)\) and \(G(a)\) interlace, there must be a root of \(F(a)\) on \([\pi/2, \pi]\). But obviously \(F(a) \geq q > 0\) on \([\pi/2, \pi]\). This contradiction yields the desired result, \(p < \pi/2\).

It follows that there is a root of
\[
\sin a = \frac{p}{a}
\]
on \((0, \pi/2)\) other than 0. As in the statement of the theorem, we denote it by \(a_p\).

Finally let us show that \(q < a_p^2 \cos a_p\). The function \(a^2 \cos a\) increases steadily from zero to a maximum at a point on the interior of \((0, \pi/2)\), and then decreases steadily to zero at \(\pi/2\). Since \(G(a) = 0\) has three roots on the closed interval \([0, \pi]\), \(F(a) = 0\) has two roots on \([0, \pi]\), and therefore on \([0, \pi/2]\). These roots are distinct, for otherwise both \(F(a)\) and \(F'(a)\) would be zero, which contradicts the condition \(-F'(a)G(a) > 0\) of Theorem 6.5. In between these roots, \(a^2 \cos a > q\). But since the roots interlace, there must be a root of \(G(a)\) on the interior of the interval joining them. This root is precisely \(a_p\), which proves the assertion.

We have thus proved the necessity of the condition.

The Condition Is Sufficient: From \(0 < p < \pi/2\) and Theorem 6.3, we see that all the roots of \(G(a) = 0\) are real. As both \(F(a)\) and \(G'(a)\) are even functions of \(a\), it will suffice to prove that \(F(a)G'(a) > 0\) for nonnegative roots \(a\) of \(G(a) = 0\).

First, we observe that \(F(0)G'(0) = pq > 0\). For nonzero roots of \(G(a)\), we may use the formula of Eq. (6.42) for \(G'(a)\). It is clear that we now need to consider separately those roots for which the cosine is positive, which we shall call positive roots, and those for which the cosine is negative, which we shall call negative roots.

1. Positive Roots. Evidently \(G'(a) < 0\) at a positive root; hence we have to prove that \(F(a) < 0\) at such a root. All these roots appear in the first quadrant. If \(a\) and \(a'\) are two successive positive roots, it is easy to see that \(\cos a'' > \cos a'\). One has only to observe that \(p/a'' < p/a'\). Hence, if \(a\) and \(a'\) are two successive positive roots, \(a'' \cos a' 
\geq a_p^2 \cos a_p\). Hence \(F(a) = q - a^2 \cos a' \leq q - a_p^2 \cos a_p < 0\). This disposes of positive roots.

2. Negative Roots. Evidently \(F(a) > 0\) at a negative root; hence we have to prove that \(G'(a) > 0\) at such a root. All these roots appear in the second quadrant. If \(a\) and \(a'\) are two successive negative roots, it is easy to see that \(-\cos a'' > -\cos a',\) so that \(-a'' \cos a'' > -a' \cos a'\). Hence it will suffice to prove that \(G'(a_2) = a_2 - a_p^2 \cos a_2 > 0\), where \(a_2\) is the root of \(\sin a = p\) lying in the second quadrant. Evidently \(G'(a_2) \geq 0\), as \(G(a)\) is positive just to the right of \(a_2\) and is negative just to the left of \(a_2\). If \(G'(a_2) = 0\), then the sine curve, \(\sin a\), would be tangent to the curve \(p/a\), and consequently there would be no root of \(G(a) = 0\) on the open interval \((0, \pi/2)\). But we know there is such a root; accordingly, \(G'(a_2) > 0\). This disposes of negative roots.

This completes the proof of Theorem 6.11.
References


CHAPTER 7

THE STABILITY OF SOLUTIONS OF DIFFERENTIAL-DIFFERENCE EQUATIONS

7.1. Introduction

In motivating a discussion of the stability of solutions of differential-difference equations, let us begin with the observation that any mathematical formulation of physical phenomena is based on approximations of greater or lesser degree. In view of this, an essential point in the evaluation of the merits of a theory is the sensitivity of numerical results to slight changes in hypotheses and data.

The study of the changes in effect which are due to the changes in cause is stability theory. For reasons which are of both physical and mathematical origin, the majority of mathematical theories are linear. The physical origin derives from the fact that it is reasonable to begin by studying small deviations from steady-state behavior. Under the assumption of small deviations, it is permissible to use linear equations. Mathematically, the appeal of linearity lies in its analytic simplicity as opposed to the great difficulties of nonlinearity.

It is, however, necessary to justify the neglect of small terms in any particular application, since it is very easy to construct examples where it is not permissible. The classical result of this type is due to Poincaré and Liapounoff (see [1] and [5]), which is given as

Theorem 7.1. Consider the system of equations

\[
\frac{dy_i}{dt} = \sum_{j=1}^{N} a_{ij} y_j + f_i(y_1, y_2, \cdots, y_N), \\
y_i(0) = c_i, \quad i = 1, 2, \cdots, N,
\]

(7.1)

where

(a) the coefficients \( a_{ij} \) are constants,

(b) all solutions of \( dx_i/dt = \sum_{j=1}^{N} a_{ij} x_j \), \( i = 1, 2, \cdots, N \), approach zero as \( t \to +\infty \),

(c) the \( f_i \) are nonlinear terms in the sense that

\[
|f_i| / \sum_i |y_i| \to 0 \text{ as } y_i \to 0.
\]

Under these circumstances, every solution of Eq. (7.1) for which \( |c_i| \) is sufficiently small approaches zero as \( t \to \infty \).

It is this result which justifies the neglecting of nonlinear terms in studying the stability of systems under small perturbations.
Poincaré also began the study of the asymptotic behavior of linear systems with coefficients which are very close to constant for large \( t \) (see Poincaré [4] and Bellman [3]). Here the results are more complicated, and no single theorem can be singled out as being representative.

In the following section we shall discuss the analogue of Theorem 7.1 for differential-difference equations, due to Wright [10] and Bellman [2], and shall sketch its proof. In the concluding section we shall mention briefly the work of Wright [6-10] in connection with the linear differential-difference equation.

### 7.2. The Analogue of the Poincaré-Liapounoff Theorem

To simplify the presentation, let us consider a very simple type of nonlinear differential-difference equation,

\[
\frac{du(t)}{dt} = au(t) + bu(t - 1) + f[u(t)], \quad t > 1, \quad u(t) = \phi(t), \quad 0 \leq t \leq 1.
\]

The result corresponding to Theorem 7.1 is

**Theorem 7.2.** Consider Eq. (7.3) where

(a) \( a \) and \( b \) are constants,

(b) all solutions of \( \frac{du(t)}{dt} = au(t) + bu(t - 1) \) approach zero as \( t \to +\infty \),

(c) \( |f(u)| / |u| \to 0 \) as \( u \to 0 \).

Under these assumptions, every solution of Eq. (7.3) for which \( \max_{0 \leq t \leq 1} |\phi| \) is sufficiently small approaches zero as \( t \to +\infty \).

The condition that all solutions of \( \frac{du(t)}{dt} = au(t) + bu(t - 1) \) approach zero is equivalent to the requirement that all zeros of the characteristic function \( ze^a - ae^b - b \) have negative real parts.

The proof depends on the conversion of the nonlinear equation of Eq. (7.3) into a nonlinear integral equation by means of Theorem 7.2,

\[
u(t) = v_0(t) + \int_{t_1}^{t} K(t - t_1) f[u(t_1)] \, dt_1,
\]

where \( v_0 \) is the solution of the linear equation. The simplest continuation is that which uses the method of successive approximations, namely,

\[
u_0 = v_0,
\]

\[
u_{n+1} = v_0 + \int_{t_1}^{t} K(t - t_1) f[u_n(t_1)] \, dt_1.
\]

To complete the proof with this method, we require an additional restriction upon \( f(u) \), namely, a Lipschitz condition.
SOLUTIONS OF DIFFERENTIAL-DIFFERENCE EQUATIONS

\[ \frac{f(u_1) - f(u_2)}{u_1 - u_2} \to 0 \]  
(7.7)

as \( u_1, u_2 \to 0 \). For the details, see Wright [10] and Bellman [2].

It was this result which stimulated Hayes (Ref. [3] of Chapter 6) to determine the necessary and sufficient conditions that all the zeros of \( ze^z - ae^z - b = 0 \) have negative real parts (see Sec. 6.3 of Chapter 6).

7.3. Linear Differential-difference Equations with Almost Constant Coefficients

A very detailed discussion of the relation between the asymptotic behavior of the solutions of differential-difference equations of the form

\[ \frac{dv}{dt} = [a + g_1(t)]v(t) + [b + g_2(t)]v(t - 1), \]  
(7.8)

where \( g_1 \) and \( g_2 \) approach zero as \( t \to \infty \), and of the solutions of the unperturbed equation has been given by Wright in a series of papers (see [6–10]).

In [6], Wright shows that the subject of asymptotic behavior of the solutions of linear functional equations is intimately connected with Tauberian theory and may be attacked most elegantly in very general cases by means of the fundamental Tauberian theorem of Wiener and its extensions due to Pitt (see also the article by Pitt, 1947, in the Bibliography).

A start on the theory of the asymptotic development of the solutions of the case in which \( g_1 \) and \( g_2 \) possess asymptotic developments has been made by Cooke [3].

REFERENCES

5. ———, Méthodes nouvelles de la Mécanique Celeste, Vols. 1 and 3, Gauthier-Villars et Fils, Paris, 1892.
CHAPTER 8
CONTROL PROBLEMS

8.1. Introduction

By "control problems" we shall mean the mathematical and physical problems encountered in the course of controlling or improving the performance of a physical system. In general, the problem is that of damping out unwanted oscillations, e.g., those in the form of airplane flutter, in the rolling of a ship, in the "hunting" of a tracking instrument, and in a hundred other everyday appliances.

We shall begin by stating the bare mathematical problem, and then we shall discuss some applications. After a discussion of instantaneous control, we shall turn to some problems involving time lags. These questions will enter into theory and application with greater and greater frequency and will assume a more and more important role in the near future.

8.2. Mathematical Preliminaries

To begin with, let us assume that we are dealing with a system's small perturbations from equilibrium. If the state variables are $x_1, x_2, \cdots, x_N$, the usual assumptions of linearity yield the set of linear differential equations with constant coefficients,

$$
\frac{dx_i}{dt} = \sum_{j=1}^{N} a_{ij} x_j, \quad i = 1, 2, \cdots, N.
$$

The equilibrium position is $x_1 = x_2 = \cdots x_N = 0$, assuming that $|a_{ij}| \neq 0$.

Occasionally situations arise where the $a_{ij}$ are periodic functions of time. Problems of this type are an order of magnitude more difficult (see Stoker [14] and Bellman [2]).

The vector-matrix form of Eq. (8.1) is

$$
\frac{dx}{dt} = Ax.
$$

The idealized initial condition is $x(0) = 0$, the null vector, which keeps $x$ at 0 for all subsequent $t$. However, in practice, random disturbances will occur which will change the state of the system from the equilibrium position, $x = 0$, to one specified by $x = \epsilon$.

The equilibrium state is said to be stable if the system tends to return to this equilibrium state, otherwise it is unstable; but see Bellman [2] for a more extended discussion.

From the elementary theory of linear systems of differential equations, we obtain readily

**Theorem 8.1.** The necessary and sufficient condition that the equilibrium solution,
\( x = 0 \), be stable is that all the characteristic roots of \( A \) have negative real parts.

If even one root with positive real part exists, we will have explosive instability, in the sense that small random disturbances will grow in magnitude. It is important to point out that, in practice, these disturbances do not become unbounded, since eventually the physical assumptions giving rise to Eq. (8.1) are violated. As in the case of airplane flutter, the new physical state may be highly undesirable.

If there exist roots with zero real part, oscillatory motion will occur which will be of either a periodic or almost periodic nature. Even if these solutions do not constitute a menace, they are almost always wasteful and of nuisance value.

The question arises as to how one can determine when the roots of \( A \) all have negative real parts, without actually computing them.

In response to precisely this question, Hurwitz derived the following result:

**Theorem 8.2.** Consider the equation

\[ \lambda^n + a_1 \lambda^{n-1} + a_2 \lambda^{n-2} + \cdots + a_{n-1} \lambda + a_n = 0 \]

(8.3)

and the infinite array

\[
\begin{align*}
& a_1 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\
& a_2 & a_1 & 1 & 0 & 0 & 0 & \cdots \\
& a_3 & a_2 & a_1 & 1 & 0 & 0 & \cdots \\
& a_4 & a_3 & a_2 & a_1 & 1 & \cdots \\
& \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\end{align*}
\]

(8.4)

The necessary and sufficient conditions that all the roots of (8.3) have negative real parts is that the determinants

\[
H_1 = |a_1|, \\
H_2 = \begin{vmatrix} a_1 & 1 \\ a_2 & a_1 \end{vmatrix}, \\
H_3 = \begin{vmatrix} a_1 & 1 & 0 \\ a_2 & a_1 & a_2 \end{vmatrix}, \\
H_4 = \begin{vmatrix} a_1 & 1 & 0 & 0 \\ a_2 & a_1 & a_2 & a_3 \end{vmatrix},
\]

(8.5)

and so on, be positive.

As some examples of the application of this result, let us consider the quadratic, cubic, and quartic equations.

**8.2.1. Quadratic Equation:** \( x^2 + a_1 x + a_2 = 0 \). The relations here are

\[
H_1 = |a_1| > 0, \\
H_2 = \begin{vmatrix} a_1 & 1 \\ 0 & a_1 \end{vmatrix} > 0,
\]

(8.6)
which reduce to \( a_1 > 0 \), \( a_1 d_2 > 0 \). The necessary and sufficient conditions are then
\[
    a_2 > 0, \quad d_2 > 0. \quad (8.7)
\]

**8.2.2. Cubic Equation:** \( x^3 + a_1 x^2 + a_2 x + a_3 = 0 \). Here
\[
    H_1 = |a_1| > 0, \\
    H_2 = \begin{vmatrix}
    a_1 & 1 \\
    a_3 & a_2
\end{vmatrix} > 0, \\
    H_3 = \begin{vmatrix}
    a_1 & 1 & 0 \\
    a_3 & a_2 & a_1 \\
    0 & 0 & a_3
\end{vmatrix} > 0. \quad (8.8)
\]

The necessary and sufficient conditions are then
\[
    a_1 > 0, \quad a_1 a_2 - a_3 > 0, \quad a_3 (a_1 a_2 - a_3) > 0, \quad (8.9)
\]
which reduce to
\[
    a_1, a_2, a_3 > 0, \quad a_1 a_2 > a_3. \quad (8.10)
\]

**8.2.3. Quartic Equation:** \( x^4 + a_1 x^3 + a_2 x^2 + a_3 x + a_4 = 0 \). The Hurwitz criteria yield, after some reduction,
\[
    a_1, a_4 > 0, \quad a_1 a_2 - a_3 > 0, \quad a_3 (a_1 a_2 - a_3) - a_2 a_4 > 0. \quad (8.11)
\]

In order to apply these criteria to problems involving systems of differential equations, the characteristic equation \( |A - \lambda I| = 0 \), must be written out in the form of Eq. (8.4). As far as is known, there seems to be no formulation of the criteria directly in terms of the elements \( a_{ij} \) of the matrix \( A \).

**8.3. Derivation of the Hurwitz Criteria**

There are many different derivations of the Hurwitz criteria, some of which are based on the classical methods of Cauchy and Hermite (see Hurwitz [7]), and some of which are based on continued fractions and linear transformations (see J. Schur [17] and L. Cremer [16]).

A presentation may be found in any of the standard works on electrical circuit theory, such as in Guillemin [18] or in the works cited above.

**8.4. Nonlinear Instantaneous Control**

In some applications, the assumption of linearity is not warranted. In these cases, the equations corresponding to Eq. (8.1) have the form
\[
    \frac{dx_i}{dt} = \sum_{j=1}^{N} a_{ij} x_j + f_i(x_1, x_2, \cdots, x_N), \quad i = 1, 2, \cdots, N, \quad (8.12)
\]
where the $f_i$ are nonlinear functions of the $x_i$, which are usually power series lacking constant and first-order terms.

As a simple example, consider the equation

$$u''(t) + a_1 u'(t) + a_2 \sin u(t) = 0,$$  \hspace{1cm} (8.13)

which describes the motion of a pendulum in the case where the amplitude of the motion is such that the approximation $\sin u \approx u$ is not justified.

The coefficient $a_1 > 0$ represents the damping coefficient, and $a_2 > 0$ represents a spring constant.

The result of Poincaré and Liapounoff, quoted in Theorem 7.1, shows that the equilibrium position, $u = 0$, $u' = 0$, is stable for Eq. (8.13), since it is stable for the linear approximation

$$u''(t) + a_1 u'(t) + a_2 u(t) = 0.$$  \hspace{1cm} (8.14)

The importance of this result lies in its justification of the use of the linear approximation if we are interested in the stability properties of the solution of Eq. (8.13).

The situations in which the solutions of the approximate equation are on the borderline between stability and instability are exceedingly important, since they correspond to the periodic solutions of the nonlinear equations. The most famous example is the Van der Pol equation

$$u'' + \lambda (u^2 - 1)u' + u = 0.$$  \hspace{1cm} (8.15)

Any discussion of equations of this type is again an order of magnitude more difficult (see Minorsky [11] and Stoker [14]).

8.5. Discussion

If the equilibrium state is not stable, there are two things that may be done. Either the system must be redesigned, which is to say that various "circuit" parameters must be changed so as to satisfy the Hurwitz criteria, or additional components of a governing or controlling type must be added to the system. A departure of the original system from its desired state of operation will cause the control system to operate so as to restore the system to equilibrium.

Even if the original system is stable, it may still be necessary to improve its behavior, as, for example, in the pitching and rolling of a ship or a plane.

Let us note in passing that an analogous problem occurs in connection with computational techniques. Even though a particular process converges and is stable, it may still be necessary to improve its rate of convergence before it is usable.

The introduction of a control system is no universal panacea. In general, if a governor is introduced, the resulting equations will be nonlinear. The presence of nonlinear terms automatically introduces the possibility of autonomous oscillation. The phenomenon of "hunting," which is a parasitic oscillation observed in the use of governing devices, is a particularly important example of this.
It is further true that control devices introduce time lags. In our discussion above we assumed tacitly that the control operation was instantaneous. This is, as far as mechanical controls are concerned, in general not true, but is to be interpreted in the sense that the time lag is an appreciable fraction of basic time constants.

It is clear, then, that the control problem is one of quite subtle nature, since elimination of one undesirable feature may create two others, equally undesirable.

8.6. Some Examples

Let us now discuss three examples which will illustrate the above remarks.

8.6.1. Example 1. In connection with the damped pendulum, whose equation was given in Eq. (8.13), it is sometimes desirable to add additional velocity damping. If the damping term, $b_1 u'(t)$, is assumed to act instantaneously, we obtain Eq. (8.13) with a new damping coefficient $(a_1 + b_1)$. If, however, there is a time lag, $\tau$, because of the mechanical transmission, the resultant equation is

$$\ddot{u}'(t) + a_1 u'(t) + b_1 u'(t - \tau) + a_2 \sin u(t) = 0. \quad (8.16)$$

The linear approximation is

$$\ddot{u}'(t) + a_1 u'(t) + b_1 u'(t - \tau) + a_2 u(t) = 0, \quad (8.17)$$

whose characteristic equation is

$$\lambda^2 + a_1 \lambda + b_1 e^{-\lambda \tau} + a_2 = 0. \quad (8.18)$$

Since this equation has an infinite number of roots, the possibilities of the existence of an autonomous oscillation, under suitable choices of the coefficients, have been immensely improved. Furthermore, if $\tau$ is large, the new equation may possess roots with positive real part. In this case, the effect of the control would actually be to transform a stable system into an unstable system!


An electromechanical control system has a load which is driven so that its angular displacement $\theta_0(t)$ follows as closely as possible the changing angular displacement $\theta_1(t)$ of the input shaft. Control is provided by a synchro-transmitter on the output shaft that feeds back a signal corresponding to $\theta_0(t)$. The input differencing synchro generates a voltage, $v_0(t)$, proportional to the error angle

$$\theta_e(t) = \theta_1(t) - \theta_0(t) \quad (8.19)$$

and supplies this voltage to the amplifier. The distortionless amplifier is assumed to introduce no time delay, so that its output voltage is

$$v_0(t) = k \theta_e(t). \quad (8.20)$$

The angular speed $\theta'_0(t)$ of the output shaft is related to the amplifier output by the
differential equation

\[ k_2 v_o(t) = \theta_o'(t) + T_i \frac{d}{dt} [\theta_o'(t)]. \]  \hspace{1cm} (8.21)

Combining Eqs. (8.20) and (8.21), we obtain

\[ k_2 k_1 \theta_s(t) = \theta_s'(t) + T_i \theta_o''(t). \]  \hspace{1cm} (8.22)

By introducing the relation in Eq. (8.19), the final result is

\[ \theta_s(t) = \theta_o(t) + \frac{1}{k_1 k_2} [\theta_o'(t) + T_i \theta_o''(t)]. \]  \hspace{1cm} (8.23)

The characteristic equation which determines the stability of the control system will be

\[ 1 + \frac{s(1 + T_i s)}{k_1 k_2} = 0. \]  \hspace{1cm} (8.24)

If we assume a time lag, in place of Eq. (8.20) we have

\[ v_o(t) = k_1 \theta_s(t - \tau), \]  \hspace{1cm} (8.25)

and in place of Eq. (8.22) we have

\[ T_i \theta_o''(t) + \theta_o'(t) = k_2 k_1 [\theta_s(t - \tau) - \theta_s(t - \tau)]. \]  \hspace{1cm} (8.26)

The characteristic equation determining stability is now

\[ T_i s^2 + s + k_1 k_2 e^{-\tau} = 0. \]  \hspace{1cm} (8.27)

**8.6.3. Example 3.** Let us now discuss the pioneering work of Hartree, Porter, Callender, and Stevenson [6] in connection with the control of temperature.

They begin by pointing out the natural conflict that exists between the two desiderata of a control system: stability and quick response. The shorter the time required for the return of the temperature to its desired value, compared with the time lag, the farther the temperature will overshoot before the control responds to its reaching the normal value. Consequently, there is a danger of inciting an undesirable oscillation if the return to normal value is too rapid.

In many cases, by making the control depend on the time derivatives and the integrals of the temperature, the conflict between the two properties of stability and responsiveness can be greatly lessened.

A qualitative discussion of the effect of altering the time lag will be found in Minorsky [12].

Hartree et al. also point out that the formulation of the problem in terms of differential-difference equations is only approximate. In general, the physical problem is far more complex.

Let us now consider the problem used to illustrate their discussion. Let \( \theta(t) \) denote
the difference between the actual temperature at time $t$ and the desired temperature. Variation of $\theta(t)$ may be due to three causes:

1. Random disturbances, e.g., fluctuations in surrounding temperature;
2. Operation of control gear;
3. Effects due to $\theta$ itself being nonzero.

We write

$$\frac{d\theta}{dt} = D(t) + \epsilon(t) - m\theta(t),$$  \hspace{1cm} (8.28)

where $D(t)$ is due to item (1), $\epsilon(t)$ is due to (2), and $-m\theta(t)$ is due to (3).

Let us assume further that the mode of control is determined by the behavior of $\theta$, so that $D(t)$ affects the control indirectly through the variation of $\theta$ to which it gives rise.

For a system with a time lag, $\epsilon(t)$ depends on the behavior of $\theta$, not at time $t$, but at time $t - T$, where $T$ is the time lag which we shall take to be constant. We write

$$-\tilde{\epsilon}(t + T) = m_{1}\tilde{\theta}(t) + m_{2}\tilde{\theta}(t) + n_{1}\tilde{\theta}(t).$$  \hspace{1cm} (8.29)

In passing, let us note that the authors point out that it is also possible to obtain a control law of the form

$$-\tilde{\epsilon}(t + T) = m_{1}\tilde{\theta}(t) + n_{2}\tilde{\theta}(t) + n_{3}\tilde{\theta}(t),$$ \hspace{1cm} (8.30)

where

$$\tilde{z}(t) + B_{3}\theta(t) = B_{2}\tilde{\theta}(t) + B_{1}\theta(t).$$ \hspace{1cm} (8.31)

Omitting for the moment the inhomogeneous term $D(t)$, we obtain, combining Eqs. (8.28) and (8.29),

$$\dot{\theta}(t) + m\theta(t) = -n_{1}\theta(t - T) - n_{2}\theta(t - T) - n_{3}\theta(t - T).$$ \hspace{1cm} (8.32)

The corresponding characteristic equation is

$$\lambda^{2} + n_{1}\lambda e^{-\lambda T} + m + n_{2}\lambda e^{-\lambda T} + n_{3}\lambda e^{-\lambda T} = 0.$$ \hspace{1cm} (8.33)

The paper concludes with a discussion of the location of the roots of this equation and of the effects of the various coefficients upon these roots.

REFERENCES


Two works on the connection between the Hurwitz criteria and continued fractions are


A classic treatise on electrical circuit theory is

CHAPTER 9
SOME ECONOMIC MODELS

9.1. Introduction

In the previous chapters we have discussed the application of differential-difference equations to various types of control problems involving time lags. Let us now turn to an application of an entirely different sort—one occurring in the theory of mathematical economics.

As we shall see from a discussion of a particularly simple model, the intrinsic structure of an economic system leads inevitably to functional equations which must be of a more complicated type than differential equations. It is the complicated dependence of the future on the past and present which forces this.

We shall consider below a series of papers which appeared in *Econometrica* between the years 1935 and 1938. These papers centered around an attempt by Kalecki to forge a macrodynamic theory of business cycles.

After a brief discussion of Kalecki's ideas, together with the derivation of his basic functional equation—a differential-difference equation of the kind that we have discussed in great detail in the previous pages—we shall present criticisms of his theory. Finally, we shall close with a brief description of some of the results stimulated by his paper.


In this section we shall treat of a mathematical model of the business cycle due to Kalecki [4]. The model was designed to demonstrate that the length of the business cycle could be estimated as a function of other economic parameters.

Two quantities of primary importance in this highly "lumped" model are \( I(t) \), the investment rate, and \( U \), the rate of depreciation of capital goods. The reason for considering a highly simplified economy is a desire to keep the number of independent variables as small as possible and, consequently, the analysis as simple as possible.

The criterion function, which is to measure the state of the economy, is \( I(t) - U \), the excess of the investment rate over the depreciation rate. The oscillation of this function is to mimic the observed business cycles, and the location of the zero points is to determine the lengths of the cycles.

To simplify the situation still further, the producers are taken to be utterly devoid of speculative instinct: they produce only to meet unfulfilled orders. Consequently, there will be a time lag between the acceptance of an order and the delivery of a finished product. Let us take this time lag to be constant and to be equal to one time unit.

Let \( A(t) \) be the production rate. In view of the above, we take the production rate
to be equal to the rate of accumulation of unfilled orders; i.e.,

\[ A(t) = \int_{t-1}^{t} I(t_1) \, dt_1. \] (9.1)

Taking the process to begin at time \( t = 0 \), we see, upon comparing \( \int_{t_0}^{t} I(t_1) \, dt_1 \) with \( \int_{t_0}^{t} A(t_1) \, dt_1 \), that the producer never gets behind, which is to say, the unfilled orders at any time are all orders which have been placed within the last unit of time.

Let us now introduce the additional quantities \( K(t) \), the capital at time \( t \), and \( L(t) \), the rate of delivery of finished goods. We have

\[ K'(t) = L(t) - U. \] (9.2)

From the preceding, we also see that \( L(t) = I(t - 1) \); i.e., that the delivery rate at time \( t \) is equal to the investment rate at time \( t - 1 \). Hence

\[ K'(t) = I(t - 1) - U. \] (9.3)

It remains to make an assumption concerning the investment rate, the quantity of capital, and the production rate. The simplest assumption is that these three quantities are connected by a linear relation. We write

\[ I(t) = m[c + A(t)] - nK(t), \] (9.4)

where \( m, n, \) and \( c \) are constants which will be determined by fitting the results of the theory to the observed facts.

These equations may now be combined to obtain a differential-difference equation for \( I(t) \). Differentiating Eqs. (9.1) and (9.4) and replacing \( A'(t) \) and \( K'(t) \) by their equivalents in terms of \( I(t) \), the result is

\[ I'(t) = m[c + A(t)] - nK(t) \]

By writing \( u(t) = I(t) - U \), Eq. (9.5) becomes

\[ u'(t) = (m - n)u(t - 1), \] (9.6)

the basic equation of Kalecki's theory. Let us set

\[ p = m, \quad q = -(m + n), \] (9.7)

so that Eq. (9.6) has the form

\[ u'(t) = pu(t) + qu(t - 1). \] (9.8)

The associated characteristic equation is

\[ pe^z + q - ze^z = 0, \] (9.9)

which has been discussed in Sec. 6.2.1 of Chapter 6.
The length of the trade cycle will be determined by the pure complex roots of this equation, since we observe that the phenomenon is periodic, or cyclic, in nature.

Let us first dispose of the possibility that Eq. (9.9) has real roots. As a function of a real variable, \( f(z) = pe^z + q - ze^z \) is increasing for \( z < p - 1 \) and is decreasing for \( z > p - 1 \). Upon substituting \( z = p - 1 \), we see that if \( q \geq 0 \), there is always exactly one root, while if \( q < 0 \), there are two, one, or no real roots, depending on whether \( q/p + q \) is greater than, equal to, or less than zero. We therefore assume that this quantity is less than zero and thus ensure that all the roots of Eq. (9.9) are complex.

It follows immediately, from the results of Pontryagin presented in Chapter 6, that the real parts of the roots are uniformly bounded (since \( ze^z \) is the principal term) and actually that there are only a finite number of roots with positive real parts. There are then two roots with greatest positive real part—complex conjugates—and it is not difficult to show that there are only two roots with any particular real part.

The importance of these roots derives from the Fourier expansions we have mentioned previously. Using the contour integral representation of the solution, and assuming, as we may in this case, that it is legitimate to shift the contour to \(-\infty\), we obtain an expansion

\[
\mu(t) = \sum_{k=1}^{\infty} c_k e^{\tau_k t}. \quad (9.10)
\]

The roots are arranged according to the magnitude of the real parts, which are exactly as they occur when we shift the contour. For large \( t \), the two first terms will dominate, corresponding to complex conjugate roots \( \tau_1 \) and \( \tau_2 = \overline{\tau_1} \), and we will have, as \( t \to \infty \),

\[
\mu(t) \approx c_1 e^{\tau_1 t} + c_2 e^{\tau_2 t}. \quad (9.11)
\]

Since \( \mu(t) \) is real, we have \( c_2 = \overline{c_1} \). The approximate expression for \( \mu(t) \) may be written as

\[
\mu(t) \approx c_1 e^{\tau_1 t} \cos(\gamma_1 t + \zeta_1), \quad (\tau_1 = x_1 + iy_1). \quad (9.12)
\]

The zeros of \( \mu(t) \) will tend to become regularly spaced as \( t \to \infty \), with a common distance of \( \pi/\gamma_1 \), and the "trade cycle" will have length \( 2\pi/\gamma_1 \).

The critical parameter is now \( x_1 \). If \( x_1 > 0 \), we have explosive instability, with \( \mu \) assuming arbitrarily large positive and negative values. Since \( \mu(t) = l(t) - U \), with \( l(t) \geq 0 \), we see that at all times we must have \( \mu(t) \geq -U \). Hence \( \mu(t) \) cannot take on arbitrarily large negative values.

If \( x_1 < 0 \), \( \mu(t) \to 0 \) as \( t \to \infty \). In this case, the concept of a trade cycle is meaningless. We are thus forced to assume that \( x_1 = 0 \). As Kalecki states: "This case is of particular importance, as it appears to be nearest to actual conditions. Indeed, in reality, we do not observe any regular progression or depression in the intensity of cyclical fluctuations."

Various authors, notably Frisch and Holme [1] and Samuelson [5] (page 337), have
criticized this assumption that \( x = 0 \). There are three main points in these critiques. In the first place, for \( x_i \) to equal zero, Eq. (9.9) must have two pure complex roots which are conjugate. This means that \( p \) and \( q \) must lie on the boundary of the stability region in the \((p, q)\)-plane. This inflexibility of the values of \( p \) and \( q \) is highly undesirable.

Secondly, because of the linearity of the equation, the amplitude of the asymptotic solution, which determines the amplitude of the cycle, depends linearly on the behavior of \( u(t) \) in some initial interval. We thus have a situation where the steady-state, or stationary, solution is determined by a transient solution. Again, this is undesirable.

Finally, a small change in \( u(t) \) in an initial interval may radically change the values of the coefficients \( c_1 \) and \( c_2 \). These changes, which are of no importance for small \( t \), produce a great effect for large \( t \). In other words, we have a form of instability which is also undesirable.

We see then that, ingenious as Kalecki's formulation is, it is inadequate for the purpose of determining either the amplitude or the length of the trade cycle.

These difficulties are common to all linear theories. It is by now a well-established heuristic canon that any satisfactory theory of periodic or cyclic phenomena requires a nonlinear formulation, or else the presence of exogenous forces which provide forcing terms. A discussion of this point is given by Minorsky [13] in Chapter 8 (see also Samuelson [5]).

9.3. A Brief Survey of the Literature Related to Kalecki's Model

Following Kalecki's paper in 1935, there appeared a paper by Frisch and Holme [1], which discussed the differential-difference equation given in Eq. (9.8). Using the techniques of complex variable theory, they presented some results concerning the location of the zeros of the characteristic function, and gave some numerical techniques for determining these zeros.

E. Theiss, in his book, Dynamics of Saving and Investment, 1935, discussed various economic models, one of which leads to an equation of Kalecki's type. There is, however, no intensive mathematical discussion.

In two papers, [2] and [3], James and Belz continued the investigation of Frisch and Holme. They gave a thorough discussion of the concept of a solution of a differential-difference equation, and used the Laplace transform to determine the solution and its Fourier expansion.

They also drew attention to some further models which lead to the equations

\[
y'(t) = b[y(t - a) - y(t - a - 1)]
\]

(9.13)

and

\[
\lambda y''(t) - y'(t) + by(t) - cy(t - 1) = 0.
\]

(9.14)

The characteristic equation associated with Eq. (9.14),

\[
\lambda z^2 - z + b - ce^z = 0,
\]

(9.15)
has been discussed in Chapter 6. The characteristic equation associated with Eq. (9.13),

\[ z = b(e^{az} - e^{-(a+1)z}) \]  

(9.16)

presents considerably more difficulty. Although the general approach of Pontryagin can
still be followed, the details are now much more perplexing.

REFERENCES


A recent paper of interest is

APPENDIX

SUPPLEMENTARY BIBLIOGRAPHICAL MATERIALS TO THE ARTICLE, "GENERAL THEORY OF DIFFERENTIAL EQUATIONS WITH A RETARDED ARGUMENT"

BY

A. D. Myshkis

Translated by John Danskin, 1952

Since I wrote the paper cited in the title, there have become available to me a number of additional materials on differential-functional equations and related questions. A significant portion of these works was brought to my attention by R. Ya. Shostak, to whom I gladly express my deep gratitude. For the sake of completeness of the survey and bibliography, these materials will be spotlighted here.

Of the older works we cite the articles of Charles [1], Biot [2], and Vernier [3]. Charles obtains a formula expressing the solution of the equation

$$\gamma(x + a) - \gamma(x) = by'(x) \quad (a > 0)$$

on the interval \([ns, (n - 1)s]\) \((n\) an arbitrary integer) in terms of the values of \(\gamma(x)\) on \([0, s]\). Biot shows how to form differential-difference equations by eliminating arbitrary constants, and also gives the simplest geometrical examples; in certain cases he obtains exact solutions of such equations, containing an arbitrary periodic function. Vernier, in particular, solves by special methods the nonlinear difference-differential equation

$$\frac{\gamma(x + a) - \gamma(x)}{a} = \frac{2\gamma'(x)}{1 - (\gamma'(x))^2}$$

which has geometrical application.

In Chapter IX of the textbook of M. Vashchenko-Zakharchenko [4] there is given a survey of certain formal methods of solution of difference-differential equations. In it are described a method of factorization of differential-difference operators into the product of differential and difference operators (cf. OT [4]), the cascade method (OT [6]) and the symbolic method. Section IX, Chapter V, Vol. VI of the Traité d’Analyse of Laurent [5] is also devoted to these equations.

* "Dopolnitel’nye bibliograficheskie materialy k stat’ë Obshchaa teoria differentzial’nykh uravnenii s zapazdyvayushchim argumentom," Uspehi Matem. Nauk, IV-5(33) 99–141 (1949). This work will be designated in the sequel by the letters OT. (Translator’s note: This last paper appeared in English as American Mathematical Society Translation No. 55, New York, 1951.)
On equations in the complex plane we may mention the papers of O. Polosukhina, G. Gruzintsev, L. B. Robinson and A. F. Leont'ev. O. Polosukhina, in the second chapter of her dissertation [6], found, employing the theory of residues, entire transcendental solutions of the equations

\[ y'(x) - \lambda y(x - 1) = \sum a_n x^n. \]

Gruzintsev [7], with the help of the method of successive approximations, and employing contour integrals, found solutions to the equations

\[ y'(x) = ky(x + a) \]

\((a \text{ real})\), the solutions being holomorphic in a strip \(|Imx| < \alpha\). Carmichael [8] studied the solutions of the system

\[ y'_j(x) = \sum_{k=1}^{n} a_{jk}(x) y_k(x) \quad (j = 1, \ldots, m) \]

\[ y_i(x + a) - y_i(x) = \sum_{k=1}^{n} a_{ik}(x) y_k(x) \quad (i = m - 1, \ldots, n) \]

near the point at infinity. Robinson [9] considered various properties of the solutions of equations of the type

\[ y'(x) = \sum_{k=1}^{n} P_k(x) y(a_k x + b_k) + Q(x). \]

Leont'ev [10] investigated the solutions of the equations OT [2], existing in the strip \(\alpha < \text{Im}x < \beta\). He studied, in particular, the expansion of such solutions into series of elementary solutions (given with the help of the roots of the characteristic equation OT [3]). The paper of Leont'ev contains important complements to the work of Hilb (OT [36]): namely, with respect to the expansion of an arbitrary solution \(y(x)\) into a series of elementary solutions, in the case that the whole series does not converge to \(y(x)\) (as one had to assume in the work of Hilb) but rather only a certain subsequence of its partial sums.

We turn to linear differential-difference equations with constant coefficients.

Lecornu [11] applied the theory of residues to obtain solutions of the equation

\[ \frac{y(x + b) - y(x - b)}{2b} = p y'(x) \]

in the form of the sum of a series of functions \(e^{\lambda x}\), where the \(\lambda\) are the roots of the characteristic equation. This work contains also indications of the possibility of the application of the method of Euler and successive integration across intervals to the general equation OT [2].
Neufeld [12] solves with the help of the Laplace transform the equation
\[ \sum_{i=0}^{m} \sum_{k=0}^{k_i} a_{ik} y^{(i)}(x - k) = f(x) \quad (x > 0), \]
where \( y(x) \equiv 0 \) \((x < 0)\), and \( y^{(i)}(0) \) \((i = 0, \ldots, m - 1)\) are given in advance. This method of solution is investigated also in the book of Churchill [13] (pp. 26-28) and in the paper of Bellman [14]. In the latter work the equation
\[ y'(x) + ay(x) + by(x - 1) = f(x) \quad (x \geq 1) \quad (1) \]
is solved under the initial conditions
\[ y(x) = \phi(x) \quad (0 \leq x \leq 1). \]
The same problem (but with \( f(x) = 0 \) in (1), and also for certain equations similar to (1)) was solved by James and Belz [15], using the Fourier transform. This transform makes it possible to obtain an expansion of the solution of the given problem into a series of elementary solutions.

With regard to the characteristic equations OT [3], a large number of properties of their solutions may be found in the monograph of N. G. Chebotarev and N. N. Meiman [16]; by way of example they examine systems of linear differential-difference equations appearing in the theory of regulators. In this same book is included an extensive bibliography, on, in particular, transcendental equations of various types. In addition we call attention to the paper, which came out later, of Yu. I. Neimark [17] in which the space of "quasipolynomials"
\[ \sum_{i=0}^{m} \sum_{k=0}^{k_i} a_{ik} x^{i} e^{\lambda_i x} \quad (m \geq 0, \ k_i \geq 0, \ k_m > 0, \ a_{ik} \neq 0) \quad (2) \]
are subdivided into regions, corresponding to quasipolynomials with one or more roots having positive real parts. Also in this paper is a supplementary bibliography on quasipolynomials.

We note the following theorem of Pontryagin-Chebotarev [16] (p. 255). Let, in the quasipolynomial (2), all the \( b_{ik} \) be real, and suppose that the largest of the \( b_{mk} \) \((k = 1, \ldots, k)\) is less than at least one of the other \( b_{ik} \)'s. Then this quasipolynomial admits roots with arbitrarily large real parts. From this theorem follows a violation of the continuous dependence of the solution on the initial conditions, discussed in OT (cf. OT Sec. 9) for the equation
\[ \sum_{i=0}^{m} \sum_{k=0}^{k_i} a_{ik} y^{(i)}(x + b_{ik}) = 0 \quad (m > 0, \ k_1 \geq 0, \ k_m > 0, \ a_{ik} \neq 0) \]
(where the \( a_{ik} \) and \( b_{ik} \) are all constant and real) with the same conditions on \( b_{ik} \), that is,
with at least one "advancing" argument. Indeed, if we put \( x_k = \alpha_k + i\beta_k \), then there is a certain sequence of roots of the quasipolynomial (2) for which \( \alpha_k \to +\infty \). The sequence of solutions

\[
y_k(x) = \frac{1}{\alpha_k} \text{Re} \left[ \frac{e^{(\alpha_k + i\beta_k)x}}{i\alpha_k + i\beta_k} \right]^{-1}
\]

converges uniformly to zero on the interval \((-\infty, 0)\), although, for an arbitrarily fixed \( \epsilon > 0 \),

\[
\lim_{k \to \infty} \max_{0 \leq x \leq \epsilon} y^{(m-1)}(x) = +\infty,
\]

as desired.

We pass to linear differential-difference equations with variable coefficients. In the case of polynomial coefficients and in the absence of the right-hand side, an original method for finding special solutions was proposed by I. R. Braitshev [18]. Employing the substitution

\[
y(x) = \int_L e^{x-v} f(v) \, dv
\]

(\( L \), some contour) he obtains an ordinary differential equation which must be satisfied by the function \( f(v) \). The order of this equation is equal to the highest degree of the polynomials appearing as coefficients in the original equation. If this order is unity, then \( f(v) \) may be found by quadratures. Braitshev considers also systems of equations of the same type.

O. Poloskhina, in her dissertation [6] (Sec. 3), studies the solutions of equations of the type

\[
y'(x) + a(x)y(x-1) = b(x)
\]

by means of integral equations with infinite limits; she studies the asymptotic properties of the solutions as \( x \to -\infty \). The cascade method of Laplace-Poisson (its invariants, and the possibility of carrying it through) for the equation

\[
y'(x) + a(x)y(x-1) + b(x)y(x) + c(x)y(x-1) = 0
\]

is considered by Borden [19] and [20], although formally. A rigorous foundation for his procedure was given by Williams [21].

An important development in the symbolic method appeared in the work of R. Ya. Shostak [22]. He considers the general equations

\[
\sum_{i=0}^{m-1} \sum_{k=0}^{n} a_{ik}(x)y^{(i)}(x - kb) = f(x) \quad (m \geq 0, n \geq 0, b > 0, a_{nn}(x) = 1)
\]

under initial conditions analogous to those given in OT (Sec. 2). By inverting the
operations of taking differences and differentiation, on the basis of the symbolic method there results an explicit formal solution of this equation, the validity of which is assured by the convergence of the series involved. This work appears far advanced in its direction.

De Bruijn [23] considered the asymptotic behavior of the solutions of the equation

\[ y'(x) + a(x)y(x) + b(x)y(x - 1) = c(x) \]

as \( x \to +\infty \) (properties of boundedness, periodicity, and so forth).

Turning to nonlinear differential-difference equations, we mention first of all the paper of Bennett [24], in which appear certain very preliminary considerations concerning the existence of a solution to the equation

\[ y(x + b + 1) = F[x, y(x), y'(x), \cdots, y^{(k)}(x), y(x - 1), \cdots, y^{(k)}(x - 1), \cdots, y^{(k)}(x - b)] \]

where \( b \) is an integer \( \geq 0 \), \( k > 0 \), and \( k \) exceeds all the \( k_i \). An algebraic theory of such equations, analogous to the theory of Ritt and Doob for pure differential and difference equations, will be found in the paper of Herzog [25].

Bellman, in the paper [14] mentioned above, studies equations

\[
y'(x) = a_1y(x) + a_2y(x - 1) + b_1(x)y(x) + b_2(x)y(x - 1) + \sum_{n \geq 1} b_{1n}(x)[y(x)]^n[y(x - 1)]^n \tag{3}
\]

under assumptions on the smallness of the functions \( b_1(x) \), \( b_2(x) \), and under corresponding restrictions on the series appearing on the right side of equation (3). The idea of the investigator consists in that if we consider provisionally the sum of the terms on the right containing the \( b(x) \) as coefficients as known, then we may, by employing operational methods, obtain an expression for \( y(x) \) as a certain integral. In this way he obtains an integral-functional equation, which is then solved by successive approximations. He studies also the behavior of the solutions as \( x \to +\infty \).

The dissertation of Poloukhtina [6], already mentioned, contains in Sec. 1 investigations of the general differential-functional equations

\[ y'(x) + a(x)y[a(x)] = f(x) \quad (a \leq x \leq b, a \leq a(x) \leq b) \]

and analogous equations of higher order, by means of reducing them to integral equations (cf. OT, Sec. 1, pp. 105–106) and applying the theory of Fredholm. She succeeds in proving the existence and uniqueness of a solution under the assumption \( a(x) \leq x \) (and corresponding conditions on the functions involved). This assumption is not fortuitous; it ought to be expected from OT.

Elementary considerations concerning the degree of arbitrariness in a general solution of certain differential-functional equations will be found in the work of Popovici [26].

\* Translator's note: In American Mathematical Translation No. 55, pp. 11–12.
The papers [27]–[32] contain investigations of differential-difference equations in connection with their applications outside mathematics. Among the new methods of investigation we mention the passage from the equation
\[ y''(x) + ay'(x) + by(x) = cy(x - b) \]
to an integro-functional equation, by means of the variation of arbitrary constants, carried out in the paper of Collatz [31]. Solutions of linear differential-difference equations met with in the applications are expounded in detail and with due rigor in the monograph of B. V. Bulgakov [32]. Using operational methods he obtains a solution of the initial conditions (coinciding with that given in OT) in the form of the sum of a series, certain residues serving as its coefficients. The paper discusses the reduction of errors in examples under various initial conditions, and also gives corresponding graphs, investigations of the roots of the characteristic equation and so forth.

Among the papers on equations closely related to differential-functional equations, we mention the paper of Germany [33] on the solution of the equation
\[ y'(x) = F \left[ x, y(x), \int_{s_1}^{s_2} f_1(x, s, y(s)) \, ds, \ldots, \int_{s_n}^{s_{n+1}} f_n(x, s, y(s)) \, ds \right], \]
and the paper of James and Belz [34], dealing with the equation
\[ y'(x) = ay(x) + \int_{0}^{x} f(t) y(x - t) \, dt \]
in connection with its applications.

**LITERATURE CITED**


BIBLIOGRAPHY


       ———, "Bemerkung zu meiner Arbeit Über die Funktional-Differentialgleichung \( f(x + 1) = af(x) \)," ibid., pp. 239–246.


BIBLIOGRAPHY


Volterra, V., Teoria de los funcionales y de las ecuaciones integrales e integro-differentiales, Imprenta Clásica Española, Madrid, 1927.

1928


1929


1930


Bruwier, L., "Sur l’équation fonctionnelle \( y^{(n)} = \sum a_n y^{(n-1)}(x+c) + \cdots + a_0(x+c)^r \) + \( a_0(x+c)^r \) = 0," C. R. Congr. Nat. Sci., Bruxelles, 1930, 1931, pp. 91-97.


1931


Volterra, Vito, Leçons sur la théorie mathématique de la lutte pour la vie, Gauthier-Villars & Cie, Paris, 1931.


1932


1933


1934


---

**1935**


"Sur les solutions des systèmes d'équations intégro-différentielles normales considérées comme fonctions des valeurs initiales $x_0$, $x_0$, $x_0$, $x_0$, *ibid.*, Vol. 4, 1935, pp. 329–336; Vol. 5, 1936, pp. 5–11.


**1936**


BIBLIOGRAPHY


1937


König, H., "Periodische und aperiodische Schwingungen an emtpitlichen Regelankor-


Vikker, D. A., "Effekt zapadzianyma w protsessaah automaticheskiego regulirovaniia" (The Effect of Retardation in Processes of Automatic Regulation), Automatika i Teleme-
Tekhnik, No. 6, 1937, pp. 59–76.

1938


BIBLIOGRAPHY


BIBLIOGRAPHY


Svirkov, I. V., "Oprødelenie chislа kornei, lezhashchikh v pravoi poluploskosti, dlia funktsii vla]; F(s, z), gde F(s, z) = rational'naia funktsii ot argumentov s i z i primenenie rezultatov k issledovaniu avtomaticheskogo regulirovaniia parovыхh turbин" (Determination of the Number of Roots Lying in the Right Half-plane for Functions of the Form F(s, z), Where F(s, z) Is a Rational Function of the Arguments s and z, and Application of the Results to the Investigation of the Automatic Regulation of Steam Turbines), Izv. Kazanskogo Fil. AN SSSR (F.M.I.T.), No. 1, 1948, pp. 51-61.


1949


BIBLIOGRAPHY 107


