UNIVERSAL
LIBRARY

OU_162062

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 517 | S65M Accession No. 36315

Author Smith, Lloyd P

Title Mathematical methods for Scientists and Engineers, 1953

This book should be returned on or before the date last marked below.

MATHEMATICAL METHODS

for

SCIENTISTS AND ENGINEERS

Prentice-Hall Physics Series

DR. DONALD H. MENZEL, *Editor*

MATHEMATICAL METHODS

for

SCIENTISTS AND ENGINEERS

BY

Lloyd P. Smith

Chairman, Department of Physics,
Director, Department of Engineering Physics,
Cornell University

PRENTICE-HALL, INC. NEW YORK

1953

Copyright, 1953, by Prentice-Hall, Inc., 70 Fifth Avenue, New York. All rights reserved. No part of this book may be reproduced, by mimeograph or any other means, without permission in writing from the publishers.

Printed in the United States of America.

Library of Congress Card Catalog Number: 53-10561

PREFACE

For several years the author gave a one-year course in Mathematical Methods for first-year graduate students and capable seniors majoring in physics. The chief aim of this course was to provide a means by which a student could obtain a knowledge of the more important mathematical functions and their properties and the facility to apply the more important and useful mathematical methods to problems which he would encounter in theoretical and applied physics. The content of these lectures has been enlarged and broadened to constitute the content of this book. With the exception of the subject of differential equations it is felt that the mathematical methods and functions discussed herein rep esent the minimum knowledge in this field that any Ph.D. in physics should acquire. It has been the author's experience that students who have become thoroughly acquainted with the mathematical methods given here have been able to handle courses in advanced theoretical physics with much greater facility.

It will be evident at once that this book is not written from the standpoint of the mathematician. It has been the foremost objective of the author to present the substance of the mathematical methods together with clear and understandable statements of the conditions which must be satisfied in order that the method or mathematical manipulation in question can be validly carried through. For the most part, no effort has been made to re-do the mathematician's work by providing the proofs for the various and sundry mathematical operations. To do this would have indeed been beyond the scope of any one book covering the variety of fields of mathematics to which the methods outlined in this book belong. In fact the inclusion of all the proofs upon which the various methods are based would not at all have been within the spirit of this book since its principal objective is to provide the means by which a student may acquire facility and practice in the use of available methods of solving physical problems. Those who are interested in the proofs are referred to the general mathematical literature or to selected references which appear at the end of each chapter.

With the exception of the subject of differential equations, the author has attempted to include the methods and information in those branches of mathematics which in his experience are widely used in problems of engineering and science. Differential equations are only incidentally treated as they arise in connection with various important functions or they become involved in applying some of the mathematical methods such as asymptotic solutions. A treatment of differential equations was omitted for two reasons, the most important being the large number of good and accessible treatises on this subject and the other being the simple fact that this book would have become much too long if the extensive subject of differential equations had been adequately treated.

The author has included a number of important illustrative problems which are typical of those which appear in the physical and engineering literature and are therefore representative of such problems. A number of practice problems are also included at the end of each chapter, and many of these have been taken from the scientific literature directly or in modified form and are thus representative of those that would be actually encountered in practice. It will be noticed that as one proceeds through the book the problems may require the use of methods discussed throughout the book up to that point.

It is hoped that the arrangement of the subject matter and methods discussed will be such that the book will be of continuous use to workers in science and engineering as a reference and handbook as well as a textbook in mathematical methods for advanced seniors and graduate students.

The author is deeply indebted to Professor Mark Kac who has written the last chapter on the Elements of Probability Theory. Professor Kac is an authority in this field and has had much experience in the practical application of modern probability theory. In addition the author is indebted to Professor Kac for many stimulating discussions and valuable advice during the preparation of the book.

There remains only the pleasant task of thanking those who have otherwise been of help in the preparation of the book. In particular the author wishes to thank Mr. Paul Kellogg for considerable help in the preparation of the practice problems and also Dr. William Parkins for executing a considerable number of the figures. For a great amount of help connected with the early stages in the preparation of the manuscript the author is indebted to his wife, Florence H. Smith. The author also wishes to express his thanks to Mrs. DeWitt Zeen and Miss Martha Hennessey, who did much of the arduous work of manuscript preparation and arrangement of details.

Finally the author wishes to express his gratitude to Prentice-Hall, Inc., for its helpfulness and fine spirit of cooperation throughout the entire course of publication.

LLOYD P. SMITH

Ithaca, New York

CONTENTS

Chapter 1. Elements of Function Theory **1**

1. Abstract sets, *1*. 2. Point sets and n-dimensional space, *2*. 3. Function of a single variable, *3*. 4. Limiting value of a function, *5*. 5. Continuity of a function, *8*. 6. Limit and continuity of a function of several variables, *9*. 7. Classes of functions, *12*.

Chapter 2. Differential Calculus **14**

8. Derivative and differential of functions of one variable, *14*. 9. Change of variable and higher order derivatives, *16*. 10. Mean value theorems, *17*. 11. Evaluation of indeterminate forms, *18*. 12. Partial derivatives, *22*. 13. Change of variable, *23*. 14. Directional derivative, *23*. 15. Total differential, *26*. 16. Differentiation of functions in parametric form, *29*. 17. Differentiation of implicit functions, *30*. 18. Dependent and independent functions, *32*. 19. Properties of Jacobians, *34*. 20. Taylor's theorem, *35*. 21. Maxima and minima, *36*. 22. Maxima and minima with auxiliary conditions, *39*.

Chapter 3. Integral Calculus **46**

23. The Riemann integral defined, *46*. 24. Classes of bounded integrable functions, *47*. 25. Fundamental properties of Riemann integrals, *48*. 26. Integration by change of variable, *49*. 27. Integration of rational functions, *49*. 28. Integrals reducible to rational functions, *52*. 29. Integrals of irrational functions and elliptic integrals, *53*. 30. Linear transformations of elliptic integrals, *56*. 31. Normal forms when $r(x)$ has real roots, *59*. 32. Normal forms when $r(x)$ has complex roots, *62*. 33. Definite integrals of functions containing parameters, *64*. 34. Inversion of the order of integration, *66*. 35. Improper integrals, *66*. 36. Tests for convergence of infinite integrals, *67*. 37. Uniform convergence of infinite integral with a parameter, *69*. 38. Tests for uniform convergence of an infinite integral, *70*. 39. Theorems concerning uniformly convergent integrals, *71*. 40. Methods of evaluating definite integrals, *71*. 41. Definite integrals with integrands which become infinite at certain points, *74*. 42. Tests for convergence, *75*. 43. Stieltjes integral, *76*.

Chapter 4. Space Geometry **82**

44. Space curves, *82*. 45. Equations and direction cosines of the tangent to a space curve, *83*. 46. Surfaces, *86*. 47. Coordinate transformations, *87*. 48. Curvilinear coordinates, *87*. 49. Parabolic coordinates, *88*. 50. Elliptic coordinates, *89*. 51. The elements of length in the image space, *90*. 52. Conformal and orthogonal transformations, *91*.

Chapter 5. Line, Surface, and Multiple Integrals . . . 93

53. Functions of bounded variation and rectifiable curves, *93*. 54. Line integrals, *95*. 55. Line integrals of total differentials, *98*. 56. Definition of a double integral, *99*. 57. Properties and evaluations of double integrals, *100*. 58. Double integrals over nonrectangular domains, *100*. 59. Connections between double and line integrals: Green's theorem, *102*. 60. Triple integrals, *105*. 61. The element of area on a transformed plane, *106*. 62. Evaluation of definite integrals by double integration, *108*. 63. Elements of length and area on a surface, *110*. 64. Surface areas, *111*. 65. Surface integrals, *113*. 66. Transformation of coordinates in a surface integral, *115*. 67. Green's theorem in three dimensions, *117*. 68. The volume element and its transformation, *118*. 69. Transformation of the Laplacian operator to orthogonal curvilinear coordinates, *121*. 70. Stokes' theorem, *123*.

Chapter 6. Theory of Functions of a Complex Variable . 128

71. Concept of a complex number, *128*. 72. Geometric representation of complex numbers, *129*. 73. Infinite series of complex numbers, *132*. 74. Complex variables and functions of the same, *134*. 75. Uniform convergence, *135*. 76. Tests for convergence of a series, *136*. 77. Power series and their radii of convergence, *139*. 78. Calculations with power series, *141*. 79. Transformations of a power series, *142*. 80. Laurent's series, *143*. 81. Derivatives of complex functions, *144*. 82. Cauchy-Riemann differential equations, *146*. 83. Complex integration, *148*. 84. Integration and differentiation of infinite series, *148*. 85. Cauchy's integral theorem, *149*. 86. Cauchy's integral formula, *151*. 87. The derivative of an analytic function defined by a Cauchy integral, *152*. 88. Taylor's theorem, *153*. 89. Continuation of a power series, *155*. 90. Nature of singularities, *156*. 91. Determination of coefficients in a Laurent series, *157*. 92. Many-valued functions, *158*.

Chapter 7. Residues and Complex Integration 164

93. Theory of residues, *164*. 94. Determination of the residue, *165*. The definite integral $\int_0^{2\pi} f(\theta)\, d(\theta)$, *166*. 96. The integral $\int_{-\infty}^{+\infty} F(x)\, dx$, *167*. 97. Infinite integrals continued, *168*. 98. Integrals of the type $\int_{-\infty}^{+\infty} F(x)e^{i\mu x}\, dx$, *170*. 99. Integrals involving hyperbolic functions, *172*. 100. The principal value of integrals, *174*. 101. Integrals of the form $\int_0^{\infty} x^{m-1}F(x)\, dx$, *176*. 102. Contour integral representation of functions, *178*.

Chapter 8. Representation of Functions by Infinite Series of Functions 182

103. General remarks, *182*. 104. Laurent expansion of cot z and Bernouli numbers, *182*. 105. Expansion of a function in powers of another function, *183*. 106. Inversion of a power series, *185*. 107. Expansion of functions with an infinite number of poles, *186*. 108. Infinite product representation of a function, *188*. 109. Infinite product representation of the gamma function, *190*. 110. Properties of asymptotic expansions, *192*.

111. Calculations with asymptotic expansions, *194*. 112. Methods of obtaining asymptotic expansions, *195*. 113. Asymptotic form of the gamma function, *196*. 114. Method of steepest descents or saddle point method, *201*. 115. Asymptotic solution of a differential equation, *210*.

Chapter 9. Application of Functions of a Complex Variable to Potential and Flow Problems 215

116. Connection between analytic functions and potential theory, *215*. 117. Conformal mapping and the solution of the potential problem, *216*. 118. Methods for determining the proper mapping functions, *220*. 119. Constant potential over a parametrically defined curve, *220*. 120. Potential distributed over a circle, *222*. 121. Polygonal boundaries. Schwarz-Christoffel formula, *225*. 122. Mapping function for rectangular poles, *228*. 123. Electric field in the cyclotron dees, *232*. 124. Mapping of a polygon on a circle, *236*. 125. Two-dimensional flow of an ideal fluid, *237*. 126. Flow of an ideal fluid around a cylinder, *241*. Complex force and moment on immersed obstacles, *242*. 128. Fluid flow with free surfaces, *245*.

Chapter 10. Algebra of Linear Equations, Transformations and Quadratic Forms 252

129. Permutations and determinants, *252*. 130. Certain properties of determinants, *253*. 131. Laplace's development by minors, *255*. 132. Linear equations, *258*. 133. Homogeneous linear equations, *260*. 134. Vectors and vector spaces, *261*. 135. Orthogonalization and completeness, *263*. 136. Orthogonal and unitary spaces, *264*. 137. Linear transformations, *265*. 138. Matrix calculus, *267*. 139. Operational point of view in linear transformations, *271*. 140. Transformation of bilinear, quadratic, and Hermitian forms, *273*. 141. Characteristic vectors and the transformation to principal axes, *267*. 142. Simultaneous reduction of a pair of quadratic forms, *282*. 143. Operations in a space of a countably infinite number of dimensions, *284*.

Chapter 11. Vector and Tensor Analysis 288

144. Vectors in a three-dimensional space, *288*. 145. Algebra of vectors, *289*. 146. Vector fields, *293*. 147. The gradient, divergence, and curl, *295*. 148. Useful vector formulas, *299*. 149. The tensor concept and definition, *300*. 150. Definition of tensors of higher order, *302*. 151. Addition, multiplication and contraction of tensors, *303*. 152. Fundamental and associate tensors, *304*. 153. The covariant derivative, *306*. 154. The nature of the covariant derivative, *307*.

Chapter 12. Orthonormal Function Systems 309

155. Definitions, *309*. 156. Approximation of a function by a linear aggregate of orthonormal functions, *311*. 157. Orthogonalization of linearly independent functions, *313*. 158. Examples of orthonormal sets, *315*. 159. Fourier series, *317*. 160. Functions of arbitrary period, *319*. 161. Properties of functions and their Fourier coefficients, *320*. 162. Mean square spectral extent of a function, *322*. 163. Harmonic analysis of curves, *323*. 164. Hermitian polynomials and orthogonal Hermitian functions, *326*. 165. Laguerre polynomials and the Laguerre orthonormal

functions, *329*. 166. Associated Laguerre polynomials, *334*. 167. Legendre polynomials, *337*. 168. Associated Legendre functions and spherical harmonics, *342*. 169. Bessel functions and Hankel functions, *348*.

Chapter 13. Orthonormal Functions with a Continuous Spectrum *362*

170. The expansion problem in the continuous spectrum, *362*. 171. Fourier integrals and transforms, *363*. 172. The orthogonality concept for functions with a continuous spectrum, *364*. 173. Relation between the mean extent of a Fourier integral and the mean spectral extent of its Fourier transform, *365*. 174. Some examples of Fourier integrals and transforms, *367*. 175. Representations of functions in terms of a general class of functions with discrete and continuous spectra, *369*.

Chapter 14. Integral Equations *371*

176. Integral equations in general, *371*. 177. Some integral equations of the first kind, *371*. 178. Integral equations of the second kind; connection with linear differential equations, *373*. 179. Types of solutions of integral equations, *373*. 180. Method of inversion for equations of the first kind, *375*. 181. Solution of integral equations of the second kind by the method of successive approximations, *378*. 182. The Fredholm method, *380*. 183. Homogeneous integral equations, *387*. 184. Homogeneous integral equations of the second kind with symmetric kernels, *390*. 185. Orthonormal character of the solutions of the homogeneous integral equation with symmetric kernel, *393*. 186. Solution of nonhomogeneous integral equation in terms of the characteristic solutions of the associated homogeneous equation, *394*. 187. Setting up the integral equation associated with a boundary value problem, *398*.

Chapter 15. Variational Methods *403*

188. Typical problems of the calculus of variations, *403*. 189. Nature of the simplest variational problem, *404*. 190. The first variation and the Euler equation, *404*. 191. Variational problems involving several unknown functions, *409*. 192. Variational problems involving several independent variables, *410*. 193. Potential theory and the associated variational problem, *412*. 194. Variational problems with an auxiliary condition, *413*. 195. Direct methods of obtaining solutions of variational problems, *415*. 196. Construction of the minimal sequence, *416*.

Chapter 16. Elements of Probability Theory *424*

197. Preliminaries, *424*. 198. Probabilities assigned to events, *424*. 199. Mutually exclusive events, *426*. 200. Independence, *426*. 201. Random variables, *428*. 202. Expectations, *429*. 203. Binomial distribution, *430*. 204. The normal (Gaussian) distribution as a limiting case, *433*. 205. The continuous case, *435*. 206. An example, *436*. 207. Expectations in the continuous case, *437*. 208. Characteristic functions, *438*. 209. Random flights, *438*. 210. The Holtzmark distribution, *441*. 211. The Poisson distribution, *442*. 212. Generating functions, *444*. 213. An example, *445*.

Index *449*

Chapter 1

ELEMENTS OF FUNCTION THEORY

1. Abstract sets. The concept of a function and the validity of the processes of differential and integral calculus rest essentially on the theory of point sets. The mathematician defines an abstract set to be any collection of objects for which the following two conditions must be fulfilled: (1) A criterion must exist to determine whether any object belongs to the set or not, and (2) those objects, often called elements, belonging to a set must be distinguishable one from another. In this sense the positive integral numbers form a set.

It is often convenient to refer to a given set by a letter to distinguish it from some other set, thus two sets σ_1 and σ_2 may be under consideration. The two sets σ_1 and σ_2 may have all, some, or no elements in common. If no elements are common, they are distinct sets; but it often happens that every element of the set σ_1 is also contained in σ_2, in which case σ_1 is said to be a subset of σ_2. An example of this would be the case where σ_1 is the set of positive integral numbers, and σ_2 is the set composed of all positive real numbers.

The idea of establishing a correspondence of the elements of one set with the elements of another is of fundamental importance to the concept of a function. If to each element of a given set we associate one and only one element of another set, a one-to-one correspondence is said to exist between the two sets. It sometimes happens that the elements of a given set can be put into one-to-one correspondence with the set of positive integers, in which case the given set is said to be *countable*. The set of numbers

$$\begin{array}{ccccccc} 1, & \frac{1}{2}, & \frac{1}{3}, & \frac{1}{4}, & \frac{1}{5}, & \ldots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ 1, & 2, & 3, & 4, & 5, & \ldots \end{array}$$

can be put into one-to-one correspondence with the set of positive integers as shown, and consequently is a countable set. On the other hand, no such correspondence can be established for the set of real numbers.

2. Point sets and n-dimensional space. It is convenient to have a definition of space which will allow one to use convenient geometrical notions even when the number of dimensions is greater than three. This may be done by defining each definite ordered collection of n real numbers $(x_1,x_2,x_3,\ldots,x_n)$ to be a point in the n-dimensional space S_n. It should be noted that the entire collection defines the point, *not* the single numbers. The individual numbers x_1 or x_2, etc. are called the coordinates of the point x. The space S_n itself is, then, the totality of all such points.

It is often convenient to divide the space into certain regions or intervals. This is usually done by considering the set of points for which the coordinates lie within definite boundaries. Thus, by a closed interval in S_n we will understand the set of points $x = (x_1,x_2,\ldots,x_n)$ for which the coordinates satisfy the relations

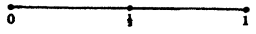
$$a_i \leq x_i \leq b_i \qquad i = 1, 2, 3, \ldots, n$$

An open interval is composed of the set of points for which the coordinates satisfy the following relations.

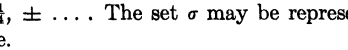
$$a_i < x_i < b_i \qquad i = 1, 2, 3, \ldots, n$$

In this case the coordinates are never allowed to assume the boundary values a_i or b_i.

It is sometimes necessary to consider the neighborhood of a point in the space S_n. Precisely, this means any open interval containing the point in question. If *every* neighborhood of a point x contains at least one point different from x which belongs to a set σ, then x is said to be a limit point of the set σ. As an example of a limit point of a one-dimensional set, let $x = \frac{1}{2}$, and let the set σ be all positive real numbers between 0 and 1. The set σ may be represented by the straight line:



Then every open interval containing the point $x = \frac{1}{2}$ also contains other points of the set; consequently $x = \frac{1}{2}$ is a limit point of the set and in this case the limit point belongs to the set. It is readily seen that any point on the line between 0 and 1 is a limit point of the set. For another example, let $x = 0$, and let the set σ be composed of the real numbers ± 1, $\pm\frac{1}{2}$, $\pm\frac{1}{3}$, $\pm\frac{1}{4}$, $\pm \ldots$. The set σ may be represented by discrete points along a line.



It is readily seen that any open interval or neighborhood containing the point $x = 0$ contains at least one of the numbers $1/n$ of the set,

so that $x = 0$ is a limit point of the set. It is also clear that no other point fulfills this condition, so that there are no other limit points to the set under consideration. In this case the limit point $x = 0$ does not belong to the set σ.

If a point set lies entirely within a certain interval it is said to be bounded. It is proven in point-set theory that every bounded infinite point set possesses at least one limit point. The second example given above is a good illustration of this theorem.

3. Function of a single variable. When a correspondence is established in any manner between points in one set with those in another, this correspondence constitutes the definition of a function. If to each number or point x of a set S_1 there is associated one, and only one, number or point y in a set T_1, then y is called a single-valued function of x. In this case x is called the independent variable, and y the dependent variable, since x must be specified before y is determined. To indicate that y is a function of x, one writes $y = f(x)$ or $y(x)$. The set S_1 containing the points x is called the domain of definition of the function. There may be instances such that for every number x belonging to a subset S_1' of S_1, there is associated more than one number in T_1, then y is said to be a many-valued function of x over the domain S_1'.

If the set T_1 (which includes all values y the function may assume as x traverses the domain of definition) is bounded, the function is said to be bounded. It will be evident that if T_1 is bounded, the points in T_1 will be contained in some smallest interval. Let the smallest interval be defined by the numbers c and d, so that

$$c < y < d$$

Then c is called the greatest lower bound of the function, and d is the least upper bound. If, on the other hand, c and d belong to the set T_1, so that

$$c \leq y \leq d$$

then c is the minimum and d the maximum of the function as x traverses the set S_1. The nonnegative difference, $d - c = v$, is known as the variation of the function over the set S_1. If the points of the set S_1 are represented on a straight line, and the points of the set T_1 represented on another straight line intersecting the first at right angles, the function $y = f(x)$ may be represented by plotting the points of S_1 and T_1 which correspond as shown in Fig. 3.1.

In speaking of functions it is convenient to use terms which describe certain simple properties which they may have. A few of these may be given at once. In the first place a number $x = a$, for which $y = f(a) = 0$,

is called a zero point of the function, or merely $y = f(x)$ has the zero point a.

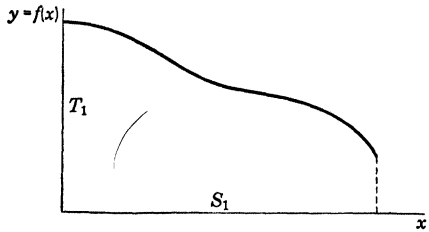


FIG. 3.1

A function is said to be periodic with the period $p \neq 0$, when $f(x + p) = f(x)$ is an identity in x, or merely that the equation is satisfied for all x in the domain of definition of $f(x)$. It should be noticed that along with p all integer multiples of p are also periods. For example, $3p$ is a period, since

$$f(x + 3p) = f(x + 2p + p) = f(x + 2p)$$
$$= f(x + p + p) = f(x + p) = f(x)$$

The smallest period is called the primitive period.

A function $f(x)$ is said to be an even function when $f(-x) = f(x)$ for all x. An example of this is the function $\cos x$, since $\cos(-x) = \cos x$. The graph of the function looks the same for negative values of x as for positive values. If, on the other hand, $f(-x) = -f(x)$, the function is called an *odd* function. An example of this is the function $\sin x$, since $\sin(-x) = -\sin x$. The graph of the function for negative x is the same as for positive x, except that it is reflected across the x axis.

If $y = f(x)$ is a single-valued function of x in the interval of definition $a \leq x \leq b$, and if x_1 and x_2 are any two points in the interval such that $x_1 < x_2$, then $f(x)$ is said to be a monotonically increasing function when $f(x_1) \leq f(x_2)$, and monotonically decreasing when $f(x_1) \geq f(x_2)$.

Thus far, x has been the independent variable, and $y = f(x)$ the dependent variable. In many cases it is possible to invert the role of x and y and let the variable y in the set T_1 be the independent variable, and let x be the dependent variable. In this case x becomes another function of y, which may be written $x = g(y)$. When the two functions $f(x)$ and $g(y)$ satisfy the equations

$$y = f[g(y)] \qquad \text{and} \qquad x = g[f(x)]$$

for all x and y in the domain of definition, one function is said to be the inverse of the other. It may be mentioned here that not all functions possess unique inverses in accordance with the definition above. No

unique inverse exists for $y = x^2$, since $x = +\sqrt{y}$ and $x = -\sqrt{y}$. For every continuous (see Sec. 5) and monotonic function there exists a unique inverse function which is continuous and monotonic throughout the interval of definition.

4. Limiting value of a function. In defining the limiting value of a function the absolute value of a quantity is used. By the absolute value of a quantity is meant the quantity taken with a positive sign regardless of the sign it actually has. The absolute value of a quantity q is indicated by the symbol $| q |$. Thus $| 2 - 3 | = 1$.

The function $f(x)$ has the limit value L for a point $x = x_0$ if for any positive number ϵ, however small, there exists a positive number δ (depending on ϵ) such that $| f(x) - L | < \epsilon$ when $0 < | x - x_0 | < \delta$. This means that when the difference between $f(x)$ and its limit value L is made arbitrarily small, x approaches sufficiently near to the point $x = x_0$. The point x_0 must therefore be a limit point of the set over which $f(x)$ is defined. The process of obtaining the limiting value of $f(x)$ is indicated by writing

$$\lim_{x \to x_0} f(x) = L$$

It should be pointed out that it is not necessary that the limiting value L at $x = x_0$ be the actual value of the function at $x = x_0$ although very commonly it is. As an example, one may consider the function

$$\begin{aligned} f(x) &= 2x \qquad \text{for } 0 \leq x < 1 \\ &= x - 1 \quad \text{for } 1 \leq x \leq 2 \end{aligned}$$

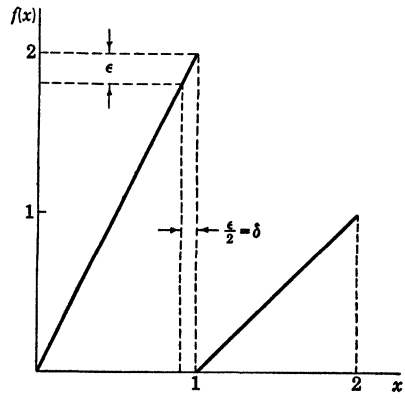


FIG. 4.1

The graph of this function is shown in Fig. 4.1. In the sense of the

above definition the function $f(x)$ does not have a limit at the point $x = 1$, since there exists no definite number L such that $|f(x) - L|$ is less than a prescribed arbitrarily small positive number ϵ for values of x inside an interval $\delta > 0$ of any length whatever, which includes the point $x = 1$. However, it is readily seen that as x approaches the value $x = 1$ from the left the limiting value of $f(x) = 2$, since on choosing an arbitrarily small positive number ϵ, and requiring that $x < 1$, then $|2x - 2| < \epsilon$ for values of x satisfying $0 < |x - 1| < \frac{1}{2}\epsilon$, so that a suitable value of δ exists, namely, $\frac{1}{2}\epsilon$ no matter how small ϵ is chosen. Although $L = 2$ is the limiting value of $f(x)$ as $x \to 1$ from the left this is not the actual value of $f(x)$ at $x = 1$ since the function is defined to be zero there. If x approaches $x = 1$ from the right, the limiting value of $f(x)$ is zero which coincides with the value of the function at that point.

As is indicated by the above example it is useful to speak of a left-hand or right-hand limit of a function at a given point. The limiting process as the point in question is approached from the left or right is indicated by

$$\lim_{x \to x_0 - 0} f(x) \qquad \text{or} \qquad \lim_{x \to x_0 + 0} f(x)$$

respectively.

If the point set S_1 over which $f(x)$ is defined is not bounded, it is often of interest to seek the value which $f(x)$ approaches when the value of x increases or decreases without limit. In the first case the limiting value of $f(x)$ is defined to be L, if for any positive number ϵ, however small there exists a positive number N (depending on ϵ) such that $|f(x) - L| < \epsilon$ when $x > N$. The limiting process for this case is written $\lim_{x \to \infty} f(x)$. The limit process when x decreases without limit is indicated by $\lim_{x \to -\infty} f(x)$.

A few useful theorems involving limits of functions are given below without proof.

(1) If $\lim_{x \to x_0} f(x) = L$, then $\lim_{x \to x_0} |f(x)| = |L|$

(2) If $\lim_{x \to x_0} f(x) = \mathrm{L}$ and $\lim_{x \to x_0} g(x) = M$ then

(a) $$\lim_{x \to x_0} f(x) \pm g(x) = \lim_{x \to x_0} f(x) \pm \lim_{x \to x_0} g(x)$$
$$= L \pm M$$

(b) $$\lim_{x \to x_0} f(x) \cdot g(x) = \lim_{x \to x_0} f(x) \cdot \lim_{x \to x_0} g(x)$$
$$= L \cdot M$$

and further if $M \neq 0$, then

$$(c)\ \lim_{x \to x_0} \frac{f(x)}{g(x)} = \frac{\lim_{x \to x_0} f(x)}{\lim_{x \to x_0} g(x)} = \frac{L}{M}$$

In certain cases the value of a function $f(x)$ may increase without limit as x approaches a point x_0. Nevertheless one speaks of the limiting value ∞ and writes $\lim_{x \to x_0} f(x) = \infty$. More precisely, if for a certain point x_0 and an arbitrarily large positive number N there exists a number δ (depending on N) for which $f(x) > N$ for all values of x in the domain $| x - x_0 | < \delta$, then $f(x)$ has the improper limiting value infinity (∞). An example of this type is furnished by the function $f(x) = 1/x^2$.

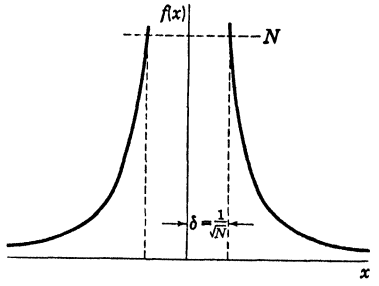


FIG. 4.2

A graph of this function is shown in Fig. 4.2. It is readily seen that, having chosen a large positive N, a number δ can be chosen so that $f(x) > N$ for all x in the interval of length δ containing the point $x = 0$. A suitable δ would be $1/\sqrt{N}$.

Certain convenient theorems involving improper limits are given without proof. If the

$$\lim_{x \to x_0} f(x) = \pm\infty \quad \text{and} \quad \lim_{x \to x_0} g(x) = L,$$

then

(1) $\lim_{x \to x_0} f(x) \pm g(x) = \pm\infty$

(2) $\lim_{x \to x_0} f(x) \cdot g(x) = \pm\infty$ or $\mp\infty$ according as the function $g(x)$ is greater than or less than zero in the neighborhood of the point $x = x_0$.

(3) $\lim_{x \to x_0} \frac{g(x)}{f(x)} = 0$

5. Continuity of a function. The concept of continuity of functions is one of the most important in analysis, and this property is very often required in order that certain manipulations with functions may be carried out. A function $y = f(x)$ is said to be continuous at a point $x = x_0$ of its domain of definition if the $\lim_{x \to x_0} f(x) = f(x_0)$. This means that the limiting value of the function at $x = x_0$ must be the actual value which the function has at that point. Referring again to Fig. 4.1, it is seen that the function graphed is not continuous at the point $x = 1$. It is continuous, however, for all other points in the interval $1 \leq x \leq 2$. A function which has a discontinuity of a different type is $f(x) = 1/x$, which fails to be continuous at the point $x = 0$.

Functions which are continuous at every point in an interval are said to be continuous throughout the interval. It is often necessary to place a greater restriction on a function than just continuity, namely, that of *uniform* continuity in an interval. A function is said to be uniformly continuous in an interval if, for a given arbitrarily small positive number ϵ, there exists *one* positive number $\delta > 0$ such that $| f(x_1) - f(x_2) | < \epsilon$ for *all* pairs of points x_1, x_2 in the interval of definition for which $0 < | x_1 - x_2 | < \delta$.

An example of a function which is continuous in the interval $0 < x \leq 1$ but not uniformly continuous is $f(x) = 1/x$. The point of discontinuity $x = 0$ has been excluded from the interval under consideration. That $f(x) = 1/x$ is not uniformly continuous in the interval may be seen by referring to Fig. 5.1. Choosing the value indicated on the graph for ϵ,

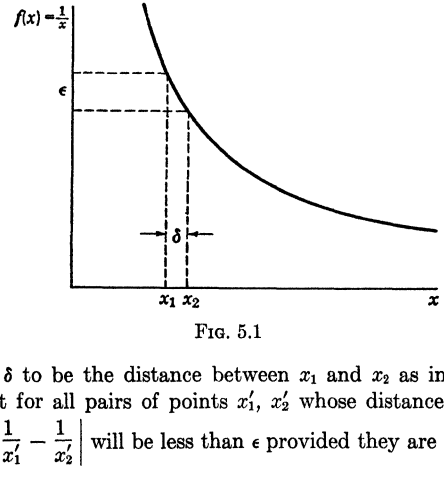


FIG. 5.1

and taking δ to be the distance between x_1 and x_2 as indicated, it is evident that for all pairs of points x_1', x_2' whose distance apart is less than δ the $\left| \frac{1}{x_1'} - \frac{1}{x_2'} \right|$ will be less than ϵ provided they are chosen to the

right of $x = x_1$. On the other hand, if these points are located to the left of $x = x_1$, it is obvious that the distance between them cannot be as large as δ if the difference $\left|\frac{1}{x_1'} - \frac{1}{x_2'}\right|$ is to be less than ϵ. Thus a smaller δ would have to be chosen, and the farther to the left the two points x_1', x_2' are taken, the smaller must δ be, and eventually δ must be vanishingly small, which the definition does not allow, since $\delta > 0$. The result of this argument may be stated more precisely by noting that for a given $\epsilon > 0$ it is always possible to find two numbers x_1, x_2 for which $\left|\frac{1}{x_1} - \frac{1}{x_2}\right| > \epsilon$ when $0 < |x_1 - x_2| < \delta$ no matter how small a value of δ is chosen. If one considers the function $f(x) = 1/x$ in the closed interval $0 < a \leq x \leq b$, it is uniformly continuous in this interval.

A function $f(x)$ defined in the interval $a \leq x \leq b$, and continuous only in the intervals $a \leq x < c_1$, $c_1 < x < c_2$, $c_2 < x < c_3$, $\ldots$, $c_{n-1} < x \leq b$ is called a piecewise continuous function in the interval ab.

A few important facts concerning continuous functions are stated below.

(1) A function continuous in any closed interval is bounded.

(2) The sums, differences, products, and quotients of continuous functions are again continuous functions provided in the case of the quotients the zero points of the denominator are excluded.

(3) The absolute value of a continuous function is again continuous although the converse is not always true.

(4) A function which is continuous in any closed interval has one maximum and one minimum value.

(5) If $u = g(x)$ is continuous at the point $x = x_0$ and $y = f(u)$ is also continuous at $u = u_0 = g(x_0)$, the function $y = f[g(x)]$ is continuous at the point $x = x_0$, and the

$$\lim_{x \to x_0} f[g(x)] = \lim_{u \to u_0} f(u).$$

6. Limit and continuity of a function of several variables. The concept of a function of more than one variable is similar to that given in Sec. 3 for a function of one variable. The essential difference is that in establishing a correspondence between the elements of the two sets S_n and T, the elements of the set S_n are collections of numbers which may be denoted by $(x_1, x_2, \ldots, x_n)$. In other words, the elements of the set S_n of definition of the function are points in the n-dimensional space S_n as defined in Sec. 2. Thus in a function of n variables, indicated by writing

$$u = f(x_1, x_2, \ldots, x_n)$$

n numbers must be specified before the number u is determined. Here $x_1, x_2, \ldots, x_n$ are called the independent variables.

For simplicity, only functions of two independent variables will be referred to, since the extension of the discussion to functions of more than two variables is straightforward. For convenience let the independent variables be denoted by x and y, and consider the function $u = f(x,y)$ which is defined for all values of x and y lying in a given region of the xy plane of a Cartesian coordinate system. Then in some cases the values of the function may be plotted in the z direction, in which case the function $u = f(x,y)$ would be represented by a surface above or below the region of definition S_2, as shown in Fig. 6.1.

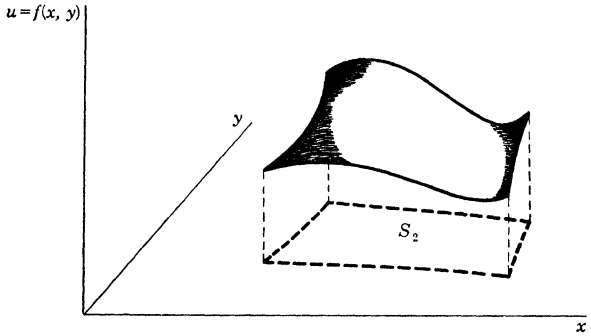


FIG. 6.1

The function $f(x,y)$ has the limiting value L at a point (x_0,y_0) if, for any positive number ϵ however small, there exists a number $\delta > 0$ such that $|f(x,y) - L| < \epsilon$ when the inequalities $0 < |x - x_0| < \delta$ and $0 < |y - y_0| < \delta$ are satisfied simultaneously. In this type of limit only values of x and y in the immediate neighborhood of (x_0,y_0) are involved. The process of finding this limit is indicated by writing

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = L \tag{6.1}$$

Care must be taken to distinguish between the simultaneous type of limit defined above and the limit obtained by a successive process which will now be defined. If, in the function $f(x,y)$ one of the variables, x for instance, is held fixed, one may try to find the $\lim_{y\to y_0} f(x,y)$. If this limit exists for different values of x, it defines a function of x which may be denoted by $\varphi(x)$. Then the limit of $\varphi(x)$, as $x \to x_0$, if such exists, is written

$$\lim_{x\to x_0} \varphi(x) = \lim_{x\to x_0} [\lim_{y\to y_0} f(x,y)]$$
$$= \lim_{x\to x_0} \lim_{y\to y_0} f(x,y) = L_1 \tag{6.2}$$

In a similar way, the variable y could be held fixed while the limit of $f(x,y)$ as $x \to x_0$ is taken, thus defining a function $\psi(y)$. Then the limit of $\psi(y)$ as $y \to y_0$ would be written

$$\lim_{y\to y_0} \psi(y) = \lim_{y\to y_0} [\lim_{x\to x_0} f(x,y)]$$
$$= \lim_{y\to y_0} \lim_{x\to x_0} f(x,y) = L_2 \tag{6.3}$$

If both repeated limits L_1 and L_2 exist, they need not be equal. Even if L_1 and L_2 exist and are equal, it does not follow that L exists. If L and L_1 exist, then $L = L_1$, or if L and L_2 exist, then $L = L_2$.

As an illustration of some of the above statements consider the function

$$f(x,y) = \frac{x^2 - y^2 + 2x^3 + y^3}{x^2 + y^2}$$

which is defined for all values of x and y except $x = y = 0$. The two repeated limits at $x = 0$ and $y = 0$ are

$$\lim_{x\to 0} \lim_{y\to 0} f(x,y) = \lim_{x\to 0} (1 + 2x) = 1$$
$$\lim_{y\to 0} \lim_{x\to 0} f(x,y) = \lim_{y\to 0} (-1 + y) = -1$$

Here the limits L_1 and L_2 are found to exist but are not equal. It will be shown that the limit L defined in (6.1) does not exist at the point $x = y = 0$. To do this it is necessary to demonstrate that given a positive number $\epsilon > 0$ there exist no number L and positive number δ for which

$$\left| \frac{x^2 - y^2 + 2x^3 + y^3}{x^2 + y^2} - L \right| < \epsilon \tag{6.4}$$

when $0 < |x - 0| < \delta$ and $0 < |y - 0| < \delta$. The inequality above may be written in the two ways:

$$\left| \frac{1 - y^2/x^2 + x(2 + y^3/x^3)}{1 + y^2/x^2} - L \right| < \epsilon \tag{6.4a}$$

$$\left| \frac{x^2/y^2 - 1 + y(2x^3/y^3 + 1)}{1 + x^2/y^2} - L \right| < \epsilon \tag{6.4b}$$

It is clear that there are values of x and y satisfying the inequalities $0 < x < \delta$ and $0 < y < \delta$ and at the same time satisfying (6.4a) when $L = 1$, no matter what choice of δ is made. On the other hand, there will also be values of x and y satisfying the same inequalities for which (6.4b) will be fulfilled when $L = -1$. Since the inequality (6.4) must

be satisfied for all values of x and y in the region $0 < x < \delta$ and $0 < y < \delta$, it follows that L does not exist at $x = y = 0$.

A function $f(x,y)$ is said to be continuous at a point (x_0,y_0) of its domain of definition provided

$$\lim_{x,y \to x_0,y_0} f(x,y) = f(x_0,y_0)$$

This means that the limiting value of the function as $(x,y \to x_0,y_0)$ must coincide with the defined value of the function at the point x_0,y_0. All of the concepts and theorems given in Sec. 5 also apply to functions of two or more variables.

7. Classes of functions. Functions may be classified according to their general properties or according to the manner in which they are defined. Thus all piecewise continuous functions in a certain domain would form a class, or all differentiable functions defined in a certain interval. Classes which arise from the method of definition include entire rational functions, rational functions, algebraic functions, etc. In the latter category an important class is that of *entire rational functions*, or *polynomials*. In the case of one variable such functions may always be written in the form

$$\sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + \ldots + a_n x^n$$

where the a's are constants independent of x. In the case of r independent variables this class of functions assumes the form

$$\sum_{p_1 p_2 \ldots p_r = 0}^{n_1 n_2 \ldots n_r} a_{p_1 p_2 \ldots p_r} x_1^{p_1} x_2^{p_2} \ldots x_r^{p_r}$$

where $p_1, p_2, \ldots, p_r$ assume positive integer values up to $n_1, n_2, \ldots, n_r$. The greatest value of the sum $p_1 + p_2 + \ldots + p_r$ appearing in the above sum is called the degree of the polynomial. In the particular case when the sum of the exponents in each term is the same, the polynomial is said to be homogeneous. Homogeneous polynomials are usually called forms. The sum of the exponents is the degree of the form. When the degree of the form is 2, one speaks of a quadratic form. The theory of quadratic forms is of considerable importance in applications, and is referred to again in Sec. 139.

Rational functions are those which can be expressed as the quotient of two polynomials. If the degree of the polynomial in the numerator is less than the degree of that in the denominator, the function is called a proper rational function; otherwise it is improper.

Algebraic functions can in general be defined implicitly only by equat-

ing a polynomial to zero. Rational functions and polynomials are special cases of algebraic functions.

The name *transcendental* is given to all functions which are not algebraic. The elementary transcendental functions are x^a, where a is an irrational number, b^x, x^x, and logarithmic, trigonometric, and hyperbolic functions or combinations of these.

SUGGESTED REFERENCES

E. W. Hobson, *The Theory of Functions of a Real Variable*, London, Cambridge University Press, 1927.

E. J. P. Goursat and E. R. Hedrick, *Mathematical Analysis*, Boston, Ginn & Company, 1904.

Handbuch der Physik Vol. 3, Berlin, Julius Springer, 1928.

E. T. Whittaker and G. N. Watson, *Modern Analysis*, London, Cambridge University Press, 1920.

G. A. Gibson, *Advanced Calculus*, London, Macmillan & Co., Ltd., 1931.

PROBLEMS

4.1. Investigate the limits of the function

$$y = \frac{x}{|x|} e^x \quad \text{at} \quad x = 0$$

4.2. What is the limit of $y = \sin(1/x)$ at $x = 0$?

4.3. Investigate the limits of $\log x$ and $(\sin x)/x$ as $x \to \infty$.

5.1. Determine the domain for which $y = \tan x$ is (1) continuous, (2) uniformly continuous.

5.2. If the value of $y = (\sin x)/x$ is defined to be unity at $x = 0$, is y: (1) a continuous function of x for all x; (2) a uniformly continuous function for all x?

6.1. If $f(x,y) = xy/(x+y)^2$ for all x and y except $x = y = 0$, show that

$$\lim_{x\to 0} \lim_{y\to 0} f(x,y) = \lim_{y\to 0} \lim_{x\to 0} f(x,y) = 0,$$

but that $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist.

6.2. If $f(x,y) = x \sin(1/y)$ for all x, y, except $y = 0$, investigate the limits:

$$L_1 = \lim_{y\to 0} \lim_{x\to 0} f(x,y)$$

$$L_2 = \lim_{x\to 0} \lim_{y\to 0} f(x,y)$$

$$L_3 = \lim_{(x,y)\to(0,0)} f(x,y)$$

6.3. Given $f(x,y) = \begin{cases} -x \text{ for } y \geqslant 0 \\ +x \text{ for } y < 0 \end{cases}$. Determine the domain of continuity of $f(x,y)$.

Chapter 2

DIFFERENTIAL CALCULUS

DERIVATIVES AND DIFFERENTIALS

8. Derivative and differential of functions of one variable. Let $y = f(x)$ be a continuous function in a given interval, and let x_0 and x_1 be two arbitrary points of the interval, and $y_0 = f(x_0)$ and $y_1 = f(x_1)$ be the values of the function at these points. The following quotient is known as the difference quotient

$$\frac{y_1 - y_0}{x_1 - x_0} = \frac{\Delta y}{\Delta x} = \frac{\Delta f(x)}{\Delta x}$$

Geometrically this quantity is the slope of the straight line P_1P_0 in Fig. 8.1, or the mean increase in $f(x)$ in the interval (x_0,x_1). If the dif-

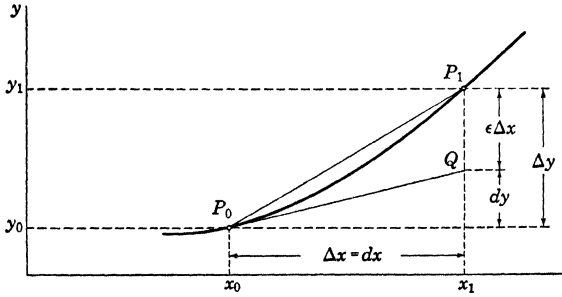


FIG. 8.1

ference quotient as defined above has a limit as $x_1 \to x_0$ or as $\Delta x \to 0$, that limit is called the derivative of $f(x)$ at the point x_0, or sometimes the differential quotient. One finds this limiting process indicated in any of the following ways.

$$\lim_{x_1 \to x_0} \frac{y_1 - y_0}{x_1 - x_0} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}$$
$$= \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$
$$= y' = f'(x) = \frac{dy}{dx}$$

Although not indicated in the notation, the value of the limit refers to a certain point, in this case $x = x_0$. When it is important to indicate the point at which the limit is taken, the notation is written in the modified form $y'(x_0)$, $f'(x_0)$, or $\left.\frac{dy}{dx}\right\}_{x=x_0}$. The limit dy/dx of the difference quotient also has a geometrical significance, namely, the slope of the tangent line P_0Q, to the curve at the point $x = x_0$.

The quantities dx and dy are called the differentials of the independent and dependent variable, respectively. Since the two differentials have not the same meaning, it is extremely important to understand clearly the difference between them. In the first place, the differential dx of the independent variable merely denotes an arbitrary interval and is consequently not dependent on the value of x. On the other hand, the value of the differential dy depends on the value of x for which the limit dy/dx is taken, as well as the differential dx, and is completely specified by the expression $dy = f'(x)dx$ when a point $x = x_0$ and the interval dx are chosen. Since dx is arbitrary, it may be taken equal to Δx, but dy is not in general equal to Δy. This may be clearly shown by writing

$$\Delta y - dy = \epsilon\, dx = \epsilon\, \Delta x$$

The quantities Δy, dy, dx are shown in Fig. 8.1. If the above relation is divided by dx, an expression for the value of ϵ results.

$$\epsilon = \frac{\Delta y}{\Delta x} - \frac{dy}{dx} = \frac{\Delta y}{\Delta x} - y'$$

Referring to the figure, it is clear that the quantity on the right is not zero for $\Delta x = dx > 0$ and consequently $\Delta y \neq dy$. The limiting value of ϵ as $\Delta x \to 0$ is found to be

$$\lim_{\Delta x \to 0} \epsilon = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} - y' = 0$$

from which it is to be concluded that the correction term $\epsilon\, \Delta x$ vanishes faster than Δx, so that for small values of Δx or dx the value of dy furnishes an approximate value for Δy, i.e.,

$$\Delta y \cong dy \quad \text{or} \quad f(x) - f(x_0) \cong f'(x_0)dx$$

The last equation indicates that an approximate value for the function

$f(x)$ can be found in terms of its value at the point x_0 and the value of its first derivative there.

It should be pointed out in this connection that if we have to do with a graph of a function $y = f(x)$, the point (x_0,y_0) lies on the curve $y = f(x)$, but the point $(x_0 + dx, y_0 + dy)$ does *not* lie on the curve but denotes a point on the tangent to the curve at the point (x_0,y_0).

The function $f(x)$ is said to be differentiable at a point (or in an interval) if it possesses a derivative at this point (or at every point in this interval). Every differentiable function is continuous, but the converse is not necessarily true.

The actual formulas for the derivatives of simple functions are well known and need not be repeated here.

9. Change of variable and higher order derivatives. It often happens that a function is not expressed in terms of the independent variable directly, but is expressed in terms of an intermediate variable, so that $u = f(y)$, where $y = g(x)$, and x is the independent variable. One may obtain u as a function of x by writing $u = f[g(x)] = F(x)$, but the value of du/dx can be found without computing $F(x)$ by making use of the formula for change of variable. This formula is

$$\frac{du}{dx} = \frac{du}{dy}\frac{dy}{dx} = f'(y)g'(x) \tag{9.1}$$

More generally, if $u = f_1(y_1)$, $y_1 = f_2(y_2)$, $\ldots$, $y_{n-1} = f_n(y_n)$, and $y_n = g(x)$

then
$$\frac{du}{dx} = \frac{du}{dy_1}\frac{dy_1}{dy_2}\frac{dy_2}{dy_3}\cdots\frac{dy_n}{dx} \tag{9.2}$$

If $y = f(x)$ and $x = g(y)$ are the inverse (see Sec. 3) of one another,

$$\frac{dx}{dy} = g'(y) = \frac{1}{dy/dx} = \frac{1}{f'(x)} \tag{9.3}$$

Again if the function $y = f(x)$ is given in the so-called parametric form so that $x = x(t)$ and $y = y(t)$, then

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{y'(t)}{x'(t)} \tag{9.4}$$

The formal computation of higher order derivatives of simple functions can be carried out by successive application of the above formulas. The nth derivative of the product of two functions can be conveniently computed by means of the Leibnitz formula

$$(u\cdot v)^{(n)} = \binom{n}{0}u^{(n)}v + \binom{n}{1}u^{(n-1)}v^{(1)} + \cdots\binom{n}{n-1}u^{(1)}v^{(n-1)} + \binom{n}{n}uv^{(n)} \tag{9.5}$$

where $\binom{n}{r}$ denotes the rth binomial coefficient

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \tag{9.6}$$

Where higher order derivatives are required, when the function is not a function of the independent variable directly, some care is required to keep in mind of what variable the different functions and derivatives are functions. Thus, if $u = f(y)$ and $y = g(x)$, then

$$\frac{d^2u}{dx^2} = \frac{d}{dx}\left[\frac{du}{dy}\frac{dy}{dx}\right] = \frac{d^2u}{dy^2}\left(\frac{dy}{dx}\right)^2 + \frac{du}{dy}\frac{d^2y}{dx^2} \tag{9.7}$$

No difficulty is encountered when it is recognized that in the above bracket du/dy is a function of y, while dy/dx is a function of x.

When one is dealing with inverse functions $x = g(y)$ and $y = f(x)$, the higher order derivatives of x may be found as follows.

$$\frac{d^2x}{dy^2} = \frac{d}{dy}\frac{1}{dy/dx} = \frac{d}{dx}\left[\frac{1}{dy/dx}\right]\frac{dx}{dy} = \frac{-d^2y/dx^2}{(dy/dx)^3} \tag{9.8}$$

When the function $y = f(x)$ is given in the parametric form $y = y(t)$ and $x = x(t)$, a similar method may be used to compute d^2y/dx^2.

$$\begin{aligned}\frac{d^2y}{dx^2} &= \frac{d}{dx}\left(\frac{y'(t)}{x'(t)}\right) = \frac{d}{dt}\left(\frac{y'}{x'}\right)\frac{dt}{dx} \\ &= \frac{y''x' - y'x''}{x'^2} \cdot \frac{1}{x'}\end{aligned} \tag{9.9}$$

where the prime indicates differentiation with respect to t.

10. Mean value theorems. It is shown in elementary calculus that if a function $f(x)$ is differentiable throughout the open interval $a < x < b$, there exists at least one point $x = \xi$ lying between a and b, for which

$$f(b) - f(a) = (b - a)f'(\xi) \tag{10.1}$$

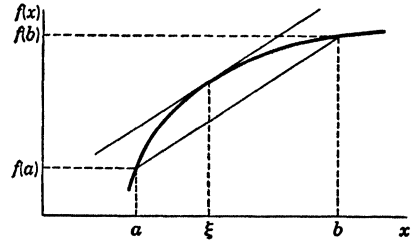


FIG. 10.1

The meaning of this law can be made clear by examining its geometrical significance. Since $f(x)$ possesses a derivative in the open interval, it is

continuous and can be represented by some curve as shown in Fig. 10.1. There will be at least one point $x = \xi$ between a and b such that the slope of the tangent to the curve at that point will be the same as that of the straight line joining $f(a)$ and $f(b)$, which fact is exactly what eq. 10.1 expresses. Taking the length of the interval $b - a$ as h, and denoting the point a by x_0 the law of the mean may be written in a slightly different form, namely,

$$f(x_0+h) - f(x_0) = hf'(x_0+\theta h) \qquad 0 < \theta < 1 \qquad (10.1a)$$

The above theorem, usually referred to as the first mean value theorem, can be generalized in the following way. Let $y = G(x)$ be a continuous and monotonic function in the closed interval $a \leq x \leq b$, which in addition possesses a nonvanishing derivative in the open interval $a < x < b$, and let a function $F(x) = f(y) = f\{G(x)\}$ be defined in the above closed interval, which also possesses a derivative in the open interval. We may take y as the independent variable in the interval $y_a \leq y \leq y_b$, where $y_a = G(a)$ and $y_b = G(b)$, and apply the mean value theorem (10.1) to $f(y)$. This gives

$$f(y_b) - f(y_a) = (y_b - y_a)f'(\zeta) \qquad y_a < \zeta < y_b \qquad (10.2)$$

Since, according to formula (9.1)

$$f'(y) = \frac{df}{dx}\frac{dx}{dy} = \frac{df/dx}{dy/dx} = \frac{F'(x)}{G'(x)}$$

eq. (10.2) may be written

$$\frac{F(b) - F(a)}{G(b) - G(a)} = \frac{F'(\xi)}{G'(\xi)} \qquad a < \xi < b \qquad (10.3)$$

This is called the generalized theorem of the mean. It may be written in a form to correspond to eq. (10.1a).

$$\frac{F(x_0+h) - F(x_0)}{G(x_0+h) - G(x_0)} = \frac{F'(x_0+\theta h)}{G'(x_0+\theta h)} \qquad 0 < \theta < 1 \qquad (10.4)$$

11. Evaluation of indeterminate forms. In practice one often has to deal with a function which may be written as the quotient of two functions, i.e., $F(x) = f(x)/g(x)$, which at a certain point $x = x_0$ fails to have a definite value. Planck's radiation law, which is of the form $Cv^3/(e^{av} - 1)$, is an illustration. This function has a definite value for every value of v except $v = 0$ at which point it gives merely 0/0. The function has a definite limiting value, however, as $v \to 0$. It is the object of the following discussion to provide a method of determining the limiting value of $F(x)$ for several important cases.

(A) $F(x) = \dfrac{f(x)}{g(x)}$ and $F(x_0) = \dfrac{0}{0}$, i.e., $f(x_0) = g(x_0) = 0$

If $f(x)$ and $g(x)$ satisfy the conditions required for the generalized law of the mean, eq. (10.4) applied to $f(x)$ and $g(x)$ yields

$$\frac{f(x_0+h)}{g(x_0+h)} = \frac{f'(x_0+\theta h)}{g'(x_0+\theta h)}$$

If the limit is taken as $h \to 0$, this is equivalent to writing

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{f'(x)}{g'(x)} \tag{11.1}$$

Consequently, if $g'(x) \neq 0$, then

$$\lim_{x\to x_0} F(x) = \lim_{x\to x_0} \frac{f'(x)}{g'(x)}$$

On the other hand, if $g'(x_0) = 0$ while $f'(x_0) \neq 0$, then the $\lim_{x\to x_0} F(x) = \pm\infty$.

It may happen that $f'(x_0) = g'(x_0) = 0$, in which case the above process may be repeated, giving

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{f'(x)}{g'(x)} = \lim_{x\to x_0} \frac{f''(x_0)}{g''(x_0)}$$

provided $f'(x)$ and $g'(x)$ satisfy the conditions specified for $f(x)$ and $g(x)$ in Sec. 10.

As an illustration of the method explained above, let it be required to find $\lim_{x\to x_0} (\sin x)/x$. Since the above conditions are satisfied,

$$\lim_{x\to 0} \frac{\sin x}{x} = \lim_{x\to 0} \frac{\cos x}{1} = 1$$

(B) $F(x) = \dfrac{f(x)}{g(x)}$; $F(x_0) = \dfrac{\infty}{\infty}$, i.e., $\lim_{x\to x_0} f(x) = \lim_{x\to x_0} g(x) = \infty$

where x_0 may be finite or infinite. Taking the case where x_0 is finite, eq. (10.4) is written for a value of x different from x_0.

$$\frac{f(x+h) - f(x)}{g(x+h) - g(x)} = \frac{f'(x+\theta h)}{g'(x+\theta h)}$$

This may be rewritten in the form

$$\frac{f(x)}{g(x)} = \frac{f(x)}{g(x)}\left[\frac{g(x+h) - g(x)}{f(x+h) - f(x)}\right]\frac{f'(x+\theta h)}{g'(x+\theta h)}$$

or

$$\frac{f(x)}{g(x)} = \left[\frac{1 - \dfrac{g(x+h)}{g(x)}}{1 - \dfrac{f(x+h)}{f(x)}}\right]\frac{f'(x+\theta h)}{g'(x+\theta h)}$$

Since $\lim_{x\to x_0} |\, g(x)\,| = \lim_{x\to x_0} f(x) = \infty$, then x and h may be made to approach x_0 and zero, respectively, in such a way that

$$\lim_{h\to 0, x\to x_0} \frac{g(x+h)}{g(x)} = 0 \quad \text{and} \quad \lim_{h\to 0, x\to x_0} \frac{f(x+h)}{f(x)} = 0$$

and consequently

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{f'(x)}{g'(x)}$$

It may readily be shown that the same relation holds if $x_0 = \infty$. As in case (A), if $f'(x_0)/g'(x_0) = \infty/\infty$, the process of differentiation may be repeated. As an illustration let it be required to find $\lim_{x\to\infty} (\log x)/x$.

$$\lim_{x\to\infty} \frac{\log x}{x} = \lim_{x\to\infty} \frac{1/x}{1} = 0$$

(C) $F(x) = f(x)\cdot g(x)$; $F(x_0) = 0\cdot\infty$, i.e., $f(x_0) = 0$ and $\lim_{x\to x_0} | g(x) | = \infty$.

In this case $F(x)$ may be written as either

$$F(x) = \frac{f(x)}{\dfrac{1}{g(x)}} \qquad \text{or} \qquad F(x) = \frac{g(x)}{\dfrac{1}{f(x)}}$$

which are the forms treated under cases (A) and (B), respectively. As an illustration take the $\lim_{x\to 0} x \log x$. Rewriting we require $\lim_{x\to 0} \dfrac{\log x}{1/x}$ which falls under case (B), and hence

$$\lim_{x\to 0} x \log x = \lim_{x\to 0} \frac{1/x}{-1/x^2} = \lim_{x\to 0} (-x) = 0$$

(D) $F(x) = f(x) - g(x)$;
$F(x_0) = \infty - \infty$, i.e., $\lim_{x\to x_0} f(x) = \lim_{x\to x_0} g(x) = \infty$

This situation is covered by case (A) on writing

$$F(x) = \frac{\dfrac{1}{g(x)} - \dfrac{1}{f(x)}}{\dfrac{1}{f(x)g(x)}}$$

As an illustration, let it be required to find the $\lim_{x\to\pi/2} \sec x - \tan x$. Rewriting as above,

$$\lim_{x\to\pi/2} \sec x - \tan x = \lim_{x\to\pi/2} \frac{\dfrac{1}{\tan x} - \dfrac{1}{\sec x}}{\dfrac{1}{\sec x \tan x}}$$

$$= \lim_{x \to \pi/2} \frac{\dfrac{\cos x}{\sin x} - \cos x}{\dfrac{\cos^2 x}{\sin x}}$$

$$= \lim_{x \to \pi/2} \frac{1 - \sin x}{\cos x}$$

$$= \lim_{x \to \pi/2} \frac{-\cos x}{-\sin x} = 0$$

(E) $F(x) = f(x)^{g(x)}$, where the substitution of $x = x_0$ may lead to the following three situations.

$$F(x_0) = 0^0, \quad \text{i.e., } \lim_{x \to x_0} f(x) = \lim_{x \to x_0} g(x) = 0$$

$$F(x_0) = \infty^0, \quad \text{i.e., } \lim_{x \to x_0} f(x) = \infty \quad \text{and} \quad \lim_{x \to x_0} g(x) = 0$$

$$F(x_0) = 1^{\pm\infty}, \quad \text{i.e., } f(x_0) = 1 \quad \text{and} \quad \lim_{x \to x_0} | g(x) | = \infty$$

In this case $F(x)$ may be written $F(x) = e^{g(x) \log f(x)}$ and the $\lim_{x \to x_0} g(x) \log f(x)$ may be examined as for case (C), whereupon

$$\lim_{x \to x_0} F(x) = e^{\lim_{x \to x_0} g(x) \log f(x)}$$

As an illustration let it be required to find the $\lim_{x \to 0} (1 + x)^{1/x}$. Rewriting as above,

$$\lim_{x \to 0} (1 + x)^{1/x} = e^{\lim_{x \to 0} (1/x) \log (1+x)}$$

then since $[\log (1 + x)]/x$ falls under case (A), the limit becomes

$$\lim_{x \to 0} \frac{\log (1 + x)}{x} = \lim_{x \to 0} \frac{1/(1 + x)}{1} = 1$$

and consequently $\lim_{x \to 0} (1 + x)^{1/x} = e^1 = e$

In evaluating limits of indeterminate forms as outlined above where the limit is not obtained after the first differentiation, care should be taken to see that the original conditions are fulfilled before differentiating a second time.

It is not always advisable to apply one of the above methods directly, since it is often easier to rewrite the function in its simplest form and proceed to the limit directly. As an illustration, let it be required to find

$$\lim_{x \to \infty} \sqrt{x^2 - 1} - x.$$

This case is covered under case (D) but it is easier to rewrite the function as

$$\lim_{x\to\infty} x\left[\sqrt{1-\frac{1}{x^2}}-1\right]$$

Then since $$\sqrt{1-\frac{1}{x^2}} = 1-\frac{1}{2}\cdot\frac{1}{x^2}+\frac{1}{8}\cdot\frac{1}{x^4}\cdots$$

the above limit can be found at once, since

$$\lim_{x\to\infty} x\left[-\frac{1}{2}\cdot\frac{1}{x^2}+\frac{1}{8}\cdot\frac{1}{x^4}+\text{higher powers of }\frac{1}{x}\right] = 0$$

12. Partial derivatives. Let $u = f(x,y)$ be a function defined in a region S_2 of the xy plane containing the point (x_0,y_0) and consider u as a function of x alone in the neighborhood of the point (x_0,y_0). This may be indicated by writing $u = f(x_0+h,y_0)$. If the limit

$$\lim_{h\to 0}\frac{f(x_0+h,y_0)-f(x_0,y_0)}{h}$$

exists, it is called the partial derivative of $f(x,y)$ with respect to x at the point (x_0,y_0), and is indicated by the symbol $\partial f/\partial x$. The partial derivative $\partial f/\partial y$ is defined in a similar manner, as are also the higher order partial derivatives $\partial^2 f/\partial y^2$, $\partial^2 f/\partial x\,\partial y$, $\partial^2 f/\partial y\,\partial x$, etc. If the limits indicated above exist at all points in an interval, the function is said to possess its first-order partial derivatives throughout the interval.

An important difference between the derivative du/dx of a function of one variable and the partial derivative $\partial u/\partial x$ of a function of more than one variable should be noted. In the former case du/dx may be formally expressed as the quotient of the dependent and independent differentials du and dx. In the latter case, there is no such meaning to be attached to what one might call the partial differentials ∂u and ∂x. The expression $\partial u/\partial x$ must be regarded as a homogeneous symbol denoting the limiting process referred to above.

Another important difference between the derivative of a function of one variable and partial derivatives, and their relations to the original functions should be noted. If in the case of a function of one variable the function has a derivative at a given point, the function is continuous there, but if a function $f(x_1,x_2,\ldots,x_n)$ possesses all its partial derivatives $\partial f/\partial x_i$ at a point, the function is not necessarily continuous there. If, however $f(x_1,x_2,\ldots,x_n)$ is uniquely defined at the point and if all its first-order partial derivatives in the *neighborhood* of the point exist and are bounded, $f(x_1,\ldots,x_n)$ is continuous at the point in question.

It is not always true that the mixed partial derivatives $\partial^2 f/(\partial x\,\partial y)$ and $\partial^2 f/(\partial y\,\partial x)$ are equal at a given point. If, however, the partial derivatives $\partial f/\partial x$, $\partial f/\partial y$, $\partial^2 f/(\partial x\,\partial y)$ exist in the neighborhood of a point, and if the latter is continuous there, then $\partial^2 f/(\partial x\,\partial y) = \partial^2 f/(\partial y\,\partial x)$.

13. Change of variable. Consider the function $u = f(y_1,y_2,\ldots,y_n)$ of the n variables $y_1, y_2, \ldots, y_n$, and let u be continuous, together with all its first-order partial derivatives throughout a region S_n. Then suppose that each of the variables y_i is in turn a function of the p variables $x_1, x_2, \ldots, x_p$. Thus

$$y_i = y_i(x_1,x_2,\ldots,x_p) \qquad i = 1, 2, \ldots, n$$

It will be supposed that these functions are continuous, together with their first-order partial derivatives, throughout a region S'_p, and furthermore each point in the space S'_p corresponds to a point in the space S_n. When all the above conditions are satisfied

$$\frac{\partial u}{\partial x_j} = \frac{\partial u}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \frac{\partial u}{\partial y_2}\frac{\partial y_2}{\partial x_j} + \ldots + \frac{\partial u}{\partial y_n}\frac{\partial y_n}{\partial x_j}; \qquad j = 1, 2, \ldots, p \tag{13.1}$$

As an example of the use of the above formula, let

$$u(x,y) = x^2 + y^2 \qquad \text{and} \qquad \begin{aligned} x &= r\cos\theta; \\ y &= r\sin\theta \end{aligned}$$

The equations defining x and y in terms of r and θ represent the usual transformation from polar coordinates to Cartesian coordinates. Suppose that $\partial u/\partial r$ and $\partial u/\partial\theta$ be desired. Applying the above formula,

$$\begin{aligned} \frac{\partial u}{\partial r} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial r} \\ &= 2x\cos\theta + 2y\sin\theta = 2r(\cos^2\theta + \sin^2\theta) = 2r \end{aligned}$$

This is the same result that would be obtained by substituting the expressions for x and y in $u(x,y)$ and then differentiating with respect to r. Similarly,

$$\begin{aligned} \frac{\partial u}{\partial\theta} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial\theta} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial\theta} = -2xr\sin\theta + 2yr\cos\theta \\ &= 2r^2(-\sin\theta\cos\theta + \sin\theta\cos\theta) = 0 \end{aligned}$$

This result can be checked directly by noting that $x^2 + y^2$ is independent of θ.

14. Directional derivative. The partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ considered above give a measure of the rate of change of the function in the neighborhood of a point (x_0,y_0) when leaving the point either in the x or y directions. It often happens, however, that one wishes to know the rate of change of the function as one leaves the point in some arbitrary direction. This rate of change can be formulated as follows: let a function $u = f(x,y)$ be defined in a region S_2 of the xy plane, and let a curve C pass through the points P_0 with coordinates (x_0,y_0) and P

as shown in Fig. 14.1. Let Δs be the length of arc between P_0 and P and form the difference quotient

$$\frac{u_p - u_0}{\Delta s} = \frac{\Delta u}{\Delta s}$$

The limit

$$\lim_{\Delta s \to 0} \frac{\Delta u}{\Delta s} = \frac{\partial u}{\partial s}$$

is called the directional derivative in the direction of the tangent to the curve C at the point P_0. The equations of the curve C make it possible to express x and y as functions of the single variable s. The directional

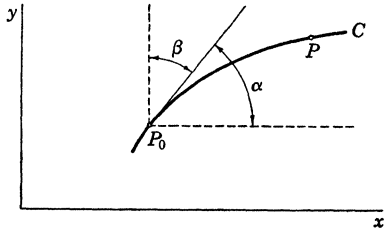


FIG. 14.1

derivative $\partial u/\partial s$ can then be expressed in terms of $\partial u/\partial x$ and $\partial u/\partial y$ by making use of the formula for change of variable given in the last section, thus

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{dx}{ds} + \frac{\partial u}{\partial y}\frac{dy}{ds}$$

but $dx/ds = \cos \alpha$, and $dy/ds = \cos \beta$, where $\cos \alpha$ and $\cos \beta$ are the direction cosines of the tangent to the curve C at the point P_0. Hence

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\cos \alpha + \frac{\partial u}{\partial y}\cos \beta + \frac{\partial u}{\partial z}\cos \gamma$$

where $\cos \alpha$, $\cos \beta$, and $\cos \gamma$ are the direction cosines of the tangent to the space curve at the point P_0.

An important special case of the directional derivative is of great use in physical problems. To obtain the special case referred to, let $u = f(x,y)$ be plotted above the xy plane. This will give a surface which we will imagine cut by equally spaced planes parallel to the xy plane. The intersections of the surface and the planes will be curves on which u will have a constant value. If these constant level curves are projected into the xy plane, a set of contour curves will result. Let P_0 be an arbitrary point on one of the contour curves as shown in Fig. 14.2. As seen above, the rate of change of the function $u = f(x,y)$ at the point P_0 depends on the direction one goes from P_0; in particular, the rate of change is zero if one goes in the direction of the contour

curve. We shall now examine the rates of increase of u in two directions, one normal to the contour curve C and one in any other direction s, which makes an angle ψ with the normal. The rate of change of u in the direction s is given by the directional derivative $\partial f/\partial s$, which

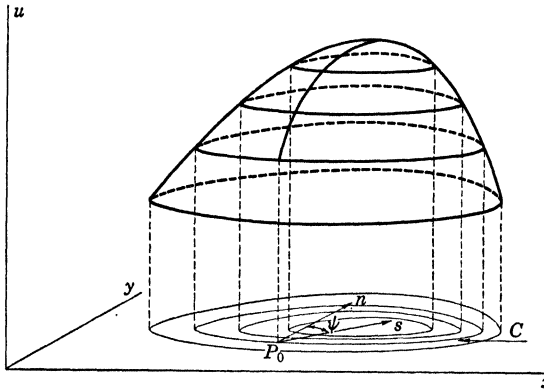


FIG. 14.2

may be expressed in terms of the directional derivative along the normal to C by using the formula for change of variable. Thus

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial n}\frac{dn}{ds}$$

but

$$\frac{dn}{ds} = \lim_{\Delta s \to 0} \frac{\Delta n}{\Delta s} = \cos \psi$$

so that

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial n} \cos \psi$$

Thus it is found that the derivative of f in any direction can be found by multiplying the derivative along the normal to a contour curve by the cosine of the angle between the normal and the arbitrary direction. It is clear that $\partial f/\partial s$ is a maximum when $\psi = 0$, so that the directional derivative along the normal gives the maximum rate of increase of the function. A curve drawn in the xy plane so that it intersects the contour lines at right angles provides the path of greatest rate of change of the function u. In applications the derivative $\partial f/\partial n$ is called the gradient of the function at the point P_0.

In mechanics and the theory of electricity it is often convenient to express the gravitational or electric force as the gradient of a function called the potential. To illustrate the use of the directional derivative in this connection, let there be a unit positive charge located at $x = -a$

and a unit negative charge at $x = a$ (see Fig. 14.3). The potential at any point in space due to these charges is $V = 1/r_1 - 1/r_2$, where r_1 and r_2 are the distances from the charges to the point in question. Let it be required to find the components of the electric force normal to the circle of radius r. As mentioned above, this force may be expressed as $F = -\partial V/\partial n$ or $-\partial V/\partial r$, since r is normal to the circle. Making use of the formula for change of variable the directional derivative becomes

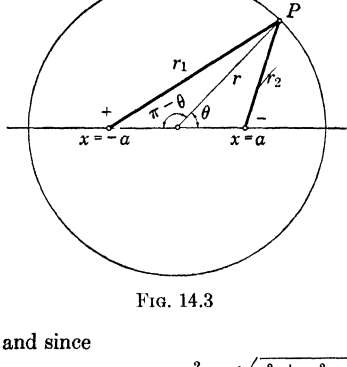


FIG. 14.3

$$\frac{\partial V}{\partial r} = \frac{\partial V}{\partial r_1}\frac{\partial r_1}{\partial r} + \frac{\partial V}{\partial r_2}\frac{\partial r_2}{\partial r}$$

and since

$$r_1 = \sqrt{r^2 + a^2 - 2ar\cos(\pi - \theta)}$$
$$= \sqrt{r^2 + a^2 + 2ar\cos\theta}$$
$$r_2 = \sqrt{r^2 + a^2 - 2ar\cos\theta}$$

then

$$\frac{\partial r_1}{\partial r} = \frac{r + a\cos\theta}{r_1} \quad \text{and} \quad \frac{\partial r_2}{\partial r} = \frac{r - a\cos\theta}{r_2}$$

and finally

$$\frac{\partial V}{\partial r} = -\frac{r + a\cos\theta}{r_1^3} + \frac{r - a\cos\theta}{r_2^3}$$

By replacing r_1 and r_2 by their expressions in terms of r and θ, we have found the normal component of the electric force at any point on the circle.

15. Total differential. Let $u = f(x,y)$ be a function defined in a region S_2 of the xy plane possessing continuous partial derivatives $\partial u/\partial x$ and $\partial u/\partial y$ in the region S_2. At the point (x_0,y_0) in S_2, $u = f(x_0,y_0)$ and at the point $(x_0+\Delta x, y_0+\Delta y)$ the value of u will be $f(x_0+\Delta x, y_0+\Delta y)$. The difference of these two values is called the *total difference* of u, and may be written

$$\Delta u = \Delta f(x,y) = f(x_0+\Delta x, y_0+\Delta y) - f(x_0,y_0)$$

Here Δx and Δy are arbitrary differentials of the independent variables and are not functions of them.

By adding and subtracting $f(x_0,y_0+\Delta y)$, the term Δu may be written in the form

$$\Delta u = f(x_0+\Delta x, y_0+\Delta y) - f(x_0, y_0+\Delta y) + f(x_0, y_0+\Delta y) - f(x_0, y_0)$$

Applying the law of the mean (Sec. 10) to the first two and last two terms, they may be written

$$f(x_0+\Delta x, y_0+\Delta y) - f(x_0, y_0+\Delta y) = \Delta x \frac{\partial f}{\partial x}(x_0+\theta\Delta x, y_0+\Delta y) \qquad 0 < \theta < 1$$

$$f(x_0, y_0+\Delta y) - f(x_0, y_0) = \Delta y \frac{\partial f}{\partial y}(x_0, y_0+\theta'\Delta y) \qquad 0 < \theta' < 1$$

The above expressions are still not convenient because the derivatives must be evaluated at an unknown point; hence the derivatives may be expressed at the point (x_0, y_0) by writing

$$\frac{\partial f}{\partial x}(x_0+\theta\Delta x, y_0+\Delta y) = \frac{\partial f}{\partial x}(x_0, y_0) + \alpha$$

$$\frac{\partial f}{\partial y}(x_0, y_0+\theta'\Delta y) = \frac{\partial f}{\partial y}(x_0, y_0) + \beta$$

where α and β are infinitesimals of a higher order than Δx and Δy since

$$\lim_{\Delta x, \Delta y \to 0, 0} \alpha = 0 \qquad \text{and} \qquad \lim_{\Delta x, \Delta y \to 0, 0} \beta = 0$$

Using the above expressions, the total difference can now be conveniently divided into two parts.

$$\Delta u = \left[\frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y\right] + \alpha\,\Delta x + \beta\,\Delta y$$

The quantity in brackets is called the principal part of the total difference Δu, or the *total differential* du of $f(x,y)$. The remaining terms constitute higher order infinitesimals. The total differential is then

$$du = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$$

This holds for all functions satisfying the conditions specified at the beginning of this section and in particular $u = x$ and $u = y$, in which cases $dx = \Delta x$ and $dy = \Delta y$. This indicates the arbitrary choice of the independent differentials. Henceforth the total differential will be written

$$du = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy$$

As in the case of a single variable, it should be noted that a point (x_0, y_0, u_0) lies on the surface representing the function while the point

(x_0+dx,y_0+dy,u_0+du) does not in general. In fact, perpendiculars erected to the xy plane at (x_0,y_0) and (x_0+dx,y_0+dy) intersect the surface at the points (x_0,y_0,u_0) and $(x_0+dx,y_0+dy,u_0+\Delta u)$, while they intersect the tangent plane to the surface at (x_0,y_0,u_0) in the points (x_0,y_0,u_0) and (x_0+dx,y_0+dy,u_0+du), as indicated in Fig. 15.1. This may be seen as follows. Let the distance from (x_0,y_0) to (x_0+dx,y_0+dy) be s. Then $du = (\partial f/\partial s)\,ds$, where $\partial f/\partial s$ is the slope of the curve produced by the intersection of a plane perpendicular to the xy plane containing the line s and the surface at the point (x_0,y_0,u_0). It is consequently the slope

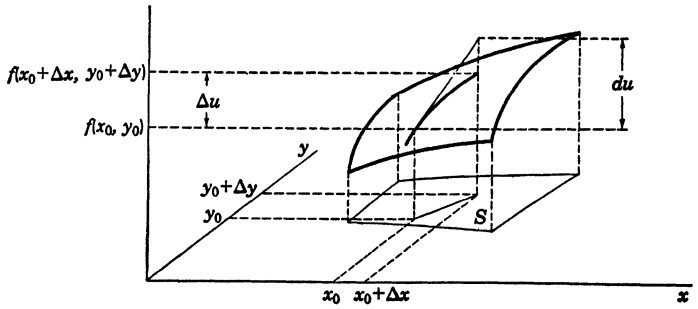


FIG. 15.1

of the tangent plane at this point in the direction of s. Therefore the perpendicular erected at (x_0+dx,y_0+dy) intersects the tangent plane at (x_0+dx,y_0+dy,u_0+du). But $\partial f/\partial s$ is the directional derivative at (x_0,y_0), i.e.,

$$du = \frac{\partial f}{\partial s}\,ds = \left(\frac{\partial f}{\partial x}\cos\alpha + \frac{\partial f}{\partial y}\cos\beta\right)ds = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy$$

In the case of functions of more than two variables the total differential becomes

$$du = \frac{\partial f}{\partial x_1}\,dx_1 + \frac{\partial f}{\partial x_2}\,dx_2 + \ \ldots\ + \frac{\partial f}{\partial x_n}\,dx_n$$

As in the case of a single variable, du may give a very good approximation to the total change Δu in the function when the independent differentials are small. Thus

$$\Delta u \cong du$$

In this connection, the total differential furnishes a convenient method of estimating the error produced in a certain quantity due to small inaccuracies in the measurement of other quantities upon which it depends. For example, the total surface area of a right circular cylinder

is $A = 2\pi r(r + a)$, where r is the radius and a is the length. If the possible error in measuring r is E_r and that in measuring a is E_a, the approximate error in A may be found from the total differential

$$dA = \frac{\partial A}{\partial r}\,dr + \frac{\partial A}{\partial a}\,da = (4\pi r + 2\pi a)dr + 2\pi r\,da$$

Substituting the possible errors for dr and da,

$$dA = 2\pi[(a + 2r)E_r + rE_a]$$

and the relative effects of the errors in r and a on the area are evident.

16. Differentiation of functions in parametric form. When the point (x,y,z) traverses a space curve, x, y, and z may be expressed in terms of one parameter or variable t, so that

$$x = x(t) \qquad y = y(t) \qquad z = z(t)$$

When the functions are given in this form, it is sometimes desirable to compute $\partial z/\partial x$, $\partial y/\partial x$ etc. This may be done by means of the differentials

$$dx = \frac{\partial x}{\partial t}\,dt, \qquad dy = \frac{\partial y}{\partial t}\,dt, \qquad dz = \frac{\partial z}{\partial t}\,dt$$

from which the differential dt can be eliminated yielding

$$\frac{dy}{dx} = \frac{\partial y/\partial t}{\partial x/\partial t}, \qquad \text{etc.} \tag{16.1}$$

A more complicated problem arises when the point (x,y,z) traverses a surface, i.e., when x, y, and z are expressed in terms of two parameters t and s; thus

$$x = x(t,s), \qquad y = y(t,s), \qquad z = z(t,s)$$

and the partial derivatives $\partial y/\partial x$, $\partial z/\partial x$, etc., are desired. The $\partial z/\partial x$ may be found by differentiating the last equation partially with respect to x, regarding t and s as functions of x; thus

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial t}\frac{\partial t}{\partial x} + \frac{\partial z}{\partial s}\frac{\partial s}{\partial x}$$

Since we do not have t and s expressed explicitly in terms of x, the partial derivatives $\partial t/\partial x$, $\partial s/\partial x$ cannot be evaluated directly. They may, nevertheless, be computed as follows. Differentiate the first and second equations with respect to x, and obtain

$$1 = \frac{\partial x}{\partial t}\frac{\partial t}{\partial x} + \frac{\partial x}{\partial s}\frac{\partial s}{\partial x} \qquad 0 = \frac{\partial y}{\partial t}\frac{\partial t}{\partial x} + \frac{\partial y}{\partial s}\frac{\partial s}{\partial x}$$

The values of $\partial t/\partial x$ and $\partial s/\partial x$ can be found from these equations by solving for these quantities. Thus

$$\frac{\partial t}{\partial x} = \frac{\dfrac{\partial y}{\partial s}}{\dfrac{\partial x}{\partial t}\dfrac{\partial y}{\partial s} - \dfrac{\partial x}{\partial s}\dfrac{\partial y}{\partial t}} \qquad \frac{\partial s}{\partial x} = \frac{\dfrac{\partial y}{\partial t}}{\dfrac{\partial x}{\partial t}\dfrac{\partial y}{\partial s} - \dfrac{\partial x}{\partial s}\dfrac{\partial y}{\partial t}}$$

The quantity appearing in the denominators can be expressed in the form of a functional determinant (see Sec. 129).

$$\begin{vmatrix} \dfrac{\partial x}{\partial t} & \dfrac{\partial x}{\partial s} \\ \dfrac{\partial y}{\partial t} & \dfrac{\partial y}{\partial s} \end{vmatrix} = \frac{\partial(x,y)}{\partial(t,s)}$$

This functional determinant is called the Jacobian determinant, and denoted by the notation given above. The formulas then become

$$\frac{\partial t}{\partial x} = \frac{\dfrac{\partial y}{\partial s}}{\dfrac{\partial(x,y)}{\partial(t,s)}} \qquad \frac{\partial s}{\partial x} = \frac{\dfrac{\partial y}{\partial t}}{\dfrac{\partial(x,y)}{\partial(t,s)}}$$

and

$$\frac{\partial z}{\partial x} = \frac{\dfrac{\partial z}{\partial t}\dfrac{\partial y}{\partial s} - \dfrac{\partial z}{\partial s}\dfrac{\partial y}{\partial t}}{\dfrac{\partial(x,y)}{\partial(t,s)}} = \frac{\dfrac{\partial(z,y)}{\partial(t,s)}}{\dfrac{\partial(x,y)}{\partial(t,s)}} = -\frac{\dfrac{\partial(y,z)}{\partial(t,s)}}{\dfrac{\partial(x,y)}{\partial(t,s)}}$$

Other partial derivatives such as $\partial y/\partial x$ can be computed in a like manner.

17. Differentiation of implicit functions. For simplicity, consider a function of two variables x, y which satisfies the identity $F(x,y) = 0$. This equation defines y as an implicit function of x. Usually this equation can be solved for y, and y then can be expressed as an explicit function $y = f(x)$, so that $F\{x,f(x)\} = 0$. It is not always possible to express y explicitly, however, as the following example will show. Let $F(x,y) = x^2 + y^2 + c^2 = 0$. It is at once obvious that no real function $y = f(x)$ can be given which, when squared, will make $F(x,y) = 0$. The conditions which must obtain in order for an explicit function $y = f(x)$ to exist in this case are that there exists a point (x_0,y_0) for which $F(x_0,y_0) = 0$ and $F(x,y)$, together with $\partial F/\partial x$ and $\partial F/\partial y$, exist and are continuous in the neighborhood of (x_0,y_0), and finally $\partial F(x_0,y_0)/\partial y \neq 0$.

If the equation $F(x,y) = 0$ does yield the explicit function $y = f(x)$, it often happens that one wishes to compute $\partial y/\partial x$ without actually finding the explicit relationship. This may be done by differentiating the equation $F(x,y) = 0$ partially with respect to x.

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial x} = 0 \qquad \text{or} \qquad \frac{\partial y}{\partial x} = -\frac{\partial F/\partial x}{\partial F/\partial y} \tag{17.1}$$

Another situation sometimes presents itself when $z = f(x,y)$ is a function of the two variables x and y, but an implicit relation $F(x,y) = 0$ connects x and y, and the value of $\partial z/\partial x$ is desired. This may be found by differentiating z with respect to x.

$$\frac{\partial z}{\partial x} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x}$$

Using the expression for $\partial y/\partial x$ derived above,

$$\frac{\partial z}{\partial x} = \frac{\partial f}{\partial x} - \frac{\partial f}{\partial y}\cdot\frac{\dfrac{\partial F}{\partial x}}{\dfrac{\partial F}{\partial y}} = \frac{\dfrac{\partial f}{\partial x}\dfrac{\partial F}{\partial y} - \dfrac{\partial f}{\partial y}\dfrac{\partial F}{\partial x}}{\dfrac{\partial F}{\partial y}} = \frac{\dfrac{\partial(f,F)}{\partial(x,y)}}{\dfrac{\partial F}{\partial y}} \tag{17.2}$$

If $F(x,y,z) = 0$, then z becomes an implicit function of the two variables x and y, and z may be expressed as an explicit function $z = f(x,y)$, provided $F(x,y,z)$ and $\partial F/\partial z$ are continuous in the neighborhood of the point (x_0,y_0,z_0), and if $F(x_0,y_0,z_0) = 0$ and $\partial F(x_0,y_0,z_0)/\partial x \neq 0$. Then in the neighborhood of (x_0,y_0) the following identity is satisfied.

$$F[x,y,f(x,y)] = 0$$

The partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$ could be computed by the method given above, but they may also be computed by using the total differential as shown below. Let $u = F(x,y,z) = 0$. Then the total differential of u is

$$du = \frac{\partial F}{\partial x}\,dx + \frac{\partial F}{\partial y}\,dy + \frac{\partial F}{\partial z}\,dz = 0$$

If an explicit function $z = f(x,y)$ exists in accordance with the above conditions, the total differential of z is

$$dz = \frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy$$

This expresses the dependent differential dz in terms of the independent differentials dx and dy. Using this expression in the preceding one, the differential of u is expressed in terms of only independent differentials; thus

$$du = \left(\frac{\partial F}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x}\right)dx + \left(\frac{\partial F}{\partial y} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial y}\right)dy = 0$$

Since dx and dy are independent, the above relation can hold only if their coefficients vanish separately, so that

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x} = 0 \qquad \text{and} \qquad \frac{\partial F}{\partial y} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial y} = 0 \tag{17.3}$$

From these equations the desired partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$ can be found.

Instead of just one equation $F(x,y,z) = 0$, which under the conditions mentioned above defines z as an explicit function of x and y, two simultaneous equations in four variables may be encountered. Consider the equations

$$F(x,y,u,v) = 0 \quad \text{and} \quad G(x,y,u,v) = 0$$

Under certain conditions it would be expected that these two equations would define both u and v as explicit functions of x, y. In order that these equations define the explicit functions $u = f(x,y)$ and $v = g(x,y)$, the functions F and G, together with their first-order partial derivatives, must be continuous in the neighborhood of a point (x_0,y_0,u_0,v_0) for which $F(x_0,y_0,u_0,v_0) = 0$ and $G(x_0,y_0,u_0,v_0) = 0$, and $\partial(F,G)/\partial(u,v) \neq 0$ at the point.

To compute $\partial u/\partial x$, differentiate both equations with respect to x.

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial F}{\partial v}\frac{\partial v}{\partial x} = 0$$

$$\frac{\partial G}{\partial x} + \frac{\partial G}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial G}{\partial v}\frac{\partial v}{\partial x} = 0$$

Solving these equations for $\partial u/\partial x$ and $\partial v/\partial x$, it is found that

$$\frac{\partial u}{\partial x} = -\frac{\partial(F,G)/\partial(x,v)}{\partial(F,G)/\partial(u,v)}, \qquad \frac{\partial v}{\partial x} = +\frac{\partial(F,G)/\partial(x,u)}{\partial(F,G)/\partial(u,v)} \tag{17.4}$$

The partial derivatives $\partial u/\partial y$, etc., can be computed in the same manner. The generalization to n equations defining a system of n functions offers no difficulty.

18. Dependent and independent functions. A set composed of p functions

$$\begin{aligned} y_1 &= f_1(x_1,x_2,\ldots,x_n) \\ y_2 &= f_2(x_1,x_2,\ldots,x_n) \\ &\cdots\cdots\cdots\cdots \\ y_p &= f_p(x_1,x_2,\ldots,x_n) \end{aligned} \tag{18.1}$$

each of which is a continuous function of n variables possessing continuous first-order partial derivatives, is said to be independent when there exists no relation between them which is zero for all values of $x_1, x_2, \ldots, x_n$ in the domain of definition. Expressed in mathematical form,

$$F(y_1,y_2,\ldots,y_p) \neq 0 \quad \text{for all} \quad x_1, \ldots, x_n$$

If a relation between the functions can be found which vanishes identically, the functions are dependent.

A criterion of independence can be given by considering the following array of all partial derivatives of the first order

$$M = \begin{Vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ \frac{\partial y_p}{\partial x_1} & \frac{\partial y_p}{\partial x_2} & \cdots & \frac{\partial y_p}{\partial x_n} \end{Vmatrix} \tag{18.2}$$

An array of quantities arranged in the manner shown is called a matrix (see Sec. 132). The matrix under consideration has p rows and n columns. In case $p = n$, the quantities in the matrix can be used to form a determinant which is just the Jacobian determinant

$$\frac{\partial(y_1, y_2, \ldots, y_n)}{\partial(x_1, x_2, \ldots, x_n)}$$

considered in the preceding section. The necessary and sufficient condition that all n of the functions y be independent is that the Jacobian determinant does not vanish identically.

When $n > p$, one must examine the determinants having p rows and columns which can be formed from any of the elements of the matrix. If any such determinant of degree p does not vanish identically, the matrix is said to be of rank p, and the set of p functions are independent. However, if all the determinants having p rows and columns vanish identically, one must consider the determinants having $p - 1$ rows and columns. If there exists one $p - 1$ rowed determinant which does not vanish identically, the rank of the matrix is $p - 1$, and only $p - 1$ of the functions can be regarded as independent, while the remaining one can be expressed in terms of the $p - 1$ independent functions. The procedure outlined above can be repeated, and it is readily seen that in order for the matrix to be of rank r it is only necessary to show that all determinants of degree $r + 1$ vanish identically and at least one determinant of degree r is different from zero.

The complete statement of the facts regarding the independence of the p functions is that the necessary and sufficient condition for the independence of the system of functions (18.1) is that the rank of the matrix M (18.2) be equal to p. Furthermore, if $p = n$ and $r < p$, then of the set of functions just r are independent, and the remaining $p - r$ can be expressed in terms of the r independent y's. Also, if $p > n$, the p functions cannot be independent, since the determinant would contain one or more columns of zeros, showing that the rank can not be greater than n.

As an illustration of the use of the above criterion, consider the following three functions y_1, y_2, y_3 of the variables x_1, x_2, x_3.

$$y_1 = x_1 + x_2 - x_3$$
$$y_2 = x_1 - x_2 + x_3$$
$$y_3 = x_1^2 + x_2^2 + x_3^2 - 2x_2x_3$$

The matrix (18.2) becomes

$$M = \begin{Vmatrix} 1 & 1 & -1 \\ 1 & -1 & +1 \\ 2x_1 & 2(x_2 - x_3) & 2(x_3 - x_2) \end{Vmatrix}$$

In order to determine the rank of M we first evaluate the determinant of degree 3

$$\begin{vmatrix} 1 & 1 & -1 \\ 1 & -1 & +1 \\ 2x_1 & 2(x_2 - x_3) & 2(x_3 - x_2) \end{vmatrix} = \begin{vmatrix} 1 & 0 & -1 \\ 1 & 0 & +1 \\ 2x_1 & 0 & 2(x_3 - x_2) \end{vmatrix}$$

The right-hand determinant follows from the first by adding the third column to the second, and since a column of zeros results, the value of the determinant is zero, and so the rank of M is less than 3. To show that it is of order 2, a nonvanishing two-rowed determinant must be found. It is at once obvious that there is such a determinant, namely,

$$\begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

Thus M is of rank 2, so that only two of the y's are independent. Taking these to be y_1 and y_2, it is easily seen that y_3 can be expressed in terms of them, since

$$y_3 = \tfrac{1}{2}(y_1^2 + y_2^2)$$

19. Properties of Jacobians. It should be pointed out that Jacobian determinants, often called Jacobians for briefness, have properties much like the derivatives of functions of one variable. A few of these useful properties will be given below in terms of functions of only two variables, though they hold for any number of variables. The proofs depend on the algebra of determinants but will not be given.

(1) Let u and v be two functions of the variables x and y, and suppose that in turn x and y are functions of two other variables s, t, so that

$$u = u(x,y) \qquad v = v(x,y)$$
$$x = x(s,t) \qquad y = y(s,t)$$

then

$$\frac{\partial(u,v)}{\partial(s,t)} = \frac{\partial(u,v)}{\partial(x,y)} \frac{\partial(x,y)}{\partial(s,t)} \tag{19.1}$$

provided all the partial derivatives involved are continuous.

(2) If $s = u$ and $t = v$, from eq. (19.1),

$$\frac{\partial(u,v)}{\partial(x,y)} = \frac{1}{\partial(x,y)/\partial(u,v)} \tag{19.2}$$

This assumes that the functions $u = u(x,y)$ and $v = v(x,y)$ have unique inverses, but the Jacobian itself furnishes a criterion for the existence of unique inverses. In terms of two variables, the conditions may be stated as follows. If $u = u(x,y)$ and $v = v(x,y)$ are continuous, together with their first-order partial derivatives, at a point (x_0,y_0), and if $\partial(u,v)/\partial(x,y) \neq 0$ at (x_0,y_0), the unique inverses $x = x(u,v)$ and $y = y(u,v)$ exist and are continuous, together with their first-order partial derivatives in the neighborhood of (x_0,y_0).

(3) A property of Jacobians analogous to the formula for change of variable in functions of more than one variable is as follows. Let $u = u(r,s,t)$ and $v = v(r,s,t)$, where the variables r, s, t are in turn functions of the two independent variables x and y. Then

$$\frac{\partial(u,v)}{\partial(x,y)} = \frac{\partial(u,v)}{\partial(r,s)}\frac{\partial(r,s)}{\partial(x,y)} + \frac{\partial(u,v)}{\partial(s,t)}\frac{\partial(s,t)}{\partial(x,y)} + \frac{\partial(u,v)}{\partial(t,r)}\frac{\partial(t,r)}{\partial(x,y)} \tag{19.3}$$

20. Taylor's theorem. Given a function $f(x)$ which possesses derivatives of the nth order in an interval (ab), Taylor's formula makes it possible to express the function $f(x)$ in terms of the value of the function and its n derivatives at a neighboring point x_0. The formula is

$$f(x_0 + h) = f(x_0) + \frac{h}{1!}f'(x_0) + \frac{h^2}{2!}f''(x_0) + \cdots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(x_0) + r_n \tag{20.1}$$

where x_0 and $x_0 + h$ lie in the interval (ab). The remainder term r_n furnishes an estimate of the error involved when the value of $f(x_0 + h)$ is calculated by means of the polynomial

$$\sum_{s=0}^{n-1} \frac{h^s}{s!} f^{(s)}(x_0)$$

There are various ways of expressing r_n, but only two will be given here.

(1) $r_n = \dfrac{h^n(1-\theta)^{n-1}}{(n-1)!}f^{(n)}(x + \theta h) \qquad 0 < \theta < 1$

(2) $r_n = \dfrac{h^n}{n!}f^{(n)}(x + \theta h) \qquad 0 < \theta < 1$

A special case of Taylor's formula arises when the point x_0 happens to be zero. Then

$$f(h) = f(0) + hf'(0) + \frac{h^2}{2!}f''(0) + \ldots + \frac{h^{n-1}}{(n-1)!}f^{n-1}(0) + r_n \tag{20.2}$$

This is known as Maclaurin's formula. In this case the two expressions for the remainder may be written

$$r_n = \frac{h^n(1-\theta)^{n-1}}{(n-1)!}f^{(n)}(\theta h) \qquad 0 < \theta < 1$$

$$r_n = \frac{h^n}{n!}f^{(n)}(\theta h) \qquad 0 < \theta < 1$$

Taylor's formula can be generalized to functions of more than one variable. It will suffice to give it for a function $f(x,y)$ of two variables which possesses all partial derivatives including those of order n in a domain of definition. Since the formula is somewhat more complicated it is more convenient to express it in symbolic operator form. Let (x_0,y_0) and (x_0+h,y_0+k) be two points in the domain S_2 under consideration; then Taylor's formula may be written

$$f(x+h,y+k) = \sum_{\nu=0}^{n-1} \frac{1}{\nu!} d^\nu f(x_0,y_0) + r_n \tag{20.3}$$

The symbol $d^\nu f(x,y)$ denotes a certain sequence of operations to be carried out on $f(x,y)$. When $\nu = 1$, then df is just the total differential

$$df = h\left(\frac{\partial f}{\partial x}\right)_{x_0y_0} + k\left(\frac{\partial f}{\partial y}\right)_{x_0y_0}$$

where the independent differentials have been replaced by h and k. The symbol d^2f is to be interpreted as $d(df)$, i.e., the total differential of the function df; thus

$$d^2f = d\left(h\frac{\partial f}{\partial x} + k\frac{\partial f}{\partial y}\right) = \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)\left(h\frac{\partial f}{\partial x} + k\frac{\partial f}{\partial y}\right)$$

$$= \left(h^2\frac{\partial^2 f}{\partial x^2} + 2hk\frac{\partial^2 f}{\partial x\,\partial y} + k\frac{\partial^2 f}{\partial y^2}\right)_{x_0y_0}$$

The general operation $d^\nu f$ can be evaluated formally by writing

$$d^\nu f = \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^\nu f$$

and expanding the operator $\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^\nu$ by means of the binomial theorem.

21. Maxima and minima. In the calculus of functions of a single variable, it is shown that the function $f(x)$ has a relative maximum or

minimum at a point x_0 if $f'(x)$ exists and is continuous at x_0 and $f'(x_0) = 0$. It is quite clear from Fig. 21.1 that at a point of discontinuity of $f'(x)$, $f(x)$ may have a maximum. Assuming a continuous derivative, the condition $f'(x_0) = 0$ merely insures that the tangent line to the curve is parallel to the x axis at $x = x_0$, but the values of the function in the neighborhood of $x = x_0$ may all lie below the tangent line as in the case of a maximum or all above it (minimum) or some above and some below (neither a maximum nor a minimum). If no distinction is to be made between these points, they are called stationary points or extrema. In order to ascertain whether an extremum is a maximum or a minimum, it is necessary to have recourse to the second derivative or to the first nonvanishing even derivative of higher order. Thus, if $f^{(n)}(x_0) > 0$, n

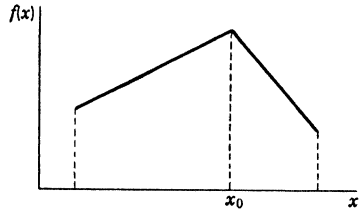


FIG. 21.1

even, $f(x)$ has a minimum, or if $f^{(n)}(x_0) < 0$, the point x_0 corresponds to a maximum of $f(x)$. On the other hand, if the lowest order nonvanishing derivative is of odd order, then the function has neither a maximum nor a minimum, but it has an inflection point at $x = x_0$. The above conditions may easily be derived from Taylor's formula. The corresponding conditions for functions of two variables will be developed.

In the case of a function of two variables, the condition for an extremum is that the tangent plane to the surface $u = f(x,y)$ at the point (x_0,y_0) be parallel to the xy plane. If u possesses continuous first partial derivatives in the neighborhood of (x_0,y_0), the condition that the tangent plane be parallel to the xy plane is that $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$ at (x_0,y_0). This may be seen from the fact that the total differential of $f(x,y)$ must be zero in order that the tangent plane be parallel to the xy plane; thus

$$du = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy = 0$$

Since dx and dy are arbitrary, it follows that

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0 \quad \text{at} \quad (x_0,y_0)$$

These constitute the necessary conditions for an extremum.

Conditions for a maximum or a minimum can be given if $f(x,y)$ possesses second-order derivatives at the point in question, for then the values of $f(x,y)$ at points in the neighborhood of (x_0,y_0) may be obtained from Taylor's formula, which becomes

$$f(x_0+h,y_0+k) = f(x_0,y_0) + df(x_0,y_0) + \frac{1}{2}d^2f(x_0+\theta h,y_0+\theta k)$$

$$f(x_0+h,y_0+k) - f(x_0,y_0) = \frac{\partial f}{\partial x}(x_0,y_0)h + \frac{\partial f}{\partial y}(x_0,y_0)k$$

$$+\frac{1}{2}\left[\frac{\partial^2 f(x,y)}{\partial x^2}h^2 + 2\frac{\partial^2 f(x,y)}{\partial x\,\partial y}hk + \frac{\partial^2 f(x,y)}{\partial y^2}k^2\right] \qquad \begin{matrix} x = x_0 + \theta h \\ y = y_0 + \theta k \end{matrix}$$

If the conditions for an extremum are satisfied, the above formula may be written

$$\Delta u = \tfrac{1}{2}(Ah^2 + 2Bhk + Ck^2)$$

where

$$A = \frac{\partial^2 f(x_0+\theta h,y_0+\theta k)}{\partial x^2} \qquad B = \frac{\partial^2 f(x_0+\theta h,y_0+\theta k)}{\partial x\,\partial y}$$

$$C = \frac{\partial^2 f(x_0+\theta h,y_0+\theta k)}{\partial y^2}$$

For a minimum the difference Δu (see Sec. 15) must be positive in the neighborhood of (x_0,y_0). The condition for a minimum thus reduces to the conditions which must be satisfied in order that the quantity $H_2 = Ah^2 + 2Bhk + Ck^2 > 0$. This is a homogeneous polynomial in h and k of degree 2, or merely a quadratic form in two variables. It is required to find the conditions for which the quadratic form H_2 will be definite, i.e., either always negative or always positive. To determine this, H_2 may be rewritten in the form

$$H_2 = \frac{1}{A}\left[(Ah + Bk)^2 + (AC - B^2)k^2\right]$$

It is clear that H_2 cannot vanish for any values of h and k not equal to zero, provided

$$B^2 - AC < 0$$

This condition cannot be fulfilled if A and C have different signs; thus A and C must necessarily have the same sign, and H_2 will have the same sign as A and C. For a minimum in $f(x,y)$ it is required that $B^2 - AC < 0$, and A and C must be positive, while for a maximum A and C must be negative. The complete theorem may be stated as follows.

Theorem. If at the point (x_0,y_0), $\partial u/\partial x = 0$, $\partial u/\partial y = 0$, and

$$[\partial^2 u/(\partial x\,\partial y)]^2 - (\partial^2 u/\partial x^2)(\partial^2 u/\partial y^2) < 0,$$

and if the second-order derivatives are continuous near (x_0,y_0), then u will have a relative maximum at (x_0,y_0) if $\partial^2u/\partial x^2 < 0$, and a relative minimum when $\partial^2u/\partial x^2 > 0$.

If, however, $[\partial^2u/(\partial x\,\partial y)]^2 - (\partial^2u/\partial x^2)(\partial^2u/\partial y^2) > 0$, the function has no maximum or minimum. When $\partial^2u/\partial x^2$ and $\partial^2u/\partial y^2$ have different signs, the function has a saddle point at (x_0,y_0).

Much the same argument is used to obtain the conditions for functions of three variables. In this case, it is necessary that

$$\partial u/\partial x = \partial u/\partial y = \partial u/\partial z = 0$$

at (x_0,y_0,z_0). The conditions for a maximum or minimum reduce to the circumstances under which a quadratic form H_3 in three variables is definite. For completeness, this condition will be stated. The quadratic form H_3 may be written

$$H_3 = a_{11}h^2 + a_{22}k^2 + a_{33}l^2 + 2a_{12}hk + 2a_{13}hl + 2a_{23}kl$$

The necessary and sufficient conditions that H_3 be positive definite are that

$$a_{11} > 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0$$

where $a_{ij} = a_{ji}$. The term H_3 will be negative definite if the inequalities in the first, third, etc., determinates are reversed.

22. Maxima and minima with auxiliary conditions. It often happens, in physical applications, that one wishes to find the maxima or minima of a function of several variables while at the same time certain relations between the variables must be maintained. To illustrate what is meant here, suppose it is desired to determine the lengths of the principal axes of the section of the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

cut out by a plane passing through the origin, whose equation is $lx + my + nz = 0$. Stated in a slightly different form the above problem is the same as that of finding the maxima and minima of $r^2 = x^2 + y^2 + z^2$ subject to the auxiliary conditions

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \qquad lx + my + nz = 0$$

This means that in the function r^2 only one variable is independent, so that by using the auxiliary conditions r^2 could be expressed in terms of one variable, and then one could proceed in the usual way to deter-

mine the value of the one independent variable for which r^2 has an extreme value. This method is usually not the most convenient in practice. A convenient method due to Lagrange will be outlined.

Consider a function of $m + n$ variables, together with n relations involving these variables. Thus only m of the variables will be independent. For purposes of explanation, denote the m independent variables by $x_1, x_2, \ldots, x_m$, and the dependent ones by $y_1, y_2, \ldots, y_n$. Then the extreme values of $f(x_1,x_2,\ldots,x_m,y_1,y_2,\ldots,y_n)$ are desired when n relations obtain among the variables, namely,

$$\begin{aligned}
\varphi_1(x_1,x_2,\ldots,x_m,y_1,\ldots,y_n) &= 0\\
\varphi_2(x_1,x_2,\ldots,x_m,y_1,\ldots,y_n) &= 0\\
&\cdots\cdots\cdots\cdots\\
\varphi_n(x_1,x_2,\ldots,x_m,y_1,\ldots,y_n) &= 0
\end{aligned}$$

It will be supposed that all these functions are differentiable. In order that f have extreme values, its total differential must vanish, i.e., $df = 0$. Taking the total differentials of the above equations, there exist $n + 1$ equations to satisfy. When written out in detail, they are

$$\frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 + \ldots + \frac{\partial f}{\partial y_1}dy_1 + \ldots + \frac{\partial f}{\partial y_n}dy_n = 0 \tag{22.1}$$

$$\left.\begin{aligned}
&\frac{\partial \varphi_1}{\partial x_1}dx_1 + \frac{\partial \varphi_1}{\partial x_2}dx_2 + \ldots + \frac{\partial \varphi_1}{\partial y_1}dy_1 + \ldots + \frac{\partial \varphi_1}{\partial y_n}dy_n = 0\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&\frac{\partial \varphi_n}{\partial x_1}dx_1 + \frac{\partial \varphi_n}{\partial x_2}dx_2 + \ldots + \frac{\partial \varphi_n}{\partial y_1}dy_1 + \ldots + \frac{\partial \varphi_n}{\partial y_n}dy_n = 0
\end{aligned}\right\} \tag{22.2}$$

In principle the n dependent differentials $dy_1, \ldots, dy_n$ can be eliminated by using the last n equations, and the first equation would assume the form

$$G_1\,dx_1 + G_2\,dx_2 + \ldots + G_m\,dx_m = 0$$

which now contains only independent differentials. The left-hand side can be made to vanish for all choices of the differentials only if

$$G_1 = G_2 = \ldots = G_m = 0$$

These m equations, together with the n equations

$$\varphi_i(x_1,x_2,\ldots,y_1,\ldots,y_n) = 0 \qquad i = 1, 2, \ldots, n$$

serve to determine the values of the $m + n$ unknowns $x_1, x_2, \ldots, x_m$, $y_1, \ldots, y_n$ for which f assumes an extreme value.

In actual practice, the independent differentials can be eliminated by the method of undetermined multipliers. The method consists in multiplying the first of equations (22.2) by a constant λ_1, the second by λ_2, etc. The resulting equations are added to eq. (22.1), and the result may be written

$$X_1\,dx_1 + X_2\,dx_2 + \ldots + Y_1\,dy_1 + Y_2\,dy_2 + \ldots + Y_n\,dy_n = 0 \quad (22.3)$$

where

$$X_i = \frac{\partial f}{\partial x_i} + \lambda_1 \frac{\partial \varphi_1}{\partial x_i} + \lambda_2 \frac{\partial \varphi_2}{\partial x_i} + \ldots + \lambda_n \frac{\partial \varphi_n}{\partial x_i} \qquad i = 1, 2, \ldots, m$$

$$Y_j = \frac{\partial f}{\partial y_j} + \lambda_1 \frac{\partial \varphi_1}{\partial y_j} + \lambda_2 \frac{\partial \varphi_2}{\partial y_j} + \ldots + \lambda_n \frac{\partial \varphi_n}{\partial y_j} \qquad j = 1, 2, \ldots, n$$

Now the n as yet undetermined multipliers $\lambda_1, \lambda_2, \ldots, \lambda_n$ can be determined in such a way that the n coefficients Y_j all vanish. Then only independent differentials appear in eq. (22.3), and in order that eq. (22.3) be satisfied,

$$X_1 = 0, \qquad X_2 = 0, \ldots, X_m = 0$$

These equations, together with the n equations $Y_1 = 0, \ldots, Y_n = 0$, and the n equations $\varphi_1 = 0, \varphi_2 = 0, \ldots, \varphi_n = 0$, completely determine the unknown constants and the unknown variables. In practice it is best to write down all $2n + m$ equations, because simplifications may be apparent as in the case of the problem stated at the beginning of this section which will now be solved.

To find the maximum and minimum of the function $r^2 = x^2 + y^2 + z^2$ subject to the conditions

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1 = 0 \qquad lx + my + nz = 0$$

Choosing x as the independent variable, the equations $X_1 = 0$, $Y_1 = 0$, $Y_2 = 0$ are

$$X_1 = 2x + \lambda_1 \frac{2x}{a^2} + \lambda_2 l = 0$$

$$Y_1 = 2y + \lambda_1 \frac{2y}{b^2} + \lambda_2 m = 0$$

$$Y_2 = 2z + \lambda_1 \frac{2z}{c^2} + \lambda_2 n = 0$$

Multiplying the equations by x, y, and z, respectively, and adding, yields the equation

$$2x^2 + 2y^2 + 2z^2 + 2\lambda_1\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}\right) + \lambda_2(lx + my + nz) = 0$$

or simply

$$\lambda_1 = -r^2$$

Using the above equations, it is possible to solve for x, y, and z in terms of λ_2. Thus

$$x = \frac{1}{2}\frac{\lambda_2 l}{r^2/a^2 - 1}, \quad y = \frac{1}{2}\frac{\lambda_2 m}{r^2/b^2 - 1}, \quad z = \frac{1}{2}\frac{\lambda_2 n}{r^2/c^2 - 1}$$

Then $lx + my + nz = \frac{\lambda_2}{2}\left(\frac{l^2a^2}{r^2 - a^2} + \frac{m^2b^2}{r^2 - b^2} + \frac{n^2c^2}{r^2 - c^2}\right) = 0$

Since λ_2 cannot be zero, it follows that the maximum and minimum values of r are given by the above quadratic equation in r^2.

SUGGESTED REFERENCES

W. F. Osgood, *Advanced Calculus*, New York, The Macmillan Company, 1925.

E. W. Hobson, *The Theory of Functions of a Real Variable*, London, Cambridge University Press, 1927.

R. S. Burington and C. C. Torrance, *Higher Mathematics*, New York, McGraw-Hill Book Company, Inc., 1939.

E. J. P. Goursat and E. R. Hedrick, *Mathematical Analysis*, Boston, Ginn & Company, 1904.

R. Courant, *Differential and Integral Calculus*, Vol. 2, New York, Nordemann Publishing Company, 1936.

PROBLEMS

9.1. If $u = a^v$ and $y = \log x$, determine du/dx.

9.2. If $u = \sqrt{1 - y^2}$ and $y = \cos x$, determine du/dx.

9.3. When $x = \sinh t$, $y = \cosh t$, determine dy/dx. Check the result by direct differentiation of $y = \sqrt{1 + x^2}$.

9.4. Given $x = \tan y$, calculate dy/dx and d^2y/dx^2.

9.5. Given $u = f(y)$, $y = g(x)$, compute d^3u/dx^3.

9.6. Given $y = f(x)$ and the inverse exists, compute d^3x/dy^3.

11.1. Determine the limits

(1) $\lim_{x\to 0} \frac{x - \sin x}{x^3}$. (2) $\lim_{x\to -1} \frac{\sin \pi x}{1 + x}$.

11.2. Determine the limits

(1) $\lim_{x\to\infty} \frac{(\log x)^a}{x^\beta}$. (2) $\lim_{x\to\infty} \frac{x}{\sqrt{x^2 - 1}}$.

11.3. Evaluate

$$\frac{x(\pi^2 - x^2) - 2\pi^2 \sin x}{(\pi^2 - x^2)^2} \quad \text{at} \quad x = \pi.$$

11.4. Evaluate the indeterminate forms

(1) $\frac{x^3}{e^{ax} - 1}$, (2) $\frac{\sin^2 x}{x^2}$, (3) $\frac{1 - \cos 2x}{x^2}$, at $x = 0$.

11.5. Evaluate the indeterminate forms

(1) $(x - 1)^n \log (x - 1)$, $n > 0$;

(2) $\frac{a - a^x}{1 - x}$, at $x = 1$.

11.6. Evaluate the indeterminate forms

$x^n \log x$, $n < 0$, $\frac{\cosh qx}{\sinh \pi x}$, $q < \pi$ for $x = \infty$.

11.7. Evaluate the quantity

$$\frac{1}{2}\frac{[(a^2+x^2)/(2ax)]\log[(a+x)/(a-x)]-1}{[(a^2-x^2)/(2ax)]\log[(a+x)/(a-x)]+1}$$

at $x = 0$.

11.8. Evaluate the quantity $(\log x)^{\sin 1/x}$ at $x = \infty$.

11.9. Determine the limits.

(1) $\lim_{x\to 0}\frac{a^x - b^x}{x}$. (2) $\lim_{x\to \pi/4}\frac{\tan x - 1}{\pi/4 - x}$.

11.10. Determine the limits:

(1) $\lim_{x\to\infty} x^n e^{-x^2}$. (2) $\lim_{x\to\infty} x^n e^{-x}\log x$.

(3) $\lim_{x\to 0} x^\alpha\left[\log\left(\frac{1}{x}\right)\right]^\beta$. (4) $\lim_{x\to\pi/2}(\sin x)^{\tan x}$.

13.1. If $u = e^{xy}$ and $x = \log\sqrt{r^2+s^2}$ and $y = \tan^{-1} s/r$, compute $\partial u/\partial r$ and $\partial u/\partial s$.

13.2. Let $u = x^2 - y^2$, where $x = 2r - 3s + 7$ and $y = -r + 8s - 9$. Compute $\partial u/\partial r$ and $\partial u/\partial s$.

13.3. If $u = f(x,y)$ and $x = r\cos\varphi$ and $y = r\sin\varphi$, show that

$$\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 = \left(\frac{\partial u}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial u}{\partial \varphi}\right)^2$$

13.4. Let $u = \frac{x+y}{1-xy}$, where $x = \tan(2r - s^2)$, $y = \cot(r^2 s)$. Compute $\partial u/\partial r$ and $\partial u/\partial s$.

13.5. If $u = f(xy)$ show that $x(\partial u/\partial x) = y(\partial u/\partial y)$.

13.6. Let $u = f(x/y)$. Show that $x(\partial u/\partial x) + y(\partial u/\partial y) = 0$.

13.7. If $u = f(y+ax) + g(y-ax)$, show that $\partial^2 u/\partial x^2 = a^2(\partial^2 u/\partial y^2)$.

13.8. If $u = f(x,y,z)$ and $x = r\sin\theta\cos\varphi$, $y = r\sin\theta\cos\varphi$, $z = r\cos\theta$, compute $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 + \partial^2 u/\partial z^2$ in terms of r, θ, and φ.

13.9. If $u = f(x,y,z)$ and $x = \sqrt{\xi\eta}\cos\varphi$, $y = \sqrt{\xi\eta}\sin\varphi$, $z = \frac{1}{2}(\xi - \eta)$, compute

$$\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 + \partial^2 u/\partial z^2$$

in terms of ξ, η, and φ.

14.1. The electric potential at any point due to a charge q located at the origin of coordinates and a uniform electric field of intensity E is $V = q/r - Ez$. Determine the electric intensity along the outer normal to a sphere of radius R with center at the charge, i.e., determine directional derivative of V in the direction of a normal to the sphere. In which direction from a point on the sphere is V changing most rapidly?

14.2. The electric field in any direction outside of a conducting sphere of radius R in a uniform field is given by the directional derivative of the potential function

$$V = V_0 - Er\cos\theta + ER^3(\cos\theta)/r^2.$$

Find the directional derivative of the above expression in a direction perpendicular to the plane $\theta = \pi/2$.

14.3. The gravitational potential around a thin rod of length $2l$, density per unit length σ, center at origin, and lying along the z axis, is

$$V = \sigma\log\frac{-(z-l)+\sqrt{(z-l)^2+x^2+y^2}}{-(z+l)+\sqrt{(z+l)^2+x^2+y^2}}$$

Find the directional derivative in the outward direction along a line through the origin, in the xz plane.

15.1. The deflection of a cantilever beam of length l with moment of inertia I made of material whose modulus of elasticity is E is given by

$$y = -\frac{P}{EI}\left(\frac{x^3}{6} - \frac{l^2x}{2} + \frac{l^3}{3}\right)$$

where P is the load applied at the free end $x = 0$. If one wished to determine the constant EI of the beam by measurements of the other quantities, what would be the per cent error in EI if errors Δx, Δl, and Δy were made in measuring x, l, and y?

15.2. Using the concept of the differential of a function, obtain an approximate value for the volume of a spherical shell of small thickness.

15.3. Using the concept of the differential of a function, obtain an approximate expression for the volume of a conical shell of small thickness bounded by two circular cones having a common vertex, a common axis, and equal heights.

16.1. If $x = a \sin\theta$ and $y = a \cos\theta$, compute dy/dx.

16.2. The parametric equations for a spherical surface are $x = a \sin\theta \cos\varphi$, $y = a \sin\theta \sin\varphi$, $z = a \cos\theta$. Compute $\partial z/\partial x$, $\partial z/\partial y$, and $\partial y/\partial x$.

16.3. The parametric equations for a certain parabolic surface are $x = r \cos\theta$, $y = r \sin\theta$, $z = \frac{1}{2}(r^2 - 1)$. Compute $\partial z/\partial x$ and $\partial z/\partial y$.

17.1. Given the equation $x^2/z^2 + y^2/b^2 + z^2/c^2 = 1$ of an ellipsoid, compute $\partial z/\partial x$ and $\partial z/\partial y$.

17.2. If $ye^{-xy} = \sin x$, compute dy/dx.

17.3. If $y \sin x = x \tan y$, compute dy/dx.

17.4. If $F(x,y,z) = 0$, compute $\partial^2 z/\partial x^2$.

17.5. If $xu^2 + yv^3 = xy^2$, $yu^3 - xv^2 = x^2y$, compute $\partial v/\partial y$.

17.6. If $xu + uv - v = x^3$, $v^2 - xv + u = x^2$, compute du/dx.

17.7. If $V = 2uv$ and $u^5 + v^5 + x^5 = 3y$, $u^3 + v^3 + y^3 = -3x$, compute $\partial V/\partial x$.

17.8. Given $ue^v + vx = y \sin x$, $u \cos u = x^2 + y^2$. Compute $\partial v/\partial y$.

17.9. Given $yx^y = \sin x$. Compute dy/dx.

17.10. Given $u = Ce^{-k(x^2+y^2)}$, where $x^2 + y^2 = a^2$. Compute $\partial u/\partial y$.

17.11. Given $u^3 - xv = y$, $v^3 - yu = x$. Compute $\partial v/\partial y$.

17.12. Given $x = u + vu^v$, $y = v - uv^u$. Compute $\partial u/\partial x$.

Test the following sets of functions for independence.

18.1. $u = x + y + z$, $v = x - 2xy + z$, $w = x^2 + y^2 + z^2$.

18.2. $u = \cos\theta$, $v = 2r \sin\theta$, $w = (4r^2 - 2)\cos\theta$.

18.3. Ascertain whether u, v, and w are independent where

$$u = x^3 + x^2y + x^2z - z^2(x + y + z)$$
$$v = x + z$$
$$w = x^2 - z^2 + xy - zy$$

18.4. Test the independence of the functions u, v, w when

$$u = r \sin\theta \cos\varphi, \quad v = r \sin\theta \sin\varphi, \quad w = r \cos\theta.$$

18.5. Are the functions u, v, and w independent when

$$(y - z)u = x, \quad (z - x)v = y, \quad (x - y)w = z?$$

19.1. Given $x = r \cos\theta$, $y = r \sin\theta$, and the inverse relation

$$r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1} y/x.$$

Demonstrate eq. (19.2) by direct computation.

19.2. Discuss the geometrical significance of the fact that $\partial(w,v)/\partial(x,y) = 0$ at a point.

19.3. Given $u = x - y^{2/3}$, $v = x + y^2$. Determine whether or not these functions have unique inverses over the following domain of definition.

(1) $-c \leq x \leq c, \quad -c \leq y \leq c.$

(2) $-c \leq x \leq c, \quad 0 \leq y \leq c.$

19.4. Do the functions $x = \frac{1}{2}(u - v)$, $y = \sqrt{uv}$ have unique inverses?

21.1. Does the function $x^2 + y^2 + 4x - 3y - 7$ have a relative minimum?

21.2. Test the following functions for extrema and relative maxima and minima.

(a) $u = x^3 + 3x^2 - 2xy + 5y^2 - 4y^3$

(b) $u = 2(x - y)^2 - x^4 - y^4$

(c) $u = x^2y^2 - 5x^2 - 8xy - 5y^2$

21.3. Find and investigate the extrema of

$$u = (x^2 + y^2)e^{-(x^2+y^2)}$$

and sketch the function roughly.

22.1. Find the maximum of $F(x,y,z) = xyz$, subject to the condition $x + y + z = 1$.

22.2. Find the greatest rectangular volume which can be inscribed in the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

22.3. The following is a well-known problem in statistical mechanics. Consider the molecules of a certain quantity of gas in a chamber; each molecule at a given instant will have a certain position and momentum which can be specified by six numbers, x, y, z, p_x, p_y, p_z. Thus a point in a 6-dimensional space would represent the state of a molecule at any instant, and would in the course of time move in this space as the position and momentum of the molecule changes. Suppose that the space be divided into small cells. All the molecules whose coordinates and momenta fall into one cell will have about the same energy. The cells will be numbered 1, 2, $\cdots$, and let there be n_i molecules in cell i. If the total number of molecules is N, then

$$N = n_1 + n_2 + n_3 + \cdots \tag{1}$$

If the total energy of the molecules is E, and if the energy of a molecule in the ith cell is ϵ_i, then

$$E = n_1\epsilon_1 + n_2\epsilon_2 + \cdots \tag{2}$$

The total number of ways of distributing the N molecules in the cells will be

$$S = \frac{N!}{n_1!n_2!n_3! \cdots}$$

The most probable distribution among the energy cells will be that which makes S a maximum subject to conditions (1) and (2). Find this distribution. Since all numbers can be assumed large replace $n!$ by $\Gamma(n + 1)$ and use the asymptotic form of $\Gamma(n + 1)$. See eq. (113.15).

Chapter 3

INTEGRAL CALCULUS

23. The Riemann integral defined. The process of integration is essentially a process of summing. One may find the area under a given curve by integration, since the region under such a curve may be broken up into pieces and the area of the pieces added together. To be more precise, consider the bounded function $f(x)$ defined throughout a finite interval $a \leq x \leq b$, the graph of which is shown in Fig. 23.1, and divide the interval $[ab]$ into n subintervals by the points

$$a = x_0 < x_1 < x_2, \ldots, x_{n-1} < x_n = b$$

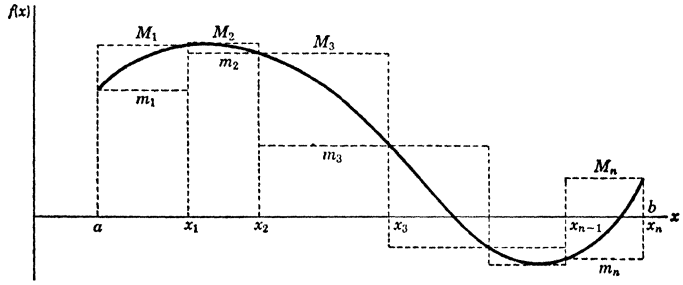


FIG. 23.1

Since the function is bounded in $[ab]$ it is bounded in every subinterval. Let the greatest lower and the least upper bounds in successive intervals be denoted by m_1M_1, m_2M_2, etc. The area contained in the rectangles whose heights are M_1M_2, etc., for any scheme of subdivision σ of the interval $[ab]$ will be

$$S_\sigma = M_1(x_1 - a) + M_2(x_2 - x_1) + \ldots + M_n(b - x_{n-1})$$

On the other hand, the area contained in the rectangles whose heights are m_1, m_2, etc. will be

$$s_\sigma = m_1(x_1 - a) + m_2(x_2 - x_1) + \ldots + m_n(b - x_{n-1})$$

It is readily seen that these two sums constitute an upper and a lower bound to the area beneath the curve. The two sums are called the upper and lower approximate sums.

As one changes the scheme of subdivision these sums change. In particular as the intervals are made smaller, i.e., as the number of subdivisions is increased, the value of S_σ decreases and the value of s_σ increases. As we go through all the modes of subdivision we obtain a set of values of S_σ and s_σ, and the set of all S_σ's has a lower bound S, and the set of all s_σ's has an upper bound s. The numbers S and s are called the upper and lower integrals of $f(x)$ in the interval $[ab]$. It is clear that $S \geq s$, and in particular if $S = s$, then $f(x)$ is said to be Riemann integrable (or integrable-R) in the interval $[ab]$, and the common value of S and s is the integral of $f(x)$ over the interval. The value of the integral is denoted by $\int_a^b f(x)dx$.

The value of the Riemann integral can be arrived at in another way. The function $f(x)$ must as before be bounded in an interval $[ab]$, and this interval divided into n subintervals. The length of the rth interval, i.e., $x_r - x_{r-1}$, will be denoted by δ_r, and of all such subintervals let the largest be denoted by δ. This largest interval is called the norm of the particular scheme of subdivision. Now choose an arbitrary point in the subinterval δ_r and call it ξ_r and form the sum

$$I_\delta^\sigma = \sum_{r=1}^{n} f(\xi_r)\delta_r$$

The value of this sum will of course depend on the scheme of subdivision and on the choice of the points ξ_r. It may happen, however, that I_δ^σ approaches a limit I as $\delta \to 0$ independent of the scheme of subdivision or choice of the points ξ_r. If this is the case, I is the Riemann integral of $f(x)$ in the interval $[ab]$ and denoted by

$$I = \lim_{\delta \to 0} \sum_{r=1}^{n} f(\xi_r)\delta_r = \int_a^b f(x)dx$$

This definition and the first one given are equivalent. The proof of this is based on a theorem of Darboux.

24. Classes of bounded integrable functions. It is convenient to know some of the classes of functions which possess Riemann integrals. A partial list of such classes is given below when the interval of definition is bounded.

1. Every continuous function.
2. Every bounded monotonic function.
3. A bounded function whose discontinuities can be enclosed in a

finite number of intervals whose total length can be made arbitrarily small.

4. If the function $f(x)$ is integrable in the interval $[ab]$, every other function $g(x)$ obtained by altering the value of $f(x)$ at a finite number of points in $[ab]$ is integrable, and

$$\int_a^b f(x)dx = \int_a^b g(x)dx$$

5. The sum, product, or quotient of two integrable functions is an integrable function, provided that in the case of the quotient the absolute value of the denominator is greater than a fixed positive number throughout the range of integration.

6. If $f(x)$ is integrable, $|f(x)|$ is integrable also.

25. Fundamental properties of Riemann integrals. Some important properties of integrals are enumerated below. It is always assumed that $f(x)$ is integrable in the interval $[ab]$.

1. $\int_a^b f(x)dx = -\int_b^a f(x)dx$
2. $\int_a^b dx = b - a$
3. $\int_a^b kf(x)dx = k\int_a^b f(x)dx; \quad k = \text{constant}$
4. $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx \quad \text{where } a \leq c \leq b$
5. If $f(x) \geq g(x)$ throughout the interval $[ab]$, then

$$\int_a^b f(x)dx \geq \int_a^b g(x)dx$$

6. $\left|\int_a^b f(x)dx\right| \leq \int_a^b |f(x)|dx$

7. If M and m are the upper and lower bounds of $f(x)$ in the interval $[ab]$, then

$$m(b - a) \leq \int_a^b f(x)\,dx \leq M(b - a)$$

8. If $f(x)$ is continuous throughout $[ab]$ there exists a number ξ such that $a \leq \xi \leq b$, and

$$\int_a^b f(x)dx = f(\xi)(b - a)$$

The above relation is called the first mean value theorem for integrals.

9. If $f(x)$ is bounded and integrable in the interval $[ab]$ and $a \leq x \leq b$, then $F(x) = \int_a^x f(t)dt$ is a continuous function of x even though $f(t)$ may not be continuous.

10. If $f(x)$ be *continuous* in the interval $[ab]$, then at every point in $[ab]$, the function $F(x)$ possesses a derivative which is the function $f(x)$, i.e., $F'(x) = f(x)$.

11. If $\varphi(x)$ is a function which at every point of $[ab]$ possesses a derivative which is the continuous function $f(x)$, then

$$\int_a^b f(x)dx = \varphi(b) - \varphi(a)$$

The last two theorems show that every continuous function is a derivative of some other continuous function which is called its primitive or *indefinite integral.*

Although it may be known that $f(x)$ is integrable it may be difficult in any given case to find the value of the integral. Since in practical applications it is often required to evaluate integrals, some useful methods will be given in the following sections. More powerful methods are given in later sections, see especially Secs. 40, 62, and Chapter 7.

26. Integration by change of variable. Suppose that $f(x)$ is continuous in the interval $a \leq x \leq b$ and that $x = \varphi(t)$ is a monotonic function in the interval $\alpha \leq t \leq \beta$ possessing a derivative at every point, and such that $a = \varphi(\alpha)$ and $b = \varphi(\beta)$. Thus as t traverses the interval $[\alpha\beta]$, x must assume all values in the interval $[ab]$ just once. Under these conditions

$$\int f(x)dx = \int f\{\varphi(t)\}\varphi'(t)dt = G(t) = G\{\psi(x)\}$$

where $\psi(x)$ is the inverse of $\varphi(t)$. The success of this method depends on one's ability to find $G(t)$.

The above restrictions imposed on $x = \varphi(t)$ should be carefully adhered to, or incorrect results may be obtained, as may be seen from the following example. First consider the simple integral

$$I = \int_{-1}^{1} \frac{dx}{1 + x^2} = \tan^{-1} x \Big]_{-1}^{+1} = \frac{\pi}{4} - \left(-\frac{\pi}{4}\right) = \frac{\pi}{2}$$

Let us evaluate I by means of a change of variable by writing $x = 1/t$ where $-1 \leq t \leq 1$. The above integral becomes

$$I = -\int_{-1}^{+1} \frac{dt}{1 + t^2} = \frac{-\pi}{2}$$

which is not the same as the result first obtained. Several things are wrong with the choice of the function $x = \varphi(t)$, namely, it does not have a derivative at $t = 0$, it is not monotonic in the interval, and finally, it allows x to take on values not in the interval $-1 \leq x \leq 1$.

27. Integration of rational functions. A rational function is one which can be expressed by the quotient of two polynomials. In the next

few sections a rational function will be denoted by $R(x)$, so that

$$R(x) = \frac{P_1(x)}{P_2(x)}$$

If the above fraction is improper, i.e., the degree of the polynomial P_1 is greater than the degree of P_2, it is always possible to write, by division,

$$R(x) = P(x) + \frac{f(x)}{F(x)} \tag{27.1}$$

where $P(x)$ is a polynomial and the remaining fraction is proper.

As far as the integral $\int R(x)dx$ is concerned, the integral $\int P(x)dx$ offers no difficulty, so that we shall deal with the method of evaluating $\int \frac{f(x)}{F(x)}dx$. A common method for doing this is to write $f(x)/F(x)$ in terms of partial fractions. To do this, however, the roots of the polynomial $F(x)$ must be known. A method due to Hermite allows one to carry out the integration even though all the roots of $F(x)$ are not known.

In order to give a clear exposition of the method, let us suppose for the moment that the roots of $F(x)$ are known, so that $F(x)$ may be written in the form

$$F(x) = (x-a)^{\alpha}(x-b)^{\beta}\dots(x-l)^{\lambda} \tag{27.2}$$

It is now possible to express the quotient $f(x)/F(x)$ in terms of partial fractions.

$$\begin{aligned}\frac{f(x)}{F(x)} = {} & \frac{A_1}{(x-a)} + \frac{A_2}{(x-a)^2} + \dots + \frac{A_\alpha}{(x-a)^\alpha} \\ & + \frac{B_1}{(x-b)} + \frac{B_2}{(x-b)^2} + \dots + \frac{B_\beta}{(x-b)^\beta} \\ & + \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ & + \frac{L_1}{(x-l)} + \frac{L_2}{(x-l)^2} + \dots + \frac{L_\lambda}{(x-l)^\lambda}\end{aligned} \tag{27.3}$$

If the integration of the right-hand side were carried out, the terms in the first column would give rise to a logarithmic term, and the remaining terms would yield a rational function. Let us denote the rational function by Y/Q, where Q is the common denominator of all the terms save those in the first column after the integration has been carried out. The function Q can be written out at once, namely,

$$Q = (x-a)^{\alpha-1}(x-b)^{\beta-1}\dots(x-l)^{\lambda-1} \tag{27.4}$$

In addition to this we know that the function Y will be a polynomial of degree one less than Q.

Before integration the terms in the first column may be added together and put over a common denominator, so that the sum of these terms will be of the form X/P, where

$$P = (x - a)(x - b) \dots (x - l) \tag{27.5}$$

and X is a polynomial of degree one less than the polynomial P. Thus far then

$$\int \frac{f(x)}{F(x)}\,dx = \int \frac{X}{P}\,dx + \frac{Y}{Q} \tag{27.6}$$

It is now our purpose to give a method for finding Y/Q. In this regard it is important to notice that Q is the highest common factor in $F(x)$ and $F'(x)$, and can thus be found by algebraic methods. Furthermore $P = F(x)/Q(x)$ and can thus be found by division. In order to determine X and Y, it is convenient to differentiate eq. (27.6) so as to obtain the relation

$$\frac{f(x)}{F(x)} = \frac{X}{P} + \left(\frac{Y}{Q}\right)' \tag{27.7}$$

Since it is known that X and Y are polynomials of degrees one less than P and Q, respectively, they may be written down with unknown coefficients and substituted in the identity (27.7) which serves to determine these coefficients. Before doing this it is convenient to alter eq. (27.7) as follows.

$$f(x) = \frac{X}{P}F(x) + \left(\frac{Y}{Q}\right)'F(x) = QX + Y'P - \frac{YQ'P}{Q} \tag{27.8}$$

when it is remembered that $F = PQ$. By equating like powers of x in this equation there will be enough equations to determine the unknown coefficients. The advantage of this method is that the rational part of the integral can be determined without solving the equation $F(x) = 0$ completely. The part X/P can always be evaluated, since the roots of $P = 0$ are all simple.

It should be noted that if $F(x)$ and $F'(x)$ have no common factor, the integrated result contains no rational part and

$$\int \frac{f(x)}{F(x)}\,dx = \int \frac{X}{P}\,dx$$

which can best be integrated by the method of partial fractions.

As an illustration of the method, we shall evaluate

$$\int \frac{3x^2 + x - 2}{(x - 1)^3(x^2 + 1)}\,dx$$

In the problem above,

$$f(x) = 3x^2 + x - 2, \quad \text{and} \quad F(x) = (x - 1)^3(x^2 + 1)$$

The first step is to find Q by differentiating $F(x)$, which gives

$$F'(x) = 3(x - 1)^2(x^2 + 1) + 2(x - 1)^3x$$
$$Q = (x - 1)^2$$

The value of P may now be found from the relation $P = F/Q$. Thus

$$P = \frac{(x-1)^3(x^2+1)}{(x-1)^2} = (x-1)(x^2+1)$$

We also know that

$$Y = A + Bx, \qquad X = C + Dx + Ex^2$$

Substituting these in eq. (27.8), we have

$$3x^2 + x - 2 = (x-1)^2(C + Dx + Ex^2) + B(x-1)(x^2+1) - \frac{(A+Bx)\cdot 2\cdot(x-1)(x-1)(x^2+1)}{(x-1)^2}$$

Equating coefficients of like powers of x, we have

$$E = 0, \quad D = -\frac{5}{2}, \quad C = -\frac{1}{2}, \quad B = -\frac{5}{2}, \quad A = 2$$

Therefore

$$\int \frac{3x^2+x-2}{(x-1)^3(x^2+1)}\,dx = \frac{1}{2}\int \frac{5x+1}{(x-1)(x^2+1)}\,dx - \frac{5x-4}{2(x-1)^2}$$

The resulting integral may be integrated by using partial fractions. The result will contain a logarithmic term.

28. Integrals reducible to rational functions. In some cases it is possible to reduce integrals of irrational functions to integrals of rational functions. These fall into three main types. The first type may be represented by the following integral.

$$\int R\left[x, \left(\frac{ax+b}{cx+d}\right)^{\alpha}, \left(\frac{ax+b}{cx+d}\right)^{\beta}, \ldots\right]dx \tag{28.1}$$

where R is a rational function of the arguments x, $[(ax+b)/(cx+d)]^{\alpha}$, etc. The exponents α, β, etc., are rational numbers, and a, b, c, and d are constants.

To evaluate integrals of this type one proceeds as follows: First find the least common multiple m of the denominators of α, β, etc., and then make the substitution $t^m = (ax+b)/(cx+d)$. This makes the integrand a rational function of t. As an example, consider the integral

$$\int \frac{x^3\,dx}{(x-1)^{1/2}(x-1)^{1/4}}$$

Here $\alpha = \frac{1}{2}$ and $\beta = \frac{1}{4}$. The least common multiple of 2 and 4 is 4 so that $m = 4$. Then $t^4 = x - 1$, and $x = 1 + t^4$, and $dx = 4t^3\,dt$, so that the integral is now

$$\int \frac{4(1+t^4)^3 t^3\,dt}{t^2\cdot t} = 4\int (1+t^4)^3\,dt$$

The resulting integrand is now rational and can be integrated at once.

The next type of integral to be considered is the so-called *bionomial integral*

$$\int x^{\alpha}(a + bx^{\beta})^{p}\,dx$$

where a and b are constants and the exponents α, β, p are rational numbers. The first step in the reduction of an integral of this type is to make the substitution

$$t = x^{\beta}, \qquad dx = \frac{1}{\beta}t^{\frac{1}{\beta}-1}\,dt$$

whereafter the integral becomes

$$\frac{1}{\beta}\int t^{\frac{\alpha+1}{\beta}-1}(a + bt)^{p}\,dt$$

Denoting $\dfrac{\alpha + 1}{\beta} - 1$ by q, the integral

$$\varphi(p,q) = \frac{1}{\beta}\int t^{q}(a + bt)^{p}\,dt$$

can be made rational in three cases, namely,

(a) If p = an integer, then

$$\varphi(p,q) = \int R(t,t^{q})dt$$

(b) if q = an integer, then

$$\varphi(p,q) = \int R[t,(a + bt)^{p}]dt$$

(c) if $p + q$ = an integer, then

$$\varphi(p,q) = \int R\left[\left(\frac{a + bt}{t}\right)^{p},t\right]dt$$

and each of the above integrals is of the first type (28.1) considered, and may be handled accordingly.

The third type is exemplified by the integral

$$\int R(\sin x, \cos x)dx$$

where R is a rational function of $\sin x$ and $\cos x$. The first step is to make the substitution $t = \tan(x/2)$, whereafter $dx = 2\,dt/(1 + t^2)$, $\sin x = 2t/(1 + t^2)$, and $\cos x = (1 - t^2)/(1 + t^2)$. The resulting integrand will be a rational function of t, so that

$$\int R(\sin x, \cos x)dx = \int R_1(t)\frac{2\,dt}{1 + t^2}$$

29. Integrals of irrational functions and elliptic integrals. In many applications one meets a somewhat broader class of irrational functions, the integrals of which cannot always be evaluated in terms of elementary functions. The class to be considered here may be denoted by

$\int R(x,y)dx$, where R is a rational function of the variables x and y, and may be written $R(x,y) = f(x,y)/F(x,y)$, and $f(x,y)$ and $F(x,y)$ are polynomials in the two variables x and y. In addition y is to be regarded as an irrational function of x of the following form.

$$y = \sqrt{r(x)} \tag{29.1}$$

where $r(x)$ is an ordinary polynomial. The first problem is to reduce the integral to its simplest form. This is accomplished by collecting all the terms in $f(x,y)$ and $F(x,y)$ which contain x alone and products of the form $x^m y^n$, where n is even, and then substitute $\{r(x)\}^{n/2}$ for $y^n(x)$. After this is done, $R(x,y)$ may be written

$$R(x,y) = \frac{f(x,y)}{F(x,y)} = \frac{g(x) + yG(x)}{h(x) + yH(x)}$$

The denominator of the fraction on the right can be rationalized giving

$$\begin{aligned} R(x,y) &= \frac{[g(x) + yG(x)][h(x) - yH(x)]}{h^2(x) - y^2H^2(x)} \\ &= \frac{g(x)h(x) - y^2G(x)H(x) + y[h(x)G(x) - g(x)H(x)]}{h^2(x) - y^2H^2(x)} \end{aligned}$$

The denominator is now a polynomial $P_0(x)$, the first two terms in the numerator constitute a polynomial $P_1(x)$, and the coefficient of y is also a polynomial $P_2(x)$. The original function can then be written in the following ways.

$$\begin{aligned} R(x,y) &= \frac{P_1(x) + yP_2(x)}{P_0(x)} = R_1(x) + y\frac{P_2(x)}{P_0(x)} \\ &= R_1(x) + \frac{r(x)P_2(x)}{yP_0(x)} = R_1(x) + \frac{R_2(x)}{y} \end{aligned}$$

In the reduction so far, we have

$$\int R(x,y)dx = \int R_1(x)dx + \int \frac{R_2(x)}{y}\,dx \tag{29.2}$$

The first integral on the right can be evaluated by the methods given in Sec. 27. The last integral can always be evaluated in terms of elementary functions when the polynomial $r(x)$ is not of higher degree than 2. When $r(x)$ is of the third degree, the integral is said to be an elliptic integral. Integrals of the latter type can be reduced to one of the following 3 forms.

1. $\int \frac{dx}{y}$ an elliptic integral of the first kind
2. $\int \frac{x\,dx}{y}$ an elliptic integral of the second kind
3. $\int \frac{dx}{(x-a)y}$ an elliptic integral of the third kind

In order to carry out the reduction of the integral

$$\int \frac{R_2(x)dx}{y}$$

proceed as follows. If

$$R_2(x) = \frac{r(x)P_2(x)}{P_0(x)}$$

is an improper rational function then divide the numerator by the denominator and obtain

$$\frac{r(x)P_2(x)}{P_0(x)} = a_0x^n + a_1x^{n-1} + \ldots + a_n + \frac{\text{remainder}}{P_0(x)}$$

Then $P_0(x)$ is written

$$P_0(x) = (x-a)^\alpha(x-b)^\beta \ldots (x-l)^\lambda$$

and the right-hand term in the above expression is written in terms of partial fractions.

$$\begin{aligned}\frac{\text{remainder}}{P_0(x)} = {} & \frac{A_1}{x-a} + \frac{A_2}{(x-a)} + \ldots + \frac{A_\alpha}{(x-a)^\alpha} \\ & + \frac{B_1}{x-b} + \frac{B_2}{(x-b)} + \ldots + \frac{B_\beta}{(x-b)^\beta} \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & \frac{L_1}{x-l} + \frac{L_2}{(x-l)} + \ldots + \frac{L_\lambda}{(x-l)^\lambda}\end{aligned}$$

Substituting the above expressions in eq. (29.2) yields the result

$$\begin{aligned}\int R(x,y)dx = {} & \int R_1(x)dx + a_0\int \frac{x^n\,dx}{y} + a_1\int \frac{x^{n-1}\,dx}{y} + \\ & \ldots + a_n\int \frac{dx}{y} + A_1\int \frac{dx}{(x-a)y} + \ldots + L_\lambda\int \frac{dx}{(x-l)^\lambda y}\end{aligned} \tag{29.3}$$

In eq. (29.3) the integrals which still must be treated are

$$\int \frac{x^l\,dx}{y} \quad \text{and} \quad \int \frac{dx}{(x-a)^l y}$$

These can be represented by an integral of the type

$$E_l = \int \frac{(x-x_0)^l\,dx}{\sqrt{r(x)}}$$

since l may be negative or positive, and x_0 may be zero.

The next step is to reduce the integral E_l to its simplest form. Henceforth it will be assumed that $r(x)$ is not of degree higher than four, and is written in the form

$$\begin{aligned}r(x) = {} & b_0(x-x_0)^4 + b_1(x-x_0)^3 + b_2(x-x_0)^2 \\ & + b_3(x-x_0) + b_4\end{aligned} \tag{29.4}$$

A reduction formula for the E_l's can be obtained with the help of the following identity.

$$\frac{b_0(x-x_0)^4+b_1(x-x_0)^3+b_2(x-x_0)^2+b_3(x-x_0)+b_4}{y}=\sqrt{r(x)}$$

Multiplying the above expression by $(x-x_0)^l$ and integrating gives

$$\begin{aligned} &b_0E_{l+4}+b_1E_{l+3}+b_2E_{l+2}+b_3E_{l+1}+b_4E_l \\ &=\int (x-x_0)^l\sqrt{r(x)}\,dx \end{aligned} \tag{29.5}$$

After a partial integration the integral on the right can also be expressed in terms of E's.

$$\begin{aligned} &\int (x-x_0)^l\sqrt{r(x)}\,dx \\ &=\frac{(x-x_0)^{l+1}}{l+1}\sqrt{r(x)}-\frac{1}{2}\int\frac{(x-x_0)^{l+1}r'(x)dx}{(l+1)\sqrt{r(x)}} \\ &=\frac{(x-x_0)^{l+1}\sqrt{r(x)}}{l+1}-\frac{1}{2(l+1)}(4b_0E_{l+4}+3b_1E_{l+3}+2b_2E_{l+2}+b_3E_{l+1}) \end{aligned}$$

Substituting this expression in eq. (29.5) and collecting the E's gives

$$\begin{aligned} &(2l+6)b_0E_{l+4}+(2l+5)b_1E_{l+3}+(2l+4)b_2E_{l+2} \\ &\quad+(2l+3)b_3E_{l+1}+(2l+2)b_4E_l \\ &\quad=2(x-x_0)^{l+1}\sqrt{r(x)} \end{aligned} \tag{29.6}$$

The recursion formula just developed holds for all positive and negative integral values of l and may be used to evaluate any E_l in terms of E_2, E_1, E_0, and E_{-1}. If it should happen that x_0 was a root or zero point for $r(x)$, then b_4 would be zero, and the recurrence formula shows that E_2 can be computed from E_1, E_0, and E_{-1}. The same situation arises when $r(x)$ is of degree 3, for then $b_0=0$.

30. Linear transformations of elliptic integrals. Before taking up the matter of transforming the standard types E_1, E_0, and E_{-1} into the normal forms which have been tabulated, it is necessary to consider a few of the properties of the general linear transformation in one variable. Such a transformation may always be written in the form

$$z=\frac{at+b}{ct+d} \tag{30.1}$$

where the determinant $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ must not be zero. Solving the above relation for t one obtains

$$t=\frac{b-dz}{cz-a} \tag{30.2}$$

The transformation eqs. (30.1) and (30.2) establish a correspondence between points on the z axis and different points on the t axis. If in particular one wishes to have specific points on the z axis go over into special points on the t axis, this serves to determine all or part of the constants a, b, c, and d. For example, it might be desirable to have the points 0 and 1 on the z axis go over into the points 1 and ∞ on the t axis. The constants must then be chosen to satisfy the equations

$$1 = \frac{b}{-a}, \qquad \infty = \frac{b-d}{c-a}$$

These relations can be satisfied if $b = 1, a = -1$, and $c = -1, d = -1$, so that

$$t = \frac{1+z}{1-z}$$

This transforms every point in the line segment $0 \leq z \leq 1$ into points on the line segment $1 \leq t \leq \infty$. The determinant of the coefficients

$$\begin{vmatrix} -1 & 1 \\ -1 & -1 \end{vmatrix}$$

does not vanish, so that the transformation is legitimate.

It is important to ascertain how an elliptic integral reacts to a linear change of variable. First of all, let the fourth-degree polynomial $r(x)$ be subjected to the general linear transformation

$$x = \frac{at+b}{ct+d}$$

The polynomial $r(x)$ may now be written

$$\begin{aligned} r(x) &= b_0x^4 + b_1x^3 + b_2x^2 + b_3x + b_4 \\ &= \frac{\beta_0t^4 + \beta_1t^3 + \beta_2t^2 + \beta_3t + \beta_4}{(ct+d)^4}(ad-bc)^2 \end{aligned}$$

Henceforth the polynomial in t will be denoted by $\varphi(t)$, so that

$$\varphi(t) = \beta_0t^4 + \beta_1t^3 + \ldots + \beta_4$$

If x_1, x_2, x_3, x_4 are the roots of the polynomial $r(x)$, then the roots t_1, t_2, t_3, t_4, of $\varphi(t)$ will be the points on the t axis into which x_1, x_2, x_3, x_4 transform, i.e.,

$$t_i = \frac{b-dx_i}{cx_i-a}, \qquad i = 1, 2, 3, 4$$

This is done even if one of the zero points of $r(x)$ or $\varphi(t)$ lies at infinity.

Since $$dx = \frac{a(ct+d) - c(at+b)}{(ct+d)^2}\,dt = \frac{ad-bc}{(ct+d)^2}\,dt$$

Then $$\frac{dx}{\sqrt{r(x)}} = \frac{ad-bc}{(ct+d)^2}\cdot\frac{dt}{\frac{ad-bc}{(ct+d)^2}\sqrt{\varphi(t)}} = \frac{dt}{\sqrt{\varphi(t)}} \tag{30.3}$$

This relation establishes the fact that a linear transformation of an elliptic integral of the first kind carries the integral over into another elliptic integral of the first kind.

In order to determine what happens to integrals of the second and third kinds it is useful to note that the integral

$$\int \frac{dx}{(x - x_0)\sqrt{r(x)}} \tag{30.4}$$

is an integral of the third kind when x_0 is not a root of $r(x)$. If x_0 is a root of $r(x)$, the integral is essentially an elliptic integral of the second kind, as will be shown below. If x_0 is a root of $r(x)$, this polynomial may be written

$$r(x) = b_0(x - x_0)^4 + b_1(x - x_0)^3 + b_2(x - x_0)^2 + b_3(x - x_0)$$

The point x_0 can be transformed into the point $t = \infty$ by the linear transformation $t = 1/(x - x_0)$, after which the integrand of (30.4) becomes

$$\frac{-dt}{\frac{1}{t}t^2\sqrt{b_0\frac{1}{t^4} + b_1\frac{1}{t^3} + b_2\frac{1}{t^2} + b_3\frac{1}{t}}} = \frac{-t\,dt}{\sqrt{b_0 + b_1 t + b_2 t^2 + b_3 t^3}}$$

$$= \frac{-t\,dt}{\sqrt{\varphi(t)}}$$

It is seen therefore that the new integrand corresponds to that for an integral of the second kind.

If the integral (30.4) is subjected to the linear transformation

$$t = \frac{b - dx}{cx - a}, \qquad x = \frac{at + b}{ct + d}$$

then

$$t_0 = \frac{b - dx_0}{cx_0 - a}, \qquad x_0 = \frac{at_0 + b}{ct_0 + d}$$

and

$$\frac{1}{x - x_0} = \frac{1}{\frac{at + b}{ct + d} - \frac{at_0 + b}{ct_0 + d}} = \frac{A}{t - t_0} + B$$

where

$$A = \frac{(ct_0 + d)^2}{ad - bc}, \quad \text{and} \quad B = \frac{ct_0 + d}{ad - bc}\,c \tag{30.5}$$

Also since $dx/\sqrt{r(x)} \to dt/\sqrt{\varphi(t)}$, the integral (30.4) becomes

$$\int \frac{dx}{(x - x_0)\sqrt{r(x)}} = A\int \frac{dt}{(t - t_0)\sqrt{\varphi(t)}} + B\int \frac{dt}{\sqrt{\varphi(t)}} \tag{30.6}$$

If x_0 is not a root of $r(x)$, then t_0 cannot be a root of $\varphi(t)$, so that the above expression shows that in general an elliptic integral of the third kind goes over into a linear combination of an integral of the third kind and an integral of the first kind under a linear transformation. On the

other hand, if x_0 is a root of $r(x)$, then t_0 is a root of $\varphi(t)$, so that it is at once clear that under a linear transformation an integral of the second kind will go over into a linear combination of integrals of the second and first kinds.

From the above considerations it is clear that when a linear transformation of any of the standard integrals E_1, E_0, E_{-1}, is made, no more complicated integrals will arise requiring further reduction.

31. Normal forms when $r(x)$ has real roots. In order to express elliptic integrals in the several normal forms which have been tabulated, one seeks a linear transformation which transforms the polynomial $r(x)$ into a particularly simple polynomial $\varphi(t)$. In doing this it will be supposed that the roots of $r(x)$ are real and denoted by e_1, e_2, e_3, e_4, so that

$$r(x) = (x - e_1)(x - e_2)(x - e_3)(x - e_4)$$

A linear transformation to the t axis is made, so that the points $x = e_1$, $x = e_2$, $x = e_3$ on the x axis go over into the points $t = 0$, $t = 1$, $t = \infty$, respectively. From the considerations in the preceding section it is easy to show that the transformation which does this is

$$t = \frac{e_2 - e_3}{e_2 - e_1}\frac{x - e_1}{x - e_3} = q\frac{x - e_1}{x - e_3} \tag{31.1}$$

Solving for x, one finds

$$x = \frac{e_3 t - q e_1}{t - q} \tag{31.2}$$

and

$$dx = \frac{(e_1 - e_3)(e_2 - e_3)}{(e_2 - e_1)(t - q)^2}\,dt \tag{31.3}$$

Using (31.2), the factors entering into $r(x)$ become

$$\begin{aligned} x - e_1 &= \frac{(e_3 - e_1)}{t - q}\,t \\ x - e_2 &= \frac{(e_3 - e_2)t + q(e_2 - e_1)}{t - q} \\ x - e_3 &= \frac{q(e_3 - e_1)}{t - q} \\ x - e_4 &= \frac{(e_3 - e_4)t + q(e_4 - e_1)}{t - q} \end{aligned} \tag{31.4}$$

Substituting the relations (31.4) in $r(x)$, one obtains

$$r(x) = \frac{q^2 t(e_3 - e_1)^2(e_2 - e_3)(e_4 - e_1)(1 - t)}{(t - q)^4}\left[1 - \frac{(e_4 - e_3)t}{q(e_4 - e_1)}\right]$$

Making use of the expression (31.3) we have

$$\frac{dx}{\sqrt{r(x)}} = \frac{e_1 - e_3}{e_2 - e_1}\frac{(e_2 - e_3)}{q(e_3 - e_1)}\frac{dt}{\sqrt{(e_2 - e_3)(e_4 - e_1)t(1-t)\left[1 - \dfrac{(e_4 - e_3)t}{q(e_4 - e_1)}\right]}}$$

$$= \frac{-dt}{\sqrt{Ct(1-t)(1-\lambda t)}} \tag{31.5}$$

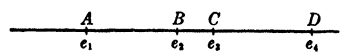
where

$$C = (e_2 - e_3)(e_4 - e_1) \quad \text{and} \quad \lambda = \frac{(e_2 - e_1)(e_4 - e_3)}{(e_2 - e_3)(e_4 - e_1)} \tag{31.6}$$

The integral

$$\int \frac{dt}{\sqrt{Ct(1-t)(1-\lambda t)}}$$

is called the Riemann normal form for the elliptic integral of the first kind. Actually this is one of six possible Riemann normal forms, since it is possible to obtain six different values of λ and six corresponding values of C by making different choices of the roots e_1, e_2, e_3, e_4, to go over into the points $t = 0, 1, \infty$. If the roots of $r(x)$ are real, it is always possible to make λ less than unity. It is convenient to represent the real numbers e_1, e_2, e_3, e_4 by distances along a line as shown below.



Now consider the following ratio of the distances, namely,

$$\frac{AB/CB}{AD/DC}$$

If the value of this ratio is -1, it is said to be harmonic, otherwise it is anharmonic. Place this ratio equal to λ and we have

$$\lambda = \frac{(e_2 - e_1)/(e_2 - e_3)}{(e_4 - e_1)/(e_4 - e_3)}$$

It is clear that whenever the above ratio is anharmonic, a value of λ can always be found which is less than unity. If the above ratio should be greater than unity, a new λ, say λ', can be defined so that $\lambda' = 1/\lambda$ by a proper adjustment of the roots e_1, e_2, e_3, e_4.

In order to transform integrals of the second and third kinds into Riemann normal forms one may start from the integral

$$\int \frac{dx}{(x - x_0)\sqrt{r(x)}}$$

This is an integral of the second kind whenever $x_0 = e_1$ is a root of $r(x)$.

In this case it is useful to make the point e_1 go over into the point $t = \infty$, and any two other roots of $r(x)$ go over into $t = 0$, and $t = 1$. When x_0 is not a root of $r(x)$ then, the integral is one of the third kind, and the transformation can be carried out in the same manner as given above for integrals of the first kind. Such transformations lead to the three Riemann normal forms.

$$\int \frac{dt}{\sqrt{Ct(1-t)(1-\lambda t)}}$$

$$\int \frac{t\,dt}{\sqrt{Ct(1-t)(1-\lambda t)}}$$

$$\int \frac{dt}{(t-t_0)\sqrt{Ct(1-t)(1-\lambda t)}}$$

where any of the six values of λ and C are used consistent with the condition $|\lambda| < 1$.

In tabulating the values of elliptic integrals, Legendre reduced the integrals to another normal form. The Legendre normal form can be obtained from the above forms by the nonlinear transformation $t = z^2$, in which case the above forms go over into

$$2\int \frac{dz}{\sqrt{C(1-z^2)(1-\lambda z^2)}}$$

$$2\int \frac{z^2\,dz}{\sqrt{C(1-z^2)(1-\lambda z^2)}}$$

$$2\int \frac{dz}{(z^2-z_0^2)\sqrt{C(1-z^2)(1-\lambda z^2)}}$$

These are the Legendre normal algebraic forms. The last two integrals are no longer integrals of the simple types due to the quadratic term. The integrals actually tabulated by Legendre are found from the above by letting $z = \sin\theta$. The first integral then becomes

$$\frac{2}{\sqrt{C}}\int \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} \quad \text{where} \quad \lambda = k^2$$

To obtain the second tabulated integral it is necessary to multiply the second integral by $-\lambda$ and add it to the first.

$$\frac{2}{\sqrt{C}}\left[\int \frac{dz}{\sqrt{(1-z^2)(1-\lambda z^2)}} - \lambda\int \frac{z^2\,dz}{\sqrt{(1-z^2)(1-\lambda z^2)}}\right]$$

$$= \frac{2}{\sqrt{C}}\int \frac{\sqrt{1-\lambda z^2}}{\sqrt{1-z^2}}\,dz$$

Making the same trigonometric substitution as before, this integral becomes

$$\frac{2}{\sqrt{C}}\int \sqrt{1 - k^2 \sin^2 \theta}\, d\theta$$

On making the trigonometric substitution in the third integral one obtains

$$\int \frac{d\theta}{(1 + p \sin^2 \theta)\sqrt{1 - k^2 \sin^2 \theta}} \quad \text{where} \quad p \neq \begin{cases} 0 \\ -1 \\ -k^2 \end{cases}$$

In this type of integral k is called the modulus of the elliptic integral. In Legendre's tables one finds the integrals

$$F(k,\theta) = \int_0^\theta \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}} \quad \text{and} \quad E(k,\theta) = \int_0^\theta \sqrt{1 - k^2 \sin^2 \theta}\, d\theta$$

for $0 \leq \theta \leq \pi/2$ and $0 < k < 1$. Legendre has also furnished a special table for the case where the upper limit is $\pi/2$.

$$K = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}}, \qquad E = \int_0^{\pi/2} \sqrt{1 - k^2 \sin^2 \theta}\, d\theta$$

32. Normal forms when $r(x)$ has complex roots. If the polynomial $r(x)$ has complex roots they must occur in conjugate pairs. Thus if e_1 is a complex root, let e_2 be its conjugate found by replacing i by $-i$ in e_1. If $r(x)$ has two further complex roots let them be denoted by e_3, e_4. In order to deal with real numbers it is convenient to transform the polynomial $r(x)$ into a polynomial $\varphi(t)$ which has the form $\varphi(t) = C(t^2 - a)(t^2 - b)$ where C, a, and b are real numbers. This means that the new polynomial should not contain odd powers of t. This may be accomplished by the linear transformation

$$x = \frac{p + qt}{1 + t}, \quad t = \frac{p - x}{x - q}, \quad dx = \frac{(q - p)dt}{(1 + t)^2}$$

provided the constants p and q are properly chosen. Writing $r(x)$ in factor form, namely,

$$r(x) = (x - e_1)(x - e_2)(x - e_3)(x - e_4)$$

we require the transformed factors $(x - e_1)$ etc. These become

$$x - e_1 = \frac{p - e_1 + (q - e_1)t}{1 + t}, \quad x - e_3 = \frac{p - e_3 + (q - e_3)t}{1 + t}$$

$$x - e_2 = \frac{p - e_2 + (q - e_2)t}{1 + t}, \quad x - e_4 = \frac{p - e_4 + (q - e_4)t}{1 + t}$$

Using these expressions, $r(x)$ becomes

$$r(x) = \frac{(p-e_1)(p-e_2)+(p-e_1)(q-e_2)t+(p-e_2)(q-e_1)t+(q-e_1)(q-e_2)t^2}{(1+t)^2} \times \frac{(p-e_3)(p-e_4)+(p-e_3)(q-e_4)t+(p-e_4)(q-e_3)t+(q-e_3)(q-e_4)t^2}{(1+t)^2}$$

The above polynomial will have the required form if p and q are determined so that the coefficients of the terms in t vanish.

$$(p - e_1)(q - e_2) + (p - e_2)(q - e_1) = 0$$
$$(p - e_3)(q - e_4) + (p - e_4)(q - e_3) = 0$$

These relations can be conveniently written in the form

$$2pq - (e_1 + e_2)(p + q) + 2e_1e_2 = 0$$
$$2pq - (e_3 + e_4)(p + q) + 2e_3e_4 = 0$$

The equations so expressed have real coefficients. To determine p and q it is convenient to regard these equations as linear in the variables pq and $p + q$. Let D be the determinant of the coefficients so that

$$D = \begin{vmatrix} 2 & -(e_1 + e_2) \\ 2 & -(e_3 + e_4) \end{vmatrix} = 2(e_1 + e_2 - e_3 - e_4)$$

Solving the equations gives

$$pq = + \frac{\begin{vmatrix} 2e_1e_2 & (e_1 + e_2) \\ 2e_3e_4 & (e_3 + e_4) \end{vmatrix}}{D} = \frac{2}{D}[e_1e_2(e_3 + e_4) - e_3e_4(e_1 + e_2)]$$

$$p + q = - \frac{\begin{vmatrix} 2 & 2e_1e_2 \\ 2 & 2e_3e_4 \end{vmatrix}}{D} = -\frac{4(e_3e_4 - e_1e_2)}{D} = 2\lambda$$

In order to find p and q it is convenient to proceed as follows.

$$(p - q)^2 = (p + q)^2 - 4pq$$
$$= \frac{16}{D^2}(e_1e_2 - e_3e_4)^2 - \frac{8}{D}[e_1e_2(e_3 + e_4) - e_3e_4(e_1 + e_2)]$$
$$= \frac{16}{D^2}(e_1 - e_3)(e_1 - e_4)(e_2 - e_3)(e_2 - e_4) = 4\mu^2$$

Then $p - q = 2\mu, \quad p + q = 2\lambda$

or $p = \mu + \lambda, \quad q = \lambda - \mu$

Since λ and μ are real, p and q are real, and finally

$$r(x) = \frac{(q-e_1)(q-e_2)(q-e_3)(q-e_4)}{(1+t)^4}\left[t^2+\frac{(p-e_1)(p-e_2)}{(q-e_1)(q-e_2)}\right]\left[t^2+\frac{(p-e_3)(p-e_4)}{(q-e_3)(q-e_4)}\right]$$

so that $\varphi(t)$ is a polynomial of the type required wherein

$$C = (q - e_1)(q - e_2)(q - e_3)(q - e_4)$$

$$a = -\frac{(p - e_1)(p - e_2)}{(q - e_1)(q - e_2)}, \qquad b = -\frac{(p - e_3)(p - e_4)}{(q - e_2)(q - e_4)}$$

The integrand of the elliptic integral will become

$$\frac{dx}{\sqrt{r(x)}} = \frac{(q - p)dt}{\sqrt{C(t^2 - a)(t^2 - b)}}$$

In order to transform the latter integrand into the Legendre algebraic normal form

$$\frac{M\,dz}{\sqrt{(1 - z^2)(1 - k^2z^2)}} \quad \text{when} \quad |k| < 1$$

it is necessary to make an appropriate change of variable for each of the following cases.

Case I

$C > 0$, $0 < a < b$, $t^2 < a$; substitute $t = z\sqrt{a}$

$C > 0$, $0 < a < b$, $t^2 > b$; substitute $t = \dfrac{\sqrt{b}}{z}$

Case II

$C > 0$, $a < 0 < b$, $t^2 > b$; substitute $t = \dfrac{\sqrt{b}}{\sqrt{1 - z^2}}$

Case III

$C > 0$, $a < 0$, $b < 0$; substitute $t = \dfrac{z\sqrt{-a}}{\sqrt{1 - z^2}}$

Case IV

$C < 0$, $0 < a < t^2 < b$; substitute $t = \dfrac{\sqrt{a}}{\sqrt{1 - (1 - a/b)z^2}}$

Case V

$C < 0$, $a < 0 < b$, $t^2 < b$; substitute $t = \dfrac{\sqrt{a}}{\sqrt{1 - z^2}}$

Case VI

$C < 0$, $a < 0$, $b < 0$.

In Case VI it is impossible to make the integrand real, since $C(t^2 - a)(t^2 - b)$ will be less than zero.

33. Definite integrals of functions containing parameters. The concept of the definite integral

$$\int_a^b f(x)dx$$

has been considered in Sec. 23. In this section a few of the properties of definite integrals of the type

$$F(x_1,x_2,\alpha) = \int_{x_1}^{x_2} f(x,\alpha)dx$$

will be considered. The limits x_1 and x_2 are values of x in the region in which the function $f(x,\alpha)$ is integrable; α is a parameter upon which $f(x,\alpha)$ depends. If $f(x,\alpha)$ be regarded as a function of the two variables x and α, it may be represented as a surface above the $x\alpha$ plane. If a

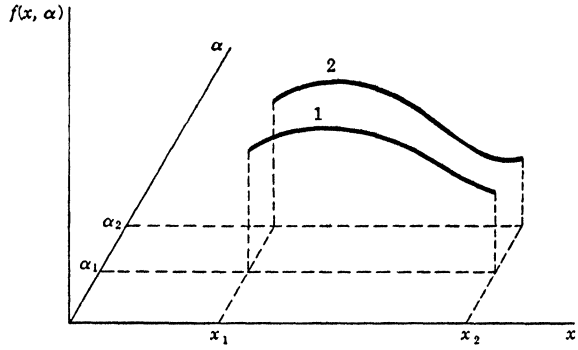


FIG. 33.1

plane is erected perpendicular to the $x\alpha$ plane and parallel to the x axis at some point $\alpha = \alpha_1$, it will cut the surface in some curve 1 as shown in Fig. 33.1. The quantity

$$F(x_1,x_2,\alpha_1) = \int_{x_1}^{x_2} f(x,\alpha_1)dx$$

represents the area under the curve 1. If α is changed slightly to the value α_2, then $F(x_1,x_2,\alpha_2)$ is the area under curve 2. If $f(x,\alpha)$ satisfies certain conditions, the area $F(x_1,x_2,\alpha)$ would be expected to vary with α in a continuous manner. In order to have a measure of the rate at which $F(x_1,x_2,\alpha)$ changes with α we require $dF/d\alpha$. Whenever $f(x,\alpha)$ is a continuous function of the two variables x and α in the range of integration and $\partial f/\partial\alpha$ is continuous in this range,

$$\frac{dF}{d\alpha} = \frac{d}{d\alpha}\int_{x_1}^{x_2} f(x,\alpha)dx = \int_{x_1}^{x_2} \frac{\partial f(x,\alpha)}{\partial\alpha}\,dx \tag{33.1}$$

and

$$\frac{\partial F}{\partial x_2} = f(x_2,\alpha) \quad \text{and} \quad \frac{\partial F}{\partial x_1} = -f(x_1,\alpha) \tag{33.2}$$

Furthermore if x_1 and x_2 are functions of α so that $x_1 = \varphi_1(\alpha)$ and $x_2 = \varphi_2(\alpha)$ and $d\varphi_1/d\alpha$ and $d\varphi_2/d\alpha$ exist,

$$\frac{dF}{d\alpha} = \frac{\partial F}{\partial \alpha} + \frac{\partial F}{\partial x_1}\cdot\frac{dx_1}{d\alpha} + \frac{\partial F}{\partial x_2}\cdot\frac{dx_2}{d\alpha}$$

$$= \int_{x_1}^{x_2} \frac{\partial f(x,\alpha)}{\partial \alpha}\, dx + f(x_2,\alpha)\frac{dx_2}{d\alpha} - f(x_1,\alpha)\frac{dx_1}{d\alpha} \qquad (33.3)$$

34. Inversion of the order of integration. It is often extremely convenient to be able to change the order of integration in an iterated integral and consequently it is important to know when this may be done. It will be supposed that $f(x,y)$ is a continuous function of the two variables within a certain finite domain. This function may be integrated with respect to y, yielding

$$G(x,y_1) = \int_b^{y_1} f(x,y)dy$$

This represents the area under the curve in which a plane perpendicular to the xy plane and parallel to the y axis through the point x cuts the surface $f(x,y)$. Then $G(x,y_1)$ will be a continuous function of the two variables x and y_1 over the domain considered. Equivalent things may be said of the function

$$F(x_1,y) = \int_a^{x_1} f(x,y)dx$$

The volume under the surface $f(x,y)$ in the domain considered will be given by either of the expressions

$$\int_a^{x_1} G(x,y_1)dx \quad \text{or} \quad \int_b^{y_1} F(x_1,y)dy$$

Since the above two expressions give the volume under the surface we have

$$\int_a^{x_1} G(x,y_1)dx = \int_b^{y_1} F(x_1,y)dy$$

or

$$\int_a^{x_1} dx \int_b^{y_1} f(x,y)dy = \int_b^{y_1} dy \int_a^{x_1} f(x,y)dx \qquad (34.1)$$

Thus with the continuity condition on $f(x,y)$ the order of integration can be changed.

The usefulness of the theorems in this and the preceding section in evaluating definite integrals will be indicated in Sec. 40.

35. Improper integrals. The term *improper integral* refers to two types. The first has to do with an interval of integration which increases without limit, and the second includes those integrals for which the integrand becomes infinite at certain points.

Integrals of the first type are defined by means of the following limit.

Given a function $f(x)$ which is bounded and integrable in a region $a \leq x \leq X$, then

$$\int_a^\infty f(x)dx \quad \text{is defined as} \quad \lim_{X\to\infty}\int_a^X f(x)dx$$

when this limit exists. When the above limit exists the integral is said to converge. On the other hand, if the above limit is infinite the integral is said to diverge.

As an example of a simple integral of this type we may consider

$$\int_1^\infty \frac{dx}{x^{3/2}}$$

To demonstrate the convergence we require the values of

$$\int_1^X \frac{dx}{x^{3/2}} = 2 - \frac{2}{X^{1/2}}$$

It is evident that

$$\lim_{X\to\infty} 2 - \frac{2}{X^{1/2}} = 2$$

so that the original integral converges to the value 2.

By applying the methods used above to the integral

$$\int_1^\infty \log x\, dx$$

it is readily found that this integral diverges, since

$$\int_1^X \log x\, dx = x \log x - x\Big]_1^X = X \log X - X + 1$$

the limit of the right-hand side as $X \to \infty$ is infinite, so that the integral is divergent.

From the above remarks it is also evident that

$$\int_{-\infty}^b f(x)dx = \lim_{X\to-\infty}\int_x^b f(x)dx$$

Furthermore if the infinite integrals

$$\int_{-\infty}^a f(x)dx \quad \text{and} \quad \int_a^\infty f(x)dx$$

are convergent,

$$\int_{-\infty}^{+\infty} f(x)dx$$

converges to the value

$$\int_{-\infty}^a f(x)dx + \int_a^\infty f(x)dx$$

36. Tests for convergence of infinite integrals. Since it is not always easy to find the integral of $f(x)$, other methods of testing for

convergence are desirable. Some of the most useful tests will be given in the following paragraphs.

Test by Direct Comparison. This test makes use of the fact that two integrable functions $\varphi(x)$ and $f(x)$ over the same domain $a \leq x < \infty$ obey the inequality

$$\int_c^\infty \varphi(x)dx \geq \int_c^\infty f(x)dx$$

whenever $\varphi(x)$ is everywhere positive and $\varphi(x) \geq |f(x)|$ in the interval $c < x < \infty$. Thus if the integral $\int_c^\infty \varphi(x)dx$ is known to converge, this guarantees the convergence of $\int_c^\infty f(x)dx$.

On the other hand, if $\varphi(x)$ is continuous in the interval $c \leq x < \infty$ and $\int_c^\infty \varphi(x)dx$ is known to diverge, and if $f(x) \geq \varphi(x) \geq 0$, the integral $\int_c^\infty f(x)dx$ certainly diverges.

Absolutely Convergent Integrals. If $f(x)$ is continuous in the interval $c \leq x < \infty$ and if the integral $\int_c^\infty |f(x)|dx$ converges, the integral $\int_c^\infty f(x)dx$ also converges. In fact the integral $\int_c^\infty f(x)dx$ is said to converge absolutely.

As an illustration of the use of the above tests, consider the integral $\int_0^\infty \frac{\sin x}{1+x^2}dx$. Since $(\sin x)/(1+x^2) \leq 1/(1+x^2)$, we have

$$\int_0^\infty \frac{\sin x}{1+x^2}dx \leq \int_0^\infty \left|\frac{\sin x}{1+x^2}\right| dx \leq \int_0^\infty \frac{dx}{1+x^2}$$

and the integral on the right is known to converge; hence the original integral converges.

Limit Tests. If $f(x)$ is continuous in the interval $a \leq x < \infty$ and if there exists a constant $k > 1$ such that $\lim_{x\to\infty} x^k f(x)$ exists, the integral $\int_a^\infty f(x)dx$ converges.

On the other hand, if $\lim_{x\to\infty} xf(x)$ exists and is not zero, the integral $\int_a^\infty f(x)dx$ certainly diverges.

Alternating Integrals. The tests for convergence thus far considered do not suffice to establish the convergence or divergence of integrals whose integrands are oscillating functions. An integral of this type is $\int_0^\infty \frac{\sin x}{x}dx$. The integrand oscillates as shown in Fig. 36.1. The method for testing the convergence may be illustrated using the above integral as an example. The range of integration is divided into those regions

in which the integrand is positive and negative. In this case

$$\int_0^\infty \frac{\sin x}{x}\,dx = \int_0^\pi \frac{\sin x}{x}\,dx + \int_\pi^{2\pi} \frac{\sin x}{x}\,dx + \ldots + \int_{(m-1)\pi}^{m\pi} \frac{\sin x}{x}\,dx + \int_{m\pi}^{(m+1)\pi} \frac{\sin x}{x}\,dx + \ldots$$

The right-hand side is an alternating series, that is, one whose terms are alternately positive and negative, and it remains to demonstrate the convergence of this series to establish the convergence of the original integral. Such a series is known to converge if $|\,u_{m+1}\,| < |\,u_m\,|$ for all m greater than a fixed number N, and if the $\lim_{m\to\infty} u_m = 0$, where u_m is the value of the mth term.

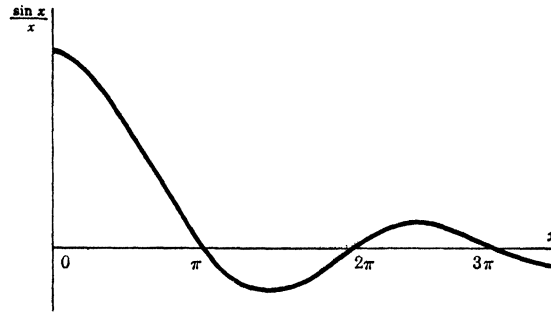


FIG. 36.1

It is obvious from the figure that the area under each arch of the curve steadily decreases, so that the condition $|\,u_{m+1}\,| < |\,u_m\,|$ is fulfilled for all m. This can easily be verified analytically by making the change of variables $x = (m-1)\pi + t$ and $x = m\pi + t$ in the mth and $(m+1)$st integrals and noting the inequality

$$\left|\int_0^\pi \frac{\sin t}{m\pi + t}\,dt\right| < \left|\int_0^\pi \frac{\sin t}{(m-1)\pi + t}\,dt\right| \text{ since } \frac{\sin t}{m\pi + t} < \frac{\sin t}{(m-1)\pi + t}$$

for the interval $0 \le t \le \pi$. The second condition $\lim_{m\to\infty} u_m = 0$ is satisfied, since

$$\int_{(m-1)\pi}^{m\pi} \frac{\sin x}{x}\,dx \le \int_0^\pi \frac{\sin t}{m\pi + t}\,dt < \frac{1}{m} \quad \text{and} \quad \lim_{m\to\infty} \frac{1}{m} = 0$$

37. Uniform convergence of infinite integral with a parameter. In Sec. 33 the definite integral over a finite range containing a parameter

was considered. The corresponding infinite integral (if convergent) is defined as

$$\int_a^\infty f(x,\alpha)dx = \lim_{X\to\infty} \int_a^X f(x,\alpha)dx$$

It may well happen that the limit on the right exists only for certain values of α. As an illustration of this situation consider the integral $\int_0^\infty e^{-\alpha x} \cos x\, dx$. By applying the methods described in the preceding sections it may readily be shown that this integral converges for all positive values of α but fails to converge for $\alpha \leq 0$. If the integral $\int_a^\infty f(x,\alpha)dx$ converges for all values of α in a certain domain, the integral is said to converge uniformly with regard to α in this range. Stated in analytical language, the integral $\int_a^\infty f(x,\alpha)dx$ converges uniformly over all α's in S if for any arbitrary positive number ϵ there exists a number N, *independent of* α, such that

$$\left| \int_{X'}^\infty f(x,\alpha)dx \right| < \epsilon \quad \text{when} \quad X' \geq N.$$

38. Tests for uniform convergence of an infinite integral. There are several tests that can be applied to see if the integral under consideration is uniformly convergent. Some of the useful tests are given below.

De La Vallee Poussin Test. This test is a kind of comparison test. Thus if a function $\varphi(x)$ is known for which $\varphi(x) \geq |f(x,\alpha)|$ for all α in a certain domain S, and the integral $\int_a^\infty \varphi(x)dx$ converges, the integral $\int_a^\infty f(x,\alpha)dx$ converges uniformly over the domain S.

The Change of Variable Test. This test can best be illustrated by an example. Suppose it is desired to test the integral $\int_0^\infty \frac{\sin \alpha x}{x} dx$ for uniform convergence. We require the values of α for which $\lim_{X\to\infty} \int_0^X \frac{\sin \alpha x}{x} dx$ exists. One then makes the change of variable $y = \alpha x$ so that

$$\int_0^X \frac{\sin \alpha x}{x} dx = \int_0^{\alpha X} \frac{\sin y}{y} dy$$

Since the integral $\int_0^\infty \frac{\sin y}{y} dy$ is convergent, the original integral $\int_0^\infty \frac{\sin \alpha x}{x} dx$ is uniformly convergent for all values of α.

Test by Integration by Parts. Again the values of α for which

$\lim_{X\to\infty}\int_a^X f(x,\alpha)dx$ exists are desired. In this test the integral $\int_a^X f(x,\alpha)dx$ is integrated once or more by parts, so that

$$\int_a^X f(x,\alpha)dx = \psi(X,\alpha) + \int_a^X \varphi(x,\alpha)dx$$

Then if $\psi(X,\alpha) \to 0$ uniformly as $X \to \infty$ and if $\int_a^\infty \varphi(x,\alpha)dx$ converges uniformly, $\int_a^\infty f(x,\alpha)dx$ converges uniformly.

39. Theorems concerning uniformly convergent integrals. The following theorems having to do with differentiation under the sign of integration and inversion of the order of integration are of very great importance in the evaluation of definite integrals.

1. If $\int_a^\infty f(x,\alpha)dx$ converges uniformly when α lies in a domain S and if $f(x,\alpha)$ is a continuous function of both variables when $x \geq a$ and α lies in S, then $\int_a^\infty f(x,\alpha)dx$ is a continuous function of α in S.

2. Inversion of the order of integration in infinite integrals. If $f(x,\alpha)$ satisfies the same conditions as in 1 and if $\alpha_1 \leq \alpha \leq \alpha_2$ lie in S, then

$$\int_{\alpha_1}^{\alpha_2} d\alpha \int_a^\infty f(x,\alpha)dx = \int_a^\infty dx \int_{\alpha_1}^{\alpha_2} f(x,\alpha)dx \tag{39.1}$$

3. Differentiation under the sign of integration in an infinite integral. The equation

$$\frac{d}{d\alpha}\int_a^\infty f(x,\alpha)dx = \int_a^\infty \frac{\partial}{\partial\alpha} f(x,\alpha)dx \tag{39.2}$$

holds when (1) the integral on the left converges, (2) the integral on the right converges *uniformly* over the values of α under consideration and (3) its integrand is a continuous function of *both* variables.

40. Methods of evaluating definite integrals. In order to show the usefulness of the preceding theorems in evaluating definite integrals, several examples will be considered.

EXAMPLE 1. Evaluate

$$\int_0^\infty x^n e^{-\alpha x}\,dx \qquad \alpha > 0.$$

Since the function under the integral sign is a continuous function of x and α in the domains under consideration and the integral converges uniformly for $\alpha > 0$, we may write

$$\int_0^\infty x^n e^{-\alpha x}\,dx = (-1)^n \int_0^\infty \left(\frac{\partial^n}{\partial\alpha^n} e^{-\alpha x}\right)dx = (-1)^n \frac{d^n}{d\alpha^n}\int_0^\infty e^{-\alpha x}\,dx$$

The last step is made in accordance with eq. (39.2). The integral on

the right can readily be evaluated so that

$$(-1)^n \frac{d^n}{d\alpha^n} \int_0^\infty e^{-\alpha x}\, dx = (-1)^n \frac{d^n}{d\alpha^n}\left(\frac{1}{\alpha}\right) = \frac{n!}{\alpha^{n+1}}$$

so that finally

$$\int_0^\infty x^n e^{-\alpha x}\, dx = \frac{n!}{\alpha^{n+1}} \tag{40.1}$$

In passing it is worth while to point out that if $\alpha = 1$, eq. (40.1) furnishes a definite integral representation of $n!$. The integral itself defines a function of n even when n is a positive number not necessarily an integer. This more general function of n is called the gamma function, and is written

$$\Gamma(n+1) = \int_0^\infty x^n e^{-x}\, dx \qquad n \geq 0 \tag{40.2}$$

Integrating the integral once by parts yields

$$\Gamma(n+1) = n\int_0^\infty x^{n-1} e^{-x}\, dx = n\Gamma(n) \tag{40.3}$$

which is the well-known recurrence formula for the gamma function. When n is a positive integer, this is just the formula satisfied by $n!$. Other representations of the gamma function will be given later which define it at all points in the complex plane except at negative real integer values.

Equation (40.1) furnishes a convenient integral representation of $1/\alpha^{n+1}$.

EXAMPLE 2. For later use we consider the following two integrals.

$$K_1 = \int_a^b e^{px} \cos qx\, dx \qquad K_2 = \int_a^b e^{px} \sin qx\, dx \tag{40.4}$$

These may be evaluated conveniently by considering both integrals simultaneously. Integrating both by parts once yields

$$K_1 = \frac{e^{px} \sin qx}{q}\Big]_a^b - \frac{p}{q} K_2$$

$$K_2 = -\frac{e^{px} \cos qx}{q}\Big]_a^b + \frac{p}{q} K_1$$

Solving these two equations for K_1 and K_2 gives

$$K_1\left(1 + \frac{p^2}{q^2}\right) = \left[\frac{e^{px} \sin qx}{q} + \frac{pe^{px} \cos qx}{q^2}\right]_a^b$$

$$K_2\left(1 + \frac{p^2}{q^2}\right) = \left[\frac{pe^{px} \sin qx}{q^2} - \frac{e^{px} \cos qx}{q}\right]_a^b$$

or

$$\left.\begin{aligned} K_1 &= e^{px} \frac{p \cos qx + q \sin qx}{p^2 + q^2}\Big]_a^b \\ K_2 &= e^{px} \frac{p \sin qx - q \cos qx}{p^2 + q^2}\Big]_a^b \end{aligned}\right\} \tag{40.5}$$

Two important special cases of these integrals are those for which $a = 0$ and $b = \infty$ and p is negative. Thus

$$\left.\begin{aligned} \int_0^\infty e^{-px} \cos qx\, dx &= \frac{p}{p^2 + q^2} \qquad p > 0 \\ \int_0^\infty e^{-px} \sin qx\, dx &= \frac{q}{p^2 + q^2} \qquad p > 0 \end{aligned}\right\} \tag{40.6}$$

EXAMPLE 3. Consider the integral $\int_0^\infty e^{-px} \frac{\sin qx}{x}\, dx$ which converges uniformly for $p \geq 0$. If it were not for the x in the denominator, the integral would correspond to one of the integrals under (40.6). This gives us a clue to a simple method of procedure. The x in the denominator can be removed for the moment by making use of eq. (40.1). This gives

$$\frac{1}{x} = \int_0^\infty e^{-xt}\, dt$$

Substituting this in the integral under consideration we have

$$\int_0^\infty e^{-px} \frac{\sin qx}{x}\, dx = \int_0^\infty e^{-px} \sin qx\, dx \int_0^\infty e^{-xt}\, dt$$

The right-hand integral converges uniformly, and the order of integration may be inverted, so that one has

$$\int_0^\infty dt \int_0^\infty e^{-(p+t)x} \sin qx\, dx = \int_0^\infty \frac{q}{(p+t)^2 + q^2}\, dt$$

when use is made of the result (40.6). A change of variable $s = p + t$ in the right-hand integral gives

$$q \int_p^\infty \frac{ds}{s^2 + q^2}$$

A further change of variable $s = q \tan \theta$ reduces the integral to

$$\int_{\tan^{-1} p/q}^{\pi/2} \frac{\sec^2 \theta\, d\theta}{\sec^2 \theta} = \int_{\tan^{-1} p/q}^{\pi/2} d\theta = \frac{\pi}{2} - \tan^{-1} \frac{p}{q}$$

so that finally

$$\int_0^\infty e^{-px} \frac{\sin qx}{x}\, dx = \frac{\pi}{2} - \tan^{-1} \frac{p}{q} \tag{40.7}$$

The special case of this integral which is obtained by placing p equal to zero and $q > 0$ is of interest, as the following will show.

$$\int_0^\infty \frac{\sin qx}{x}\, dx = \frac{\pi}{2} \quad \text{when} \quad q > 0 \tag{40.8}$$

Since the sine is an odd function,

$$\int_0^\infty \frac{\sin qx}{x}\,dx = -\frac{\pi}{2} \quad \text{when} \quad q < 0 \tag{40.9}$$

and

$$\int_0^\infty \frac{\sin qx}{x}\,dx = 0 \quad \text{when} \quad q = 0$$

Thus the definite integral is a discontinuous function of q, as shown in Fig. 40.1. This integral is occasionally of use when an expression is desired which gives a positive number for $q > 0$ and 0 for $q < 0$. Such an expression is obviously

$$\frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin qx}{x}\,dx = \begin{cases} 1 & \text{for} \quad q > 0 \\ 0 & \text{for} \quad q < 0 \end{cases}$$

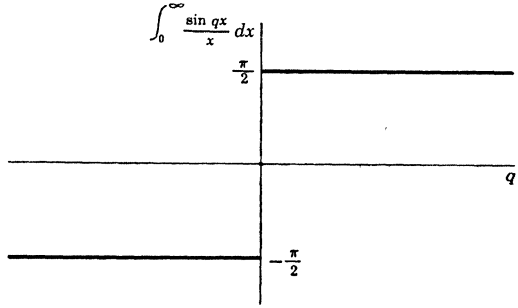


FIG. 40.1

41. Definite integrals with integrands which become infinite at certain points. First let it be supposed that the integrand becomes infinite only at the lower limit of integration. Thus the integrand $f(x)$ is bounded and integrable for the range $a + \epsilon \leq x \leq b$, where $\epsilon > 0$ and $f(x) \to \infty$ as $x \to a$. The integral $\int_a^b f(x)dx$ is said to converge if the $\lim_{\epsilon \to 0} \int_{a+\epsilon}^b f(x)dx$ exists. Then

$$\int_a^b f(x)dx = \lim_{\epsilon \to 0} \int_{a+\epsilon}^b f(x)dx \tag{41.1}$$

Similar remarks apply to the case where $f(x)$ becomes infinite at $x = b$.

Next suppose that there is a point $x = c$ lying in the interval $[ab]$ so that $a < c < b$ at which $f(x)$ becomes infinite. The definition of the integral $\int_a^b f(x)dx$ in this case becomes

$$\int_a^b f(x)dx = \lim_{\substack{\epsilon \to 0 \\ \epsilon' \to 0}} \left\{ \int_a^{c-\epsilon} f(x)dx + \int_{c+\epsilon'}^b f(x)dx \right\} \tag{41.2}$$

where ϵ and ϵ' are allowed to approach zero *independently*.

It may happen that no limit exists when ϵ and ϵ' approach zero independently, yet a limit does exist when $\epsilon' = \epsilon$ and ϵ is allowed to approach zero. When such a limit exists it is called the Cauchy principal value of the integral. This is sometimes designated by

$$P\int_a^b f(x)dx = \lim_{\epsilon \to 0} \left\{ \int_a^{c-\epsilon} f(x)dx + \int_{c+\epsilon}^b f(x)dx \right\}$$

The two types of integrals may be illustrated by the following examples.

$$\int_{-1}^{+1} \frac{dx}{|x|^{2/3}} = \lim_{\substack{\epsilon \to 0 \\ \epsilon' \to 0}} \left\{ \int_{-1}^{0-\epsilon} \frac{dx}{|x|^{2/3}} + \int_{0+\epsilon'}^{1} \frac{dx}{|x|^{2/3}} \right\}$$

but $$\lim_{\epsilon \to 0} \int_{-1}^{-\epsilon} \frac{dx}{|x|^{2/3}} = \lim_{\epsilon \to 0} \int_{\epsilon}^{1} \frac{dx}{x^{2/3}} = \lim_{\epsilon \to 0} (3 - 3\epsilon^{1/3}) = 3$$

and $$\lim_{\epsilon' \to 0} \int_{\epsilon'}^{1} \frac{dx}{|x|^{2/3}} = 3$$

Since these limits exist,

$$\int_{-1}^{+1} \frac{dx}{|x|^{2/3}} = 6$$

As an example of the second type consider $\int_{-1}^{2} dx/x$. If this exists in the first sense, the limits

$$\lim_{\epsilon \to 0} \int_{-1}^{-\epsilon} \frac{dx}{x} \quad \text{and} \quad \lim_{\epsilon' \to 0} \int_{\epsilon'}^{2} \frac{dx}{x}$$

must exist. But

$$\lim_{\epsilon \to 0} \int_{-1}^{-\epsilon} \frac{dx}{x} = -\lim_{\epsilon \to 0} \int_{\epsilon}^{1} \frac{dx}{x} = +\lim_{\epsilon \to 0} \log \epsilon = -\infty$$

$$\lim_{\epsilon' \to 0} \int_{\epsilon'}^{2} \frac{dx}{x} = \lim_{\epsilon' \to 0} \log \frac{2}{\epsilon'} = \infty$$

Thus the integral does not converge. It is readily shown, however, that it has a Cauchy principal value, since

$$\lim_{\epsilon \to 0} \int_{-1}^{-\epsilon} \frac{dx}{x} + \int_{\epsilon}^{2} \frac{dx}{x} = \lim_{\epsilon \to 0} \{\log \epsilon + \log 2 - \log \epsilon\}$$

$$= \log 2$$

Thus $$P\int_{-1}^{2} \frac{dx}{x} = 2$$

42. Tests for convergence. Simple tests can be given for the convergence of the above type of improper integral as well as for the type taken up in Sec. 35.

Comparison Test. Suppose $f(x) \to \infty$ as $x \to a$. Then if a positive function $\varphi(x)$ can be found which is continuous in the interval $a < x \leq c \leq b$,

and $\varphi(x) \geq f(x) \geq 0$ for $a < x \leq c$, and $\int_a^c \varphi(x)dx$ converges, so also does the integral $\int_a^b f(x)dx$. On the other hand if $f(x) \geq \varphi(x)$ for $a < x \leq c$, and it is known that $\int_a^c \varphi(x)dx$ diverges, the integral $\int_a^b f(x)$ certainly diverges.

Limit Test. Again suppose that $f(x) \to \infty$ when $x \to a$. Then if there exists a positive constant $0 < l < 1$ such that $\lim_{x \to a} (x - a)^l f(x)$ exists, $\int_a^b f(x)dx$ converges.

On the other hand, if the $\lim_{x \to a} (x - a)f(x)$ exists and is equal to L and $L \neq 0$, then $\int_a^b f(x)dx$ diverges.

The modification of the above tests when the infinite point is an interior point of the interval is evident.

43. Stieltjes integral. In Sec. 23, the Riemann integral of a bounded function $f(x)$, over an interval $[ab]$, with respect to x was defined. It is sometimes convenient to formulate problems with the help of a more general integral, namely, the integral of a bounded function $f(x)$, over an interval, with respect to another bounded function $\varphi(x)$ over the same interval. Such an integral is defined as follows. Let $f(x)$ and $\varphi(x)$ be two bounded functions over the interval $[ab]$ and let this interval be divided into n subintervals, the length of the rth interval being $x_r - x_{r-1}$, and the length of the longest subinterval being δ. Select an arbitrary point in the rth interval, designate it ξ_r, and form the sum

$$S_\delta^\sigma = \sum_{r=0}^{n} f(\xi_r)[\varphi(x_r) - \varphi(x_{r-1})]$$

where $x_0 = a$ and $x_n = b$. The value of the above sum will depend on the scheme of subdivision σ of the interval $[ab]$ and on the choice of the points ξ_r. It may happen that S_δ^σ approaches a limit S, as the number n of subdivisions is indefinitely increased such that $\delta \to 0$, which is independent of the scheme of subdivision and the choice of the points ξ_r. Should this be the case the limit

$$\lim_{\delta \to 0} \sum_{r=0}^{n} f(\xi_r)[\varphi(x_r) - \varphi(x_{r-1})] = S = \int_a^b f(x)d\varphi(x)$$

is called the Stieltjes integral of $f(x)$ with respect to $\varphi(x)$.

The Stieltjes integral of $f(x)$ with respect to $\varphi(x)$ over the interval $[ab]$ will exist provided that both $f(x)$ and $\varphi(x)$ are of bounded variation

in the interval $[ab]$ (see Sec. 53) and that the variation of $\varphi(x)$ over the set of points of discontinuity of $f(x)$ is zero.

Various important properties of Stieltjes integrals are briefly listed below.

(1) If the Stieltjes integral $\int_a^b f(x)d\varphi(x)$ exists, so also does

$$\int_a^b \varphi(x)df(x),$$

and $$\int_a^b f(x)d\varphi(x) + \int_a^b \varphi(x)df(x) = f(b)\varphi(b) - f(a)\varphi(a)$$

(2) If $\varphi(x) = \varphi_1(x) - \varphi_2(x)$ and the Stieltjes integrals

$$\int_a^b f(x)d\varphi_1(x) \quad \text{and} \quad \int_a^b f(x)d\varphi_2(x) \quad \text{exist,}$$

$$\int_a^b f(x)d\varphi(x) \quad \text{exists and}$$

$$\int_a^b f(x)d\varphi(x) = \int_a^b f(x)d\varphi_1(x) - \int_a^b f(x)d\varphi_2(x)$$

(3) If the Stieltjes integral $\int_a^b f(x)d\varphi(x)$ exists, the mean value theorem holds in the form

$$\int_a^b f(x)d\varphi(x) = f(\xi)[\varphi(b) - \varphi(a)]$$

where $a \leq \xi \leq b$.

(4) If $\varphi(x)$ has a continuous first derivative and

$$\int_a^b f(x)d\varphi(x) \quad \text{exists,} \quad \int_a^b f(x)d\varphi(x) = \int_a^b f(x)\varphi'(x)dx$$

and the integral on the right is an ordinary Riemann integral.

(5) If $f(x)$ has a continuous first derivative and the Stieltjes integral $\int_a^b f(x)d\varphi(x)$ exists, then from (2) and (4),

$$\int_a^b f(x)d\varphi(x) = f(b)\varphi(b) - f(a)\varphi(a) - \int_a^b \varphi(x)f'(x)dx$$

where the integral on the right is an ordinary Riemann integral.

Stieltjes integrals are of especial interest in applications where $\varphi(x)$ is discontinuous and the differential $d\varphi(x)$ consequently does not exist. In particular, Stieltjes integrals are of considerable use in the theory of probability in connection with distribution functions.

SUGGESTED REFERENCES

W. F. Osgood, *Advanced Calculus*, New York, The Macmillan Company, 1925.

E. T. Whittaker and G. N. Watson, *Modern Analysis*, London, Cambridge University Press, 1920.

R. S. Burington and C. C. Torrance, *Higher Mathematics*, New York, McGraw-Hill Book Company, Inc., 1939.

E. J. P. Goursat and E. R. Hedrick, *Mathematical Analysis*, Boston, Ginn & Company, 1904.

W. B. Fite, *Advanced Calculus*, New York, The Macmillan Company, 1938.

PROBLEMS

26.1. How may the integral $\int_{-1}^{+1} dx$ be properly evaluated by changing the variable as defined by $x = t^{3/2}$?

26.2. (a) A common method of evaluating

$$\int \frac{d\theta}{a + b \sin \theta} \qquad b < a$$

is to make a change of variable $x = \tan \theta/2$. Show that

$$\int \frac{d\theta}{a + b \sin \theta} = \frac{2 \sec \alpha}{a} \tan^{-1} \left(\sec \alpha \tan \frac{\theta}{2} + \tan \alpha \right)$$

where $\tan \alpha = a/\sqrt{a^2 - b^2}$.

(b) Evaluate $\int_0^{2\pi} \frac{d\theta}{a + b \cos \theta}$ $\qquad b < a$.

27.1. Apply Hermite's method to the evaluation of

$$\int \frac{dx}{x(x^2 + 1)^2}.$$

27.2. Use Hermite's method to determine the value of the integral

$$\int \frac{x^2\, dx}{x^4 + 3x^3 - 2x^2 + 12x + 16}.$$

27.3. Evaluate the integral $\int \frac{x^5\, dx}{x^4 + 2x^3 - 2x - 1}$.

27.4. Evaluate by Hermite's method

$$\int \frac{x^4 - x^3 - 3x^2 - x}{(x^2 + 1)^3}\, dx.$$

28.1. Perform the integration

$$\int \frac{x\, dx}{(1 + x)^{1/2} - (1 + x)^{1/3}}.$$

28.2. Find the value of the integral

$$\int \frac{x^3\, dx}{(1 + 2x)^{1/2}(\frac{1}{2} + x)^{1/4}}.$$

28.3. Find the value of the integral

$$\int \left(\frac{x + \frac{1}{2}}{3x + 1} \right)^{1/2} \left(\frac{2x + 1}{x + \frac{1}{4}} \right)^{1/3} dx.$$

28.4. Integrate $\int x^2(1 + x^{1/5})dx$.

28.5. Integrate $\int x^2(2 + x^{3/2})^{1/4}\, dx$.

28.6. Integrate $\int x^{3/2}(x^{3/2} + 1)^{1/3}\, dx$.

28.7. Obtain the appropriate value of

$$\int \frac{dx}{a + b \cos x} \qquad b \neq a$$

for the conditions

(1) $\dfrac{(a+b)}{(a-b)} > 0.$ (2) $(a+b) > 0, \quad (a-b) > 0.$

(3) $\dfrac{(a+b)}{(a-b)} < 0.$ (4) $(b+a) > 0, \quad (b-a) > 0.$

28.8. Show that

$$\int_0^{\pi/2} \sin^n x\, dx = \frac{1\cdot 3\cdot 5\cdot \ldots \cdot (n-1)}{2\cdot 4\cdot 6\cdot \ldots \cdot n}\cdot\frac{\pi}{2} \quad \text{if } n \text{ is even.}$$

$$= \frac{2\cdot 4\cdot 6\cdot \ldots \cdot (n-1)}{1\cdot 3\cdot 5\cdot \ldots \cdot n} \quad \text{if } n \text{ is odd.}$$

28.9. Integrate $\int x^3(1+x^2)^{1/3}\, dx$.

28.10. Integrate $\int x(1+x^3)^{1/2}\, dx$.

28.11. Integrate $\dfrac{x^3\, dx}{(1+2x)^{1/2} - (\frac{1}{2}+x)^{1/4}}$.

28.12. Integrate $\displaystyle\int \frac{dx}{a + b\cos x + c\sin x}$.

29.1. Compute the length of arc between two arbitrary points on the ellipse $x^2/a^2 + y^2/b^2 = 1$.

29.2. Express the integral

$$\int \frac{5x^3 - 3x^2 + x + 1}{\sqrt{x^3 - x^2 - 4x + 4}}\, dx$$

in terms of elliptic integrals of the first, second, or third kinds.

29.3. Express the integral

$$\int \frac{x^4\, dx}{\sqrt{(x-1)(x-2)(x+1)(x+2)}}$$

in terms of elliptic integrals of the first, second, or third kinds.

29.4. Express the integral

$$\int \frac{(x-1)dx}{\sqrt{x^4 - x^3 + x^2 + x - 2}}$$

in terms of elliptic integrals of the first, second, or third kinds.

31.1. Reduce the elliptic integrals of Problem 29.2 to the Riemann and Legendre normal forms.

31.2. Reduce the elliptic integrals of Problem 29.3 to the Riemann and Legendre normal forms.

31.3. Evaluate numerically $\displaystyle\int_{\pi/4}^{\pi/2} \frac{d\theta}{\sqrt{2 + \cos^2\theta}}$.

31.4. Reduce to Legendre normal form

$$\int_0^2 \frac{dy}{\sqrt{y^4 + 2y^3 + 8y^2 + 8y + 16}}.$$

32.1. Reduce the elliptic integrals of Problems 29.3 and 29.4 to Legendre normal forms.

32.2. Express the integral

$$\int_0^b \frac{\sqrt{b-y}}{(1+\frac{1}{4}y)^{3/4}}\, dy$$

in terms of elliptic integrals of the Legendre normal forms.

32.3. The length of time required for a pendulum to swing through the angle α is given by

$$t = \sqrt{\frac{l}{2g}} \int \frac{d\theta}{\sqrt{\cos\theta - \cos\alpha}}.$$

Reduce this integral to the Legendre normal form.

33.1. Verify the equation

$$\frac{d}{d\alpha}\int_0^1 e^{-\alpha x}\,dx = \int_0^1 \frac{\partial}{\partial\alpha} e^{-\alpha x}\,dx.$$

33.2. Compute

$$\frac{d}{d\alpha}\int_{-\alpha}^{+\alpha} \frac{dx}{1+\alpha^2+x}$$

by using formula (33.3) and demonstrate the correctness of the result.

33.3. Show that

$$\frac{\partial^2}{\partial\alpha^2}\int_0^\infty e^{-\alpha x}\sin\beta x\,dx = -\frac{\partial^2}{\partial\beta^2}\int_0^\infty e^{-\alpha x}\sin\beta x\,dx.$$

33.4. Compute $\dfrac{d}{d\alpha}\displaystyle\int_0^\alpha \sqrt{\alpha^2 - x^2}\,dx.$

34.1. Ascertain by computation whether the integrals

$$\int_0^1 dx\int_0^1 \frac{(x-y)}{(x+y)^3}\,dy \quad \text{and} \quad \int_0^1 dy\int_0^1 \frac{x-y}{(x+y)^3}\,dx$$

are equal.

36.1. Determine the domain of convergence of the following integrals and find their values.

(1) $\displaystyle\int_0^\infty e^{-\alpha x}\,dx.$ (2) $\displaystyle\int_0^\infty \frac{\alpha\,dx}{\alpha^2+x^2}.$

(3) $\displaystyle\int_0^\infty \frac{dx}{x^\rho}, \quad (c > 0).$

36.2. Test the convergence of the following integrals by means of the comparison test.

(1) $\displaystyle\int_0^\infty \frac{dx}{\sqrt{1+x^4}}.$ (2) $\displaystyle\int_2^\infty \frac{x\,dx}{\sqrt{x^4-1}}.$

(3) $\displaystyle\int_2^\infty \frac{x^6\,dx}{x^7-1}.$ (4) $\displaystyle\int_0^\infty e^{-x^2}\,dx.$

36.3. By using the limit test ascertain whether the following integrals are convergent or divergent.

(1) $\displaystyle\int_2^\infty \frac{dx}{x\log x}.$ (2) $\displaystyle\int_2^\infty \frac{dx}{x(\log x)^2}.$

(3) $\displaystyle\int_0^\infty \frac{\cos x\,dx}{\sqrt{1+x^p}}.$ (4) $\displaystyle\int_0^\infty \frac{\sin^2 x}{x}\,dx.$

36.4. Ascertain whether the following integrals converge.

(1) $\displaystyle\int_0^\infty \cos x^2\,dx.$ (2) $\displaystyle\int_1^\infty \frac{\sin(1/x)}{x}\,dx.$

(3) $\displaystyle\int_0^\infty \sin\sqrt{x}\,dx.$ (4) $\displaystyle\int_0^\infty \frac{x\cos x}{a^2+x^2}\,dx.$

36.5. Ascertain whether the following integrals converge.

(1) $\int_0^1 \frac{dx}{\sqrt{x}\log x}$. (2) $\int_0^1 \frac{\log x}{1-x}\,dx$.

36.6. For what n does $\int_0^\infty x^{n-1}e^{-x}\,dx$ converge?

38.1. Determine the range of the parameter α for which the following integrals converge uniformly.

(1) $\int_0^\infty e^{-\alpha x}x^2\,dx$. (2) $\int_0^\infty \frac{\alpha x\,dx}{\alpha^2+x^2}$.

(3) $\int_0^\infty \sin \alpha^2 x^3\,dx$. (4) $\int_0^\infty \sin(x^2-\alpha x)\,dx$.

39.1. Ascertain whether the following inversions of the order of integration are permissible.

(1) $\int_0^\infty dx\int_0^\infty \frac{e^{-xt}\sin xt}{x}\,dt = \int_0^\infty dt\int_0^\infty \frac{e^{-xt}\sin xt}{x}\,dx$.

(2) $\int_0^\infty dy\int_0^\infty \frac{e^{-x}\sin yx}{y^{1/2}}\,dx = \int_0^\infty dx\int_0^\infty \frac{e^{-x}\sin yx}{y^{1/2}}\,dy$.

40.1. Evaluate the integral $\int_0^\infty x^2e^{-px}\sin x\,dx$.

40.2. Evaluate the integral $\int_0^\infty e^{-x}(x^2+2x+1)dx$. *Hint*: Evaluate the integral $\int_0^\infty e^{-px}\,dx$, perform the requisite number of differentiations with respect to p, and set $p=1$ in the final result.

40.3. Evaluate the integrals

$$\int_0^\infty \frac{\cos qx\,dx}{1+x^2} \quad \text{and} \quad \int_0^\infty \frac{x\cos qx\,dx}{1+x^2}.$$

40.4. Evaluate the integral $\int_0^\infty \frac{e^{-px}\sin^2 qx\,dx}{x^2}$.

42.1. Determine the range of values of n for which

$$\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}\,dx \quad \text{converges.}$$

42.2. For what values of p and q does the integral

$$\int_0^{\pi/2} \sin^{p-1}x\cos^{q-1}x\,dx \quad \text{converge?}$$

42.3. Examine the convergence of the following integrals.

(1) $\int_0^1 \frac{dx}{\sqrt{x}\log x}$. (2) $\int_{-\infty}^{+\infty} \frac{e^{-x^2}\,dx}{\log x^2}$.

(3) $\int_0^1 \frac{\log x}{1-x}\,dx$. (4) $\int_{-1}^{+1} e^{-1/x}\,dx$.

Chapter 4

SPACE GEOMETRY

44. Space Curves. There are several ways in which a space curve may be defined. In the first place if $f(t)$, $g(t)$, and $h(t)$ are continuous single-valued functions of t in an interval $\alpha \leq t \leq \beta$ and the values of these functions are denoted by x, y, and z, so that

$$x = f(t), \qquad y = g(t), \qquad z = h(t) \tag{44.1}$$

then as t traverses the range $\alpha \leq t \leq \beta$ the aggregate of points (x,y,z) constitute a space curve.

It may happen that for two distinct values t_1 and t_2 of t, the associated values (x_1,y_1,z_1) and (x_2,y_2,z_2) are the same, so that these points coincide. In this case the point is said to be a double point of the curve. As an illustration, the plane curve defined by the relations $x = \sin t$ and $y = \sin 2t$ shown in Fig. 44.1 has a double point $x = 0$, $y = 0$ corresponding to $t = 0$ and $t = \pi$.

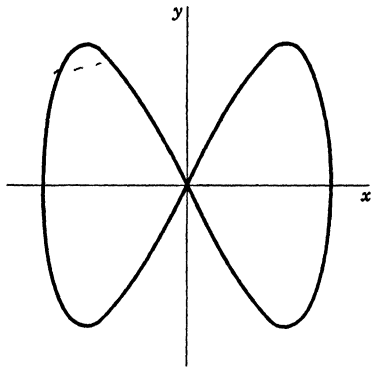


FIG. 44.1

A curve possessing no double points is called a simple curve. By eliminating t from eq. 44.1, the same curve may be defined by equations of the form

$$y = \varphi(x), \qquad z = \psi(x) \tag{44.2}$$

A space curve may also be specified by the curve of intersection of two surfaces. The two intersecting surfaces may be defined in terms of their parametric equations

$$\left.\begin{aligned} x &= f_1(u,v), \quad y = g_1(u,v), \quad z = h_1(u,v) \\ x &= f_2(u,v), \quad y = g_2(u,v), \quad z = h_2(u,v) \end{aligned}\right\} \tag{44.3}$$

or they may be defined by the implicit relations

$$F(x,y,z) = 0 \quad \text{and} \quad G(x,y,z) = 0 \tag{44.4}$$

found by eliminating the parameters u and v in eqs. (44.3).

45. Equations and direction cosines of the tangent to a space curve. In analytical geometry it is shown that the equation of a line through the point (x_0,y_0,z_0) having directions cosines $\cos\alpha$, $\cos\beta$, and $\cos\gamma$ is

$$\frac{x - x_0}{\cos\alpha} = \frac{y - y_0}{\cos\beta} = \frac{z - z_0}{\cos\gamma} \tag{45.1}$$

where α, β, γ are the angles the line makes with the x, y, and z axes, respectively.

The normal plane to this line passing through the point (x_0,y_0,z_0) has the equation

$$(\cos\alpha)(x - x_0) + (\cos\beta)(y - y_0) + (\cos\gamma)(z - z_0) = 0 \tag{45.2}$$

Thus once the direction cosines of the tangent to a space curve are known, the equation of the tangent at a given point is given by (45.1), and the equation of the plane normal to the curve at this point is given by (45.2).

If the space curve is specified by (44.1) or (44.2), the direction cosines of the tangent may readily be found from the relations

$$\cos\alpha = \frac{dx}{ds}, \qquad \cos\beta = \frac{dy}{ds}, \qquad \cos\gamma = \frac{dz}{ds} \tag{45.3}$$

where s is the length of arc on the curve.

From eqs. (45.3) and (44.1), it is readily seen that

$$\begin{aligned}(ds)^2 &= (dx)^2 + (dy)^2 + (dz)^2 \\ &= [f'(t)^2 + g'(t)^2 + h'(t)^2](dt)^2\end{aligned} \tag{45.4}$$

and the direction cosines of the tangent when the space curve is defined by (44.1) are

$$\begin{aligned}\cos\alpha &= \frac{f'(t)}{\sqrt{f'(t)^2 + g'(t)^2 + h'(t)^2}} \\ \cos\beta &= \frac{g'(t)}{\sqrt{f'(t)^2 + g'(t)^2 + h'(t)^2}} \\ \cos\gamma &= \frac{h'(t)}{\sqrt{f'(t)^2 + g'(t)^2 + h'(t)^2}}\end{aligned} \tag{45.5}$$

When the space curve is defined by the relations (44.2), the direction cosines are

$$\left.\begin{aligned}\cos\alpha &= \frac{dx}{\sqrt{(dx)^2+(dy)^2+(dz)^2}} = \frac{1}{\sqrt{1+(dy/dx)^2+(dz/dx)^2}}\\ \cos\beta &= \frac{dy}{\sqrt{(dx)^2+(dy)^2+(dz)^2}} = \frac{dy/dx}{\sqrt{1+(dy/dx)^2+(dz/dx)^2}}\\ \cos\gamma &= \frac{dz}{\sqrt{(dx)^2+(dy)^2+(dz)^2}} = \frac{dz/dx}{\sqrt{1+(dy/dx)^2+(dz/dx)^2}}\end{aligned}\right\} \quad (45.6)$$

Consequently the equations of the tangent when the space curves are specified by (44.1) and (44.2) are, respectively,

$$\frac{x-x_0}{f'(t_0)} = \frac{y-y_0}{g'(t_0)} = \frac{z-z_0}{h'(t_0)} \quad (45.7)$$

and

$$y-y_0 = \left(\frac{dy}{dx}\right)_{x=x_0}(x-x_0), \quad z-z_0 = \left(\frac{dz}{dx}\right)_{x=x_0}(x-x_0) \quad (45.8)$$

where t_0 corresponds to the point (x_0,y_0,z_0).

Similarly, the corresponding equations for the normal planes for the two specifications are

$$f'(t_0)(x-x_0)+g'(t_0)(y-y_0)+h'(t_0)(z-z_0) = 0 \quad (45.9)$$

$$(x-x_0)+\left(\frac{dy}{dx}\right)_{x=x_0}(y-y_0)+\left(\frac{dz}{dx}\right)_{x=x_0}(z-z_0) = 0 \quad (45.10)$$

In order to find the equation and direction cosines of the tangent when the space curve is defined by the intersection of two surfaces, it is convenient to use the equations of the tangent planes to the intersecting surfaces. In Sec. 15, it was shown that if a surface is defined by the relation $z = f(x,y)$, the total differential of z is

$$dz = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy$$

and the point (x_0+dx,y_0+dy,z_0+dz) lies on the tangent plane to the surface $z = f(x,y)$ at the point (x_0,y_0,z_0). Consequently the equation of the tangent plane may be written

$$z-z_0 = \left(\frac{\partial z}{\partial x}\right)_0(x-x_0)+\left(\frac{\partial z}{\partial y}\right)_0(y-y_0) \quad (45.11)$$

In order to obtain the equation of the tangent plane to the surfaces defined by (44.3) and (44.4) it is necessary only to compute $\partial z/\partial x$ and $\partial z/\partial y$ in each case. The method for doing this has been given in Secs. 16 and 17; the explicit formulas are (16.2) and (17.3), which give for the definition (44.3),

$$\left.\begin{aligned}\frac{\partial z}{\partial x} &= -\frac{\partial(y,z)/\partial(u,v)}{\partial(x,y)/\partial(u,v)} = -\frac{A_0}{C_0}\\ \frac{\partial z}{\partial y} &= -\frac{\partial(z,x)/\partial(u,v)}{\partial(x,y)/\partial(u,v)} = -\frac{B_0}{C_0}\end{aligned}\right\} \quad (45.12)$$

and for the definition (44.4),

$$\frac{\partial z}{\partial x} = -\frac{\partial F/\partial x}{\partial F/\partial z}, \qquad \frac{\partial z}{\partial y} = -\frac{\partial F/\partial y}{\partial F/\partial z} \tag{45.13}$$

These expressions together give for the equations of the tangent planes,

$$A_0(x - x_0) + B_0(y - y_0) + C_0(z - z_0) = 0 \tag{45.14}$$

$$\left(\frac{\partial F}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial F}{\partial y}\right)_0 (y - y_0) + \left(\frac{\partial F}{\partial z}\right)_0 (z - z_0) = 0 \tag{45.15}$$

By comparing eq. (45.2) with eq. (45.14) it is seen that the quantities A_0, B_0, and C_0 must be proportional to the direction cosines of the line normal to the tangent plane at the point (x_0,y_0,z_0). Therefore the direction cosines of the normal must be

$$\cos\alpha = kA_0, \qquad \cos\beta = kB_0, \qquad \cos\gamma = kC_0$$

and from the fact that

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = k^2(A_0^2 + B_0^2 + C_0^2) = 1 \tag{45.16}$$

one may determine k and consequently the direction cosines of the normal to the tangent plane to the surface (44.3) at the point (x_0,y_0,z_0). The direction cosines of the normal to the tangent plane when the surface is specified by (44.4) may be found in a similar way.

The equation of the tangent line at (x_0,y_0,z_0) to the space curve defined by the intersection of surfaces (44.3) can most easily be found by finding the equation of the line in which the tangent planes to the two surfaces at the point (x_0,y_0,z_0) intersect. The equation of the tangent line must satisfy the two equations for the tangent plane at (x_0,y_0,z_0), which from (45.14) would be

$$\left.\begin{aligned} A_{01}(x - x_0) + B_{01}(y - y_0) + C_{01}(z - z_0) &= 0 \\ A_{02}(x - x_0) + B_{02}(y - y_0) + C_{02}(z - z_0) &= 0 \end{aligned}\right\} \tag{45.17}$$

These equations may be put in the equivalent familiar form

$$\frac{x - x_0}{\begin{vmatrix} B_{01} & B_{02} \\ C_{01} & C_{02} \end{vmatrix}} = \frac{y - y_0}{\begin{vmatrix} A_{01} & A_{02} \\ C_{01} & C_{02} \end{vmatrix}} = \frac{z - z_0}{\begin{vmatrix} A_{01} & A_{02} \\ B_{01} & B_{02} \end{vmatrix}} \tag{45.18}$$

When this is compared with eq. (45.1), it is seen that the direction cosines of the tangent line are proportional to the above denominators, or

$$\cos\alpha : \cos\beta : \cos\gamma = \begin{vmatrix} B_{01} & B_{02} \\ C_{01} & C_{02} \end{vmatrix} : \begin{vmatrix} A_{01} & A_{02} \\ C_{01} & C_{02} \end{vmatrix} : \begin{vmatrix} A_{01} & A_{02} \\ B_{01} & B_{02} \end{vmatrix}$$

If one proceeds in the same way when the intersecting surfaces are defined by (44.4), one finds for the equation of the tangent line to the space curve at (x_0,y_0,z_0),

$$\frac{x - x_0}{\begin{vmatrix} \frac{\partial F}{\partial y} & \frac{\partial F}{\partial z} \\ \frac{\partial G}{\partial y} & \frac{\partial G}{\partial z} \end{vmatrix}_0} = \frac{y - y_0}{\begin{vmatrix} \frac{\partial F}{\partial z} & \frac{\partial F}{\partial x} \\ \frac{\partial G}{\partial z} & \frac{\partial G}{\partial x} \end{vmatrix}_0} = \frac{z - z_0}{\begin{vmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \\ \frac{\partial G}{\partial x} & \frac{\partial G}{\partial y} \end{vmatrix}_0} \tag{45.19}$$

and the direction cosines of the tangent line are proportional to the respective denominators.

46. Surfaces. If with every point (u,v) lying in a certain domain D on a plane, we associate a point (x,y,z) in space, then as the point (u,v) traverses the domain, the aggregate of points (x,y,z) constitutes a surface. The association of points (u,v) with (x,y,z) is accomplished by means of the continuous single-valued functions $f(u,v)$, $g(u,v)$, $h(u,v)$, so that

$$x = f(u,v), \qquad y = g(u,v), \qquad z = h(u,v) \tag{46.1}$$

An example of a familiar surface specified by a set of parametric equations of the form (46.1) is

$$x = a \sin\theta \cos\varphi, \qquad y = a \sin\theta \sin\varphi, \qquad z = a \cos\theta$$

As the point (θ,φ) traverses the rectangular domain $(0 \leq \theta \leq \pi, 0 \leq \varphi \leq 2\pi)$, the points (x,y,z) cover the surface of a sphere of radius a with its center at the origin.

The surface defined by (46.1) is said to be regular if the domain D in the uv plane can be divided into a finite number of subdomains within each of which *none* of the Jacobian determinants $\partial(x,y)/\partial(u,v)$, $\partial(y,z)/\partial(u,v)$, $\partial(z,x)/\partial(u,v)$ vanishes. The fulfillment of this condition in any domain means that eqs. (46.1) can be solved for u and v as functions of (x,y,z) or that u and v can be eliminated, yielding a definition of the surface by the implicit function $F(x,y,z) = 0$, or that z can be expressed as an explicit function $z = f(x,y)$.

It is often convenient to make use of the *sides of a curved surface.* In the first place it is necessary to tell when a surface has two distinct sides. If the surface is bounded by a definite contour, the surface has two distinct sides when, starting from a point on one side of the surface and describing any path whatever, one cannot arrive on the opposite side without having crossed the bounding contour.

An example of a surface which does not have two distinct sides can easily be had by taking a strip of paper and twisting one end through 180° and then joining the ends. One can pass from one side to the other without crossing the boundary.

In the case of a surface with distinct sides it is customary to distinguish between the upper side and the lower side. The upper side is defined as that side on which the normal to the tangent plane makes

an acute angle with the positive z axis. If the normal to the tangent planes makes an obtuse angle with the positive z axis, the side is the lower side.

47. Coordinate transformations. The solution of many problems becomes very much simpler when the formulas are expressed in terms of a suitable coordinate system, and the evaluation of many multiple integrals is facilitated by expressing the integrand as a function of convenient coordinates. In this and following sections the concept of a transformation will be discussed, and examples of certain useful coordinate transformations given.

Consider two spaces, one in which a point is specified by the numbers x, y, z, and the other in which a point is specified by the numbers ξ, η, ζ. A transformation sets up a correspondence between points in the xyz space lying in a certain domain and points in a certain domain in the $\xi\eta\zeta$ space. This correspondence may be specified by the transformation equations

$$\xi = f(x,y,z), \qquad \eta = g(x,y,z), \qquad \zeta = h(x,y,z) \tag{47.1}$$

where the functions f, g, and h are required to be single-valued so that the point (ξ,η,ζ) is uniquely determined when a point (x,y,z) is given. It is convenient to call one of the spaces, say the xyz space, the original space, and the $\xi\eta\zeta$ space the image space, and say that the point (ξ,η,ζ) is the image of the point (x,y,z). If the functions f, g, and h are single-valued, there is only one image point for every point in the original space. This does not preclude the possibility that the point (ξ,η,ζ) may not be the image of several points in the xyz space. If, however, the point (ξ,η,ζ) is the image of only one point in the original space, the transformation is said to be reversible. The condition for this is that the functions f, g, and h of (47.1) possess unique inverses which, from Sec. 19, requires that the Jacobian $\partial(\xi,\eta,\zeta)/\partial(x,y,z)$ shall not vanish over the domain under consideration.

48. Curvilinear coordinates. Suppose that we have a mathematical relationship expressed in terms of the coordinates x, y, z, which is not very convenient, and we wish to find the relationship when expressed in terms of ξ, η, ζ, having given the transformation from the xyz space to the $\xi\eta\zeta$ space. To any point P in the xyz space there correspond one or more points P_1 in the $\xi\eta\zeta$ space. The quantities ξ, η, ζ specifying the point P_1 may be taken as coordinates, of a kind, of the point P. Such coordinates are called curvilinear coordinates of P. Functions of (x,y,z) may be expressed in terms of these curvilinear coordinates in two ways; (1) by the solution of eqs. (47.1) for (x,y,z), which in some practical cases is rather difficult, or (2) from a knowledge of the surfaces in the

xyz space which correspond to constant values of the coordinates ξ,η,ζ, respectively. The three sets of surfaces in the xyz space corresponding to the surfaces $\xi = \alpha$, $\eta = \beta$, $\zeta = \gamma$ for different values of the constants α, β, and γ define a transformation just as well as eqs. (47.1). The images of the surfaces $\xi = \alpha$, $\eta = \beta$, $\zeta = \gamma$ for different α, β, γ divide the xyz space into cells which in general are not rectangular parallelepipeds. Examples of this method of setting up transformations will be given in the next two sections.

49. Parabolic coordinates. As an illustration of curvilinear coordinates in two dimensions, let the image of the line $\xi = \alpha$ in the $(\xi\eta)$ plane be the parabola $y^2 = 2\alpha(x + \alpha/2)$ in the (xy) plane. This parabola has its focus at the origin and its axis in the direction of the positive x axis, as shown by the dotted conic in Fig. 49.1. Its latus rectum is 2α. The image of the line $\eta = \beta$ is made to be a parabola confocal with that for which $\xi = \alpha$ but having a latus rectum of 2β and having its axis in the direction of the negative x axis as shown. Its equation is $y^2 = -2\beta(x - \beta/2)$. The values α, β are the curvilinear or parabolic coordinates of the point P. The complete transformation equations are

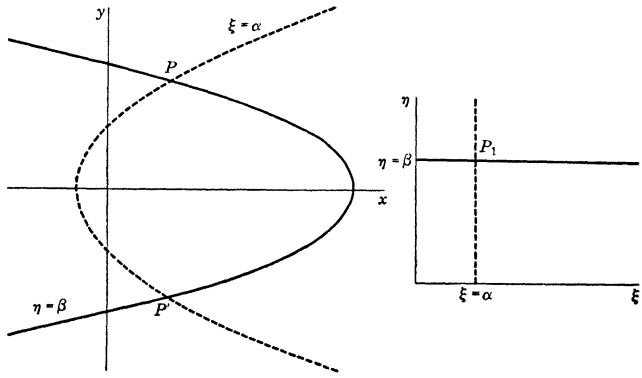


FIG. 49.1

$$y^2 = 2\xi\left(x + \frac{\xi}{2}\right), \qquad y^2 = 2\eta\left(x - \frac{\eta}{2}\right)$$

or, solving for x and y, the transformation equations connecting the $\xi\eta$ space with the xy space are

$$x = \tfrac{1}{2}(\eta - \xi), \qquad y = \sqrt{\xi\eta} \tag{49.1}$$

This transformation is not uniquely reversible because there are two points P and P' in the xy plane which correspond to P_1. The advantage of using confocal parabolas is that they intersect at right angles. The parabolic coordinates may be extended to the three-dimensional case by including an angle of rotation about the axis of symmetry of the parabolas. The transformation equations for the three-dimensional case are usually written

$$x = \sqrt{\xi\eta}\cos\varphi, \qquad y = \sqrt{\xi\eta}\sin\varphi, \qquad z = \tfrac{1}{2}(\xi - \eta) \tag{49.2}$$

50. Elliptic coordinates. In setting up this system of curvilinear coordinates in a plane, it is required that the image of the line $\xi = \alpha =$ constant in the $\xi\eta$ plane be an ellipse in the xy plane such that the foci lie on the x axis and the x and y axes correspond to the major and minor axes. Furthermore, the transformation will be defined such

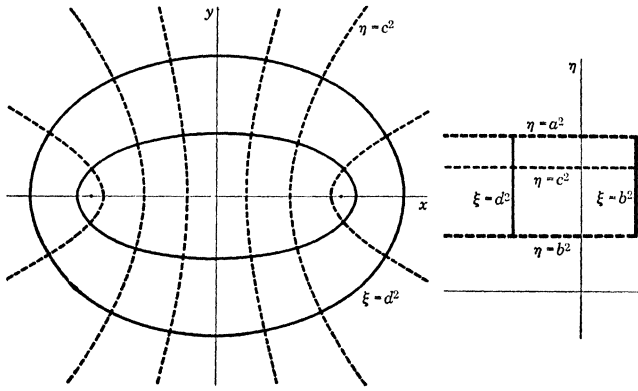


FIG. 50.1

that all lines $\xi =$ constant go over into ellipses which are confocal with that described above. This is accomplished by the transformation

$$\frac{x^2}{a^2 - \alpha} + \frac{y^2}{b^2 - \alpha} = 1 \quad \text{when} \quad a^2 > b^2 > \alpha \tag{50.1}$$

The distance from the origin to one of the foci for the ellipse defined by (50.1) is

$$\sqrt{(a^2 - \alpha) - (b^2 - \alpha)} = \sqrt{a^2 - b^2}$$

which shows that this distance is independent of α and thus the system of ellipses is confocal.

The above equation represents an ellipse as long as $-\infty < \alpha < b^2$

When $b^2 < \alpha < a^2$, eq. (50.1) represents confocal hyperbolas having the same foci as the ellipses. We are not interested in the case for which $a^2 < \alpha < \infty$, since eq. (50.1) then yields an imaginary locus. The complete transformation equations then become

$$\left.\begin{aligned} \frac{x^2}{a^2-\xi}+\frac{y^2}{b^2-\xi}&=1 \qquad -\infty<\xi<b^2 \\ \frac{x^2}{a^2-\eta}+\frac{y^2}{b^2-\eta}&=1 \qquad b^2<\eta<a^2 \end{aligned}\right\} \tag{50.2}$$

Referring to Fig. 50.1, it is to be noted that only the rectangular region in the $\xi\eta$ plane as indicated is imaged in the xy plane. The transformation (50.2) is reversible only for points in one quadrant of the image plane. Again the images of the curves $\xi = \alpha$ and $\eta = \beta$ intersect each other at right angles.

51. The elements of length in the image space. Let the transformation from the xyz space to the $\xi\eta\zeta$ space be defined by the single-valued and differentiable functions

$$x = f(\xi,\eta,\zeta), \qquad y = g(\xi,\eta,\zeta), \qquad z = h(\xi,\eta,\zeta)$$

Then

$$\begin{aligned} dx &= \frac{\partial f}{\partial \xi}\,d\xi + \frac{\partial f}{\partial \eta}\,d\eta + \frac{\partial f}{\partial \zeta}\,d\zeta \\ dy &= \frac{\partial g}{\partial \xi}\,d\xi + \frac{\partial g}{\partial \eta}\,d\eta + \frac{\partial g}{\partial \zeta}\,d\zeta \\ dz &= \frac{\partial h}{\partial \xi}\,d\xi + \frac{\partial h}{\partial \eta}\,d\eta + \frac{\partial h}{\partial \zeta}\,d\zeta \end{aligned}$$

The square of the line element in the original xyz space is defined to be

$$(ds)^2 = (dx)^2 + (dy)^2 + (dz)^2$$

which with the help of the above equations gives

$$(ds)^2 = L_1(d\xi)^2 + L_2(d\eta)^2 + L_3(d\zeta)^2 + 2M_1\,d\xi\,d\eta + 2M_2\,d\eta\,d\zeta + 2M_3\,d\zeta\,d\xi \tag{51.1}$$

where

$$\begin{aligned} L_1 &= \left(\frac{\partial f}{\partial \xi}\right)^2 + \left(\frac{\partial g}{\partial \xi}\right)^2 + \left(\frac{\partial h}{\partial \xi}\right)^2 \\ L_2 &= \left(\frac{\partial f}{\partial \eta}\right)^2 + \left(\frac{\partial g}{\partial \eta}\right)^2 + \left(\frac{\partial h}{\partial \eta}\right)^2 \\ L_3 &= \left(\frac{\partial f}{\partial \zeta}\right)^2 + \left(\frac{\partial g}{\partial \zeta}\right)^2 + \left(\frac{\partial h}{\partial \zeta}\right)^2 \\ M_1 &= \frac{\partial f}{\partial \xi}\frac{\partial f}{\partial \eta} + \frac{\partial g}{\partial \xi}\frac{\partial g}{\partial \eta} + \frac{\partial h}{\partial \xi}\frac{\partial h}{\partial \eta} \\ M_2 &= \frac{\partial f}{\partial \eta}\frac{\partial f}{\partial \zeta} + \frac{\partial g}{\partial \eta}\frac{\partial g}{\partial \zeta} + \frac{\partial h}{\partial \eta}\frac{\partial h}{\partial \zeta} \\ M_3 &= \frac{\partial f}{\partial \zeta}\frac{\partial f}{\partial \xi} + \frac{\partial g}{\partial \zeta}\frac{\partial g}{\partial \xi} + \frac{\partial h}{\partial \zeta}\frac{\partial h}{\partial \xi} \end{aligned}$$

The square root of the right-hand side of eq. (51.1) is called the element of length in the transformed space.

52. Conformal and orthogonal transformations. If two curves in $\xi\eta\zeta$ space intersect in a point P, the images of these curves will intersect in some point P' in the image or xyz space if the transformations are single-valued. Let the angle between the tangents to the two curves at P be α and the corresponding angle in the image space be α' as shown in Fig. 52.1. The transformation from $\xi\eta\zeta$ to xyz is said to be conformal if $\alpha = \alpha'$ for all intersecting curves in $\xi\eta\zeta$ space.

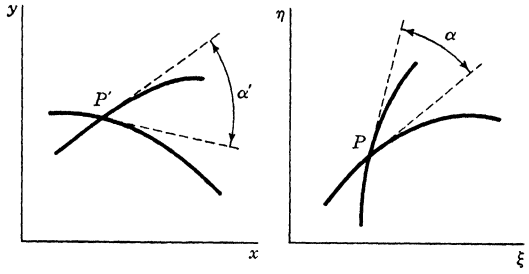


FIG. 52.1

The necessary and sufficient conditions that the magnitude of the angles be the same is that the line element in the original space be proportional to the line element in the image space independent of the direction of this line element. Since the line elements in the two spaces are

$$(ds)^2 = (dx)^2 + (dy)^2 + (dz)^2$$
$$(ds)^2 = L_1(d\xi)^2 + L_2(d\eta)^2 + L_3(d\zeta)^2 + 2M_1\, d\xi\, d\eta + 2M_2\, d\eta\, d\zeta + 2M_3\, d\zeta\, d\xi$$

it is seen that the above conditions require that $M_1 = M_2 = M_3 = 0$ and $L_1 = L_2 = L_3$. Such a transformation is said to be conformal.

If $L_1 \neq L_2 \neq L_3$ but $M_1 = M_2 = M_3 = 0$, all curves which intersect at right angles in the $\xi\eta\zeta$ space give rise to image curves which intersect at right angles in the xyz space. This type of transformation is called an orthogonal transformation. Most of the usual coordinate transformations such as that from Cartesian to polar, or the curvilinear transformations considered in Secs. 49 and 50 are examples of orthogonal transformations. All conformal transformations are necessarily orthogonal transformations.

SUGGESTED REFERENCES

W. F. Osgood, *Advanced Calculus*, New York, The Macmillan Company, 1925.

PROBLEMS

44.1. For many purposes the arc length is the most convenient parameter for the representation of space curves. Express the following curves in terms of the arc length from some convenient point.

(a) $x = a\cos\theta, \quad y = a\sin\theta, \quad z = b\theta.$

(b) $x = t, \quad y = t^2, \quad z = \tfrac{2}{3}t^3.$

(c) $x = t - \sin t, \quad y = 1 - \cos t, \quad z = 0.$

(d) $x = 2\cos\theta - \cos 2\theta, \quad y = 2\sin\theta - \sin 2\theta, \quad z = 0.$

45.1. The cosine of the angle between two lines whose direction cosines are $(\alpha_1\beta_1\gamma_1)$ and $(\alpha_2\beta_2\gamma_2)$ is

$$\cos\varphi = \alpha_1\alpha_2 + \beta_1\beta_2 + \gamma_1\gamma_2$$

Hence, if $\Delta\varphi$ is the angle between the tangents to a curve at the points s and $s + \Delta s$, show that the rate of turning is given by

$$C = \frac{d\varphi}{ds} = \left[\left(\frac{d\alpha}{ds}\right)^2 + \left(\frac{d\beta}{ds}\right)^2 + \left(\frac{d\gamma}{ds}\right)^2\right]^{1/2}$$

$$= \left[\left(\frac{d^2x}{ds^2}\right)^2 + \left(\frac{d^2y}{ds^2}\right)^2 + \left(\frac{d^2z}{ds^2}\right)^2\right]^{1/2}$$

45.2. Show that for a circle of radius r, the result of (45.1) is $C = 1/r$, so that for any curve $1/C$ is the radius of the osculating circle.

46.1. Investigate the regularity of the surface $x = a\sin\theta\cos\varphi$, $y = a\sin\theta\cos\varphi$, $z = z\cos\theta$.

49.1. What are the images in the $\xi\eta$ plane of the lines $x =$ constant, and $y =$ constant, where ξ and η are parabolic coordinates? Sketch.

49.2. An infinitesimal rectangle is bounded by the lines $x = x_0$, $x = x_0 + dx$, and $y = y_0$, $y = y_0 + dy$. What is the image of this rectangle in $\xi\eta$ space where ξ and η are parabolic coordinates? Sketch.

50.1. What are the images in the $\xi\eta$ plane of the lines $x =$ constant, and $y =$ constant, where ξ and η are elliptic coordinates?

50.2. An infinitesimal right triangle has its corners at the points (x_0,y_0), (x_0+dx,y_0), (x_0+dx,y_0+dy). Find the image of this triangle in $\xi\eta$ space, where ξ and η are elliptic coordinates Find the square of the length of the hypotenuse and hence the element of arc length. Sketch.

Chapter 5

LINE, SURFACE, AND MULTIPLE INTEGRALS

53. Functions of bounded variation and rectifiable curves. The concept of a function of bounded variation is of great importance, since many mathematical manipulations with functions are valid only in case the functions are of bounded variation. Suppose the function $f(x)$ is defined in the interval $a \leq x \leq b$ and subdivide the interval in any manner calling the points of subdivision

$$x_0 = a, x_1, x_2, \ldots, x_n = b$$

The particular scheme of subdivision is denoted by σ. If a positive number V exists which is independent of the scheme of subdivision σ for which

$$\sum_{r=1}^{n} \left| f(x_r) - f(x_{r-1}) \right| \leq V$$

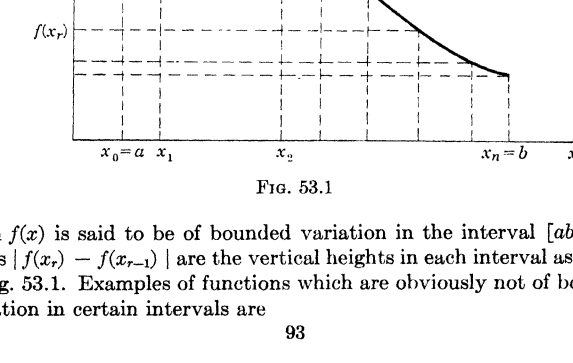


FIG. 53.1

Then $f(x)$ is said to be of bounded variation in the interval $[ab]$. The terms $|f(x_r) - f(x_{r-1})|$ are the vertical heights in each interval as shown in Fig. 53.1. Examples of functions which are obviously not of bounded variation in certain intervals are

$$f(x) = \frac{1}{x} \qquad 0 \leq x \leq 1$$

and $f(x) = \sin(\pi/x)$, $-1 \leq x \leq 1$, as shown in Figs. 53.2 and 53.3.

Three classes of functions which are of bounded variation are the following.

1. Bounded monotonic functions.
2. Sums and products of functions of bounded variation.
3. Any function which can be expressed as the difference of two monotonic functions, either both increasing or both decreasing.

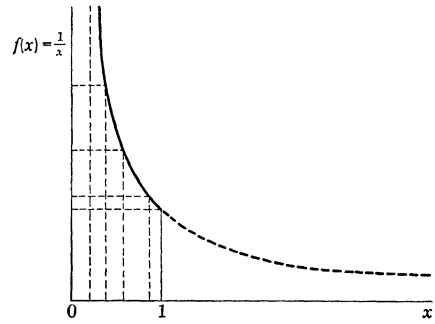


FIG. 53.2

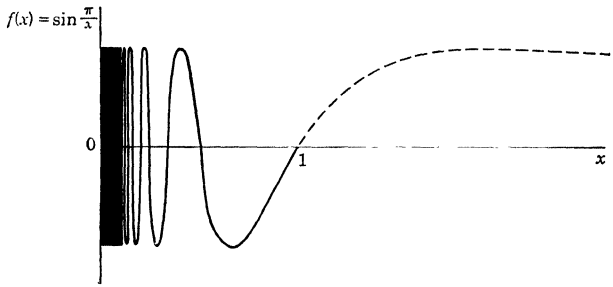


FIG. 53.3

The concept of bounded variation is of importance in connection with the determination of the length of a curve. Let $x = \varphi(t)$ and $y = \psi(t)$ be two bounded single-valued functions for $\alpha \leq t \leq \beta$ so that the simple curve of Fig. 53.4 results. The interval $[\alpha\beta]$ is now subdivided by the points $\alpha = t_0, t_1, \ldots, t_n = \beta$, and the scheme of sub-

division is denoted by σ. Each of these values of t will determine a point

$$(x_0,y_0),\ (x_1,y_1),\ \ldots,\ (x_n,y_n)$$

If the points (x_0,y_0) etc. are connected by straight lines, the points (x_0,y_0) and (x_n,y_n) will be connected by the polygonal line as shown in Fig. 53.4. The length of this line is

$$\sum_{r=1}^{n} \sqrt{(x_r - x_{r-1})^2 + (y_r - y_{r-1})^2}$$

$$= \sum_{r=1}^{n} \sqrt{[\varphi(t_r) - \varphi(t_{r-1})]^2 + [\psi(t_r) - \psi(t_{r-1})]^2}$$

This length will be different for every scheme of subdivision σ, but if the set of values for this length for all σ has a finite least upper bound L, the curve is said to be rectifiable, and has the length L.

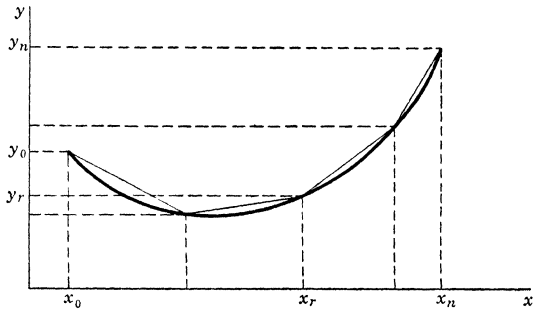


FIG. 53.4

A curve defined by the functions $\varphi(t)$ and $\psi(t)$ is rectifiable when and only when $\varphi(t)$ and $\psi(t)$ are functions of bounded variation.

The actual rectification of a curve is accomplished as follows. If the derivatives $\varphi'(t)$ and $\psi'(t)$ are continuous functions of t in the interval $\alpha \leq t \leq \beta$, the length of the curve $x = \varphi(t)$, $y = \psi(t)$ is given by

$$s = \int_{\alpha}^{\beta} \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\, dt$$

54. Line integrals. Curvilinear or line integrals are extremely important in physics and engineering because it is by means of them that the work done on a body in moving it from one position in space to another along a rectifiable space curve can be calculated. In order to see how line integrals arise, we shall calculate the amount of work done on a body when it is moved along a simple plane curve against a

force whose magnitude and direction is a continuous function of both x and y denoted by $F(x,y)$. The calculations may be carried out in two ways. First we shall proceed as follows. The force will be resolved into its components $X(x,y)$ and $Y(x,y)$ along the x and y axes, respectively. The path AB (Fig. 54.1) over which the body is transported is subdivided by the points $(x_r,y_r)(r = 0,1,\ldots,n)$, and any point on the arc between (x_r,y_r) and (x_{r-1},y_{r-1}) is selected and denoted by (ξ_r,η_r), so that $x_{r-1} \leq \xi_r \leq x_r$ and $y_{r-1} \leq \eta_r \leq y_r$. The approximate amount of work

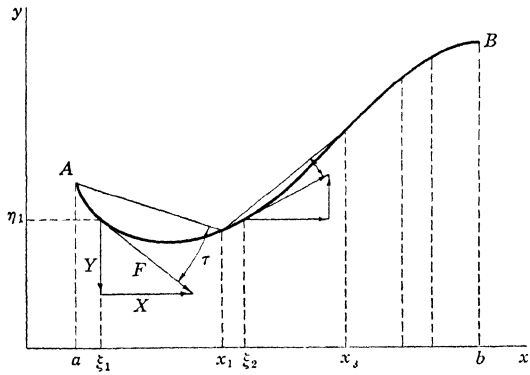


FIG. 54.1

done in moving the body from A to B against the x component of force will be

$$X(\xi_1,\eta_1)(x_1 - a) + X(\xi_2,\eta_2)(x_2 - x_1) + \ldots + X(\xi_n,\eta_n)(b - x_{n-1}) \quad (54.1)$$

Likewise the approximate work done against the y component of force is

$$Y(\xi_1,\eta_1)(y_1 - a) + Y(\xi_2,\eta_2)(y_2 - y_1) + \ldots + Y(\xi_n,\eta_n)(b - y_{n-1}) \quad (54.2)$$

The sum of the expressions (54.1) and (54.2) gives an approximate result for the total amount of work done. An increasingly accurate result is obtained as the number of subdivisions is increased. If the sums (54.1) and (54.2) possess definite limits as the number of subdivisions is indefinitely increased in such a way that the largest of them approaches zero, these limits define the line or curvilinear integrals of $X(x,y)$ or $Y(x,y)$ over the curve AB. These are then written

$$\int_{AB} X(x,y)dx \quad \text{and} \quad \int_{AB} Y(x,y)dy$$

The total work done is then

$$W = \int_{AB} X(x,y)dx + \int_{AB} Y(x,y)dy$$

Conditions for the Existence of Line Integrals. If the curve AB is defined by the equations $x = \varphi(t)$ and $y = \psi(t)$ in the interval $\alpha \leq t \leq \beta$, and if $\varphi(t)$ and $\psi(t)$ are monotonic functions in this interval so that to every pair of values of x and y there exists only one value of t, then y is a single-valued function of x, namely, $y = f(x)$. If $P(x,y)$ is a continuous function of both x and y, then $P[x,f(x)] = P_1(x)$ is a continuous function of x, and the line integral becomes the ordinary definite integral

$$\int_a^b P_1(x)dx$$

The above conditions prevent a line parallel to the x or y axis from intersecting the curve AB in more than one point.

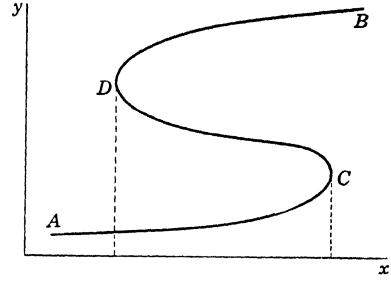


FIG. 54.2

If the curve AB is of the type shown in Fig. 54.2 where $\varphi(t)$ and $\psi(t)$ are not monotonic throughout the interval $\alpha \leq t \leq \beta$, the line integral can still be found by breaking up the curve into parts over which $\varphi(t)$ and $\psi(t)$ are monotonic. For the curve shown, the parts AC, CD, and DB may be defined by monotonic functions $\varphi(t)$ and $\psi(t)$, and the line integral becomes

$$\int_{AB} P(x,y)dx = \int_{AC} P(x,y)dx + \int_{CD} P(x,y)dx + \int_{DB} P(x,y)dx$$

In each of the integrands of the three integrals on the right, y must be replaced by a different single-valued function of x.

Entirely similar considerations hold for the line integral $\int_{AB} Q(x,y)dy$.

A few important properties of line integrals of this type are made evident in the following relations.

(1) $\int_{AB} P(x,y)dx = 0$, when AB is a line parallel to the y axis.

(2) $\int_{AB} Q(x,y)dy = 0$, when AB is a line parallel to the x axis.

(3) $\int_{AC} P(x,y)dx = \int_{AB} P(x,y)dx + \int_{BC} P(x,y)dx.$

(4) $\int_{AB} P(x,y)dx = -\int_{BA} P(x,y)dx.$

(5) If the derivatives of the functions $x = \varphi(t)$ and $y = \psi(t)$ defining the curve AB are continuous in an interval $\alpha \leq t \leq \beta$, then

$$\int_{AB} P(x,y)dx = \int_{\alpha}^{\beta} P[\varphi(t),\psi(t)]\varphi'(t)dt$$

$$\int_{AB} Q(x,y)dy = \int_{\alpha}^{\beta} Q[\varphi(t),\psi(t)]\psi'(t)dt$$

Alternative Definition of the Work Done. Referring again to Fig. 54.1, a different expression for the approximate value of the work done in moving a body along the curve AB against a force would be

$$\sum_{r=1}^{n} F(\xi_r,\eta_r) \cos \tau_r \sqrt{(x_r - x_{r-1})^2 + (y_r - y_{r-1})^2}$$

A progressively better approximation to the total work done is obtained as the number of subdivisions is increased. Hence if the above sum has a limit as the number of subdivisions is increased without limit in such a way that the largest subdivision or norm approaches zero, the limit is defined as the line integral

$$\int_{AB} F(x,y)\lambda(x,y)ds$$

where $\lambda(x,y)$ is the cosine of the angle between the direction of $F(x,y)$ at any point on AB and the tangent to the curve at that point. In terms of the parametric equations $x = \varphi(t)$ and $y = \psi(t)$ of the curve AB, the line integral becomes

$$\int_{\alpha}^{\beta} F[\varphi(t),\psi(t)]\lambda[\varphi(t),\psi(t)]\sqrt{\varphi'(t)^2 + \psi'(t)^2}\, dt$$

55. Line integrals of total differentials. If $F(x,y)$, together with $\partial F/\partial x$ and $\partial F/\partial y$, are single-valued and continuous over a certain domain S_2, then

$$\int_{AB} dF = \int_{AB} \left[\frac{\partial F}{\partial x}\, dx + \frac{\partial F}{\partial y}\, dy\right] = F(x_B,y_B) - F(x_A,y_A)$$
$$= F[\varphi(\beta),\psi(\beta)] - F[\varphi(\alpha),\psi(\alpha)]$$

where the curve AB is defined by the relations $x = \varphi(t)$, $y = \psi(t)$ for t in the interval $\alpha \leq t \leq \beta$. The above relation shows that the line integral $\int_{AB} dF$ depends only on the end points of the curve and not upon the path or curve connecting the end points A and B. In the special case where A and B may be regarded as the beginning and end of a closed curve, so that A and B are one and the same point, it is clear that $\int_{AB} dF = 0$. Line integrals over closed curves are often indicated by $\oint$.

From the above considerations, it is seen that if we have two functions $P(x,y)$ and $Q(x,y)$ which are continuous and single-valued, together with the partial derivatives $\partial P/\partial y$ and $\partial Q/\partial x$ in a domain S_2, then in order that the integral $\int_{AB} P(x,y)dx + Q(x,y)dy$ be independent

of the path of integration between two points A and P, there must be a function F of which the integrand is the total differential, i.e.,

$$dF = P(x,y)dx + Q(x,y)dy$$

where

$$P(x,y) = \frac{\partial F}{\partial x} \quad \text{and} \quad Q(x,y) = \frac{\partial F}{\partial y}$$

and

$$\frac{\partial P}{\partial y} = \frac{\partial^2 F}{\partial y\,\partial x} = \frac{\partial^2 F}{\partial x\,\partial y} = \frac{\partial Q}{\partial x}$$

The condition $\partial P/\partial y = \partial Q/\partial x$ is a necessary and sufficient condition that the original line integral be independent of the path between A and B.

The converse of the above theorem states that if

$$\oint P(x,y)dx + Q(x,y)dy = 0 \quad \text{then} \quad \frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$$

56. Definition of a double integral. Since the general concept of multiple integrals follows without difficulty from that for double integrals, it will be sufficient to define the latter class of integrals. The fundamental interval in two dimensions is a rectangle, and since this

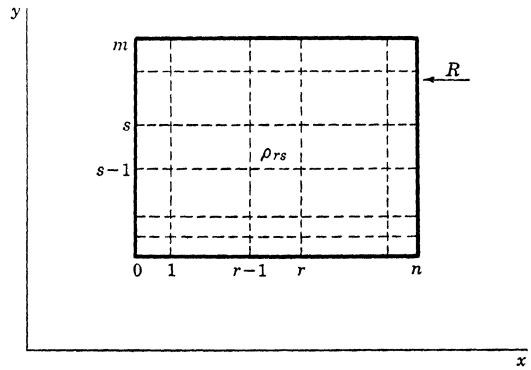


FIG. 56.1

is the case, we shall first define the double integral over the rectangle R in Fig. 56.1. The extension to integrals over other than rectangular domains will be given later.

Let $f(x,y)$ be a bounded function over the rectangle R so that $m \leq f(x,y) \leq M$. The rectangle R is subdivided by lines parallel to the x and y axes into subrectangles of area ρ_{rs}. The greatest lower and

least upper bounds of the function in the subrectangle rs may be designated m_{rs} and M_{rs}, respectively. The upper and lower approximate sums are defined similar to those for a single integral, and may be written

$$S_\sigma = \sum_{r=1}^{n}\sum_{s=1}^{m} M_{rs}\rho_{rs} \qquad s_\sigma = \sum_{r=1}^{n}\sum_{s=1}^{m} m_{rs}\rho_{rs}$$

where σ denotes the particular scheme of subdivision. Let the greatest lower bound of the set S_σ be S and the greatest upper bound of the set s_σ be s for all σ. If $S = s = I$, then I is called the Riemann double integral of $f(x,y)$ over the domain R, and is written

$$I = \int\int_R f(x,y)dx\,dy$$

It should be carefully noted that this is not a repeated integral (see next section).

Functions possessing a double integral over a finite domain R are:

(1) Every continuous function of x and y.

(2) Bounded functions whose discontinuities fall on a finite number of rectifiable curves.

57. Properties and evaluations of double integrals. Some important relations satisfied by double integrals are listed below. First

$$m\,\Delta_R \leq \int\int_R f(x,y)dx\,dy \leq M\,\Delta_R$$

where m and M are the greatest lower and least upper bounds of $f(x,y)$ over the rectangle R, and Δ_R is the area of the rectangle. Secondly, if $f(x,y)$ is a continuous function over the rectangle R, then

$$\int\int_R f(x,y)dx\,dy = f(\xi,\eta)\Delta_R$$

where ξ and η are the coordinates of some point in the rectangle R. Finally

$$\left|\int\int_R f(x,y)dx\,dy\right| \leq \int\int_R |f(x,y)|\,dx\,dy$$

The actual evaluation of a double integral for practical purposes is based on the following theorem. If $\int\int_R f(x,y)dx\,dy$ exists, where R is the rectangle $a \leq x \leq b$, $\alpha \leq y \leq \beta$, then the two repeated integrals

$$\int_\alpha^\beta dy\int_a^b f(x,y)dx \quad \text{and} \quad \int_a^b dx\int_\alpha^\beta f(x,y)dy$$

also exist and each is equal to the value of the double integral.

It should be noted that the above theorem not only provides a method for computing the double integral but furnishes a sufficient condition for the inversion of the order of integration in a repeated integral.

58. Double integrals over nonrectangular domains. The double integral over a domain D bounded by a closed curve of the type shown

in Fig. 58.1 may be computed in the following way. It will be supposed that $f(x,y)$ is continuous over the domain D and that D is bounded by a rectifiable curve. One method of computing

$$\iint_D f(x,y)dx\,dy$$

is to enclose the domain D with a rectangle for which $a' \leq x \leq b'$ and $\alpha' \leq y \leq \beta'$, and then define a new function $f_D(x,y)$ which is equal to

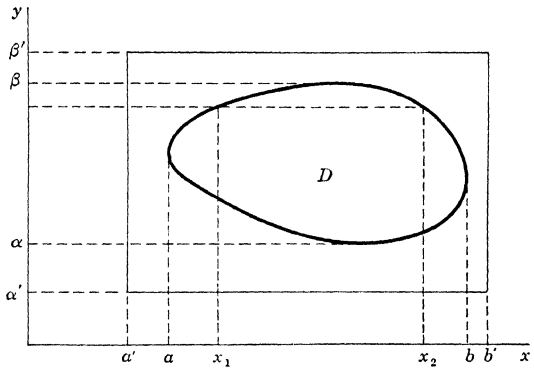


FIG. 58.1

$f(x,y)$ at all points inside D and is equal to zero everywhere else. The integral

$$\iint_D f_D(x,y)dx\,dy$$

exists as stated in Sec. 57, since the discontinuities of $f_D(x,y)$ lie on a rectifiable curve and

$$\iint_D f(x,y)dx\,dy = \iint_R f_D(x,y)dx\,dy$$

Thus the integral over the domain D can be found by defining an auxiliary integrand in the manner given above. Then

$$\iint_D f(x,y)dx\,dy = \int_{\alpha'}^{\beta'} dy \int_{a'}^{b'} f_D(x,y)dx$$

Another method of computing the double integral over D is to draw, parallel to the x axis at $y = \alpha$ and $y = \beta$, lines which are tangent to the curve bounding D as shown in Fig. 58.1. Then any other line parallel to the x axis for which $\alpha < y < \beta$ will intersect the bounding curve in not more than two points. Let the x coordinates of these two points be x_1 and x_2 as shown. For each value of y there will be two values of x and this association may be expressed by the functions $x_1 = x_1(y)$ and $x_2 = x_2(y)$. Then

$$\int_{\alpha'}^{\beta'} dy \int_{\alpha'}^{b'} f_D(x,y)dx = \int_{\alpha'}^{\beta'} dy \left[\int_{\alpha'}^{x_1} + \int_{x_1}^{x_2} + \int_{x_2}^{b'} f_D(x,y)dx\right]$$
$$= \int_{\alpha'}^{\beta'} dy \int_{x_1}^{x_2} f_D(x,y)dx = \left(\int_{\alpha'}^{\alpha} + \int_{\alpha}^{\beta} + \int_{\beta}^{\beta'} dy\right) \int_{x_1}^{x_2} f_D(x,y)dx$$
$$= \int_{\alpha}^{\beta} dy \int_{x_1}^{x_2} f(x,y)dx$$

In this case the value of the double integral has been found by choosing the limits of integration and thus avoiding the necessity of defining an auxiliary integrand.

By an exactly similar process we could obtain

$$\int_{a'}^{b'} dx \int_{\alpha'}^{\beta'} f_D(x,y)dy = \int_{a}^{b} dx \int_{y_1}^{y_2} f(x,y)dy$$

so that finally

$$\iint_D f(x,y)dx\,dy = \int_{\alpha}^{\beta} dy \int_{x_1}^{x_2} f(x,y)dx$$
$$= \int_{a}^{b} dx \int_{y_1}^{y_2} f(x,y)dy$$

If the domain D is of the type shown in Fig. 58.2, so that some lines parallel to the x axis intersect the bounding curve in more than two points, the domain D must be divided into partial domains such that in each of these a line parallel to the x axis intersects the bounding curves in only two points. The dotted line in the figure effects the division into the proper partial domains A, B, and C.

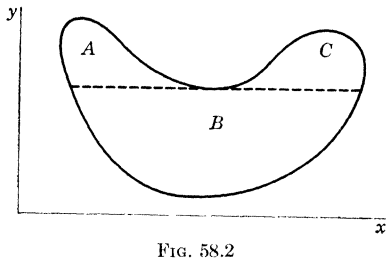


FIG. 58.2

59. Connections between double and line integrals: Green's theorem. An extremely valuable relation between a double integral and a line integral may readily be established. In the first place, consider a domain D shown in Fig. 59.1 and let it be cut by two lines parallel to the y axis at x_1 and x_2, and call the parts of the boundary cut off between these lines C_1 and C_2 as shown. These curve segments may be specified by

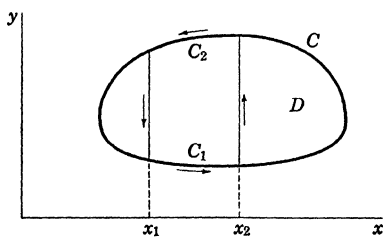


FIG. 59.1

the functions $y = \varphi_1(x)$ and $y = \varphi_2(x)$. The line integral of a function $F(x,y)$ which is continuous over the domain D taken over the boundary of the dissected domain is

$$\int_C F(x,y)dx = \int_{C_1} F(x,y)dx + \int_{C_2} F(x,y)dx$$

since the integral over the line parallel to y vanishes. The boundary is traversed in the counterclockwise direction, so that

$$\begin{aligned}\int_C F(x,y)dx &= \int_{x_1}^{x_2} F[x,\varphi_1(x)]dx + \int_{r_2}^{x_1} F[x,\varphi_2(x)]dx \\ &= \int_{x_1}^{r_2} [F_1(x) - F_2(x)]dx \end{aligned} \tag{59.1}$$

In a similar way we could have dissected the domain with lines parallel to the x axis and obtained

$$\int_{C''} G(x,y)dy = \int_{y_1}^{y_2} [G_2(y) - G_1(y)]dy \tag{59.2}$$

From these results the following important theorem may be established: *Given a domain D bounded by a regular curve C so that lines parallel to the x and y axes do not intersect C in more than two points and also given two functions $P(x,y)$ and $Q(x,y)$ which together with the partial derivatives $\partial P/\partial y$ and $\partial Q/\partial x$ are continuous throughout D. Then*

$$\int_C P(x,y)dx + Q(x,y)dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dx\,dy \tag{59.3}$$

when the bounding curve C is traversed in the positive direction. To determine the positive direction in any given case, draw a normal n to the

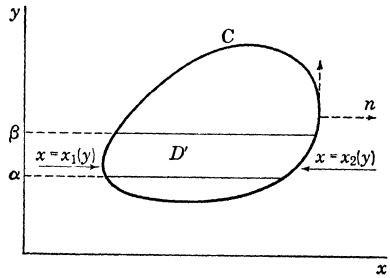


FIG. 59.2

curve C pointing away from the domain (outer normal) but having the direction of the positive x axis as shown in Fig. 59.2. Then a perpendicular to n in the positive y direction will indicate the positive direction in traversing the curve.

The above theorem may be proved by showing that

$$\int_C P\,dx = -\iint_D \frac{\partial P}{\partial y}\,dx\,dy \quad \text{and} \quad \int_C Q\,dy = \iint_D \frac{\partial Q}{\partial x}\,dx\,dy \tag{59.4}$$

Since line integrals are additive over their domains we may consider a domain between two lines parallel to the x axis, see Fig. 59.2. Then over this dissected domain D',

$$\begin{aligned}\iint_{D'} \frac{\partial Q}{\partial x}\,dx\,dy &= \int_\alpha^\beta dy \int_{x_1(y)}^{x_2(y)} \frac{\partial Q}{\partial x}\,dx \\ &= \int_\alpha^\beta [Q(y,x_2(y)) - Q(y,x_1(y))]dy \\ &= \int_\alpha^\beta (Q_2 - Q_1)dy = \int_{C''} Q(x,y)dy\end{aligned}$$

where eq. (59.2) was used in the last step. In a similar way the first relation in (59.4) can be proved.

It sometimes happens that the domain to be considered is bounded by two regular curves, one entirely enclosed in the other, as shown in Fig. 59.3. In this case the domain may be broken up into four domains by the dotted lines parallel to the y axis, in each of which Green's theorem holds. Applying Green's theorem to each domain and adding is equivalent to evaluating the line integral over the outer curve in the counterclockwise direction and the inner curve in the opposite direction. This direction of traversing the boundary will still be termed positive, since if a normal to the inner curve is drawn away from the domain and in the positive x direction, the inner curve is traversed in the direction of positive y, i.e., the arrow in Fig. 59.3.

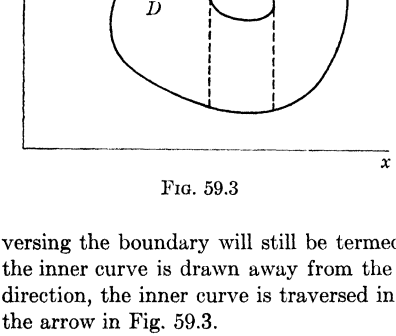


FIG. 59.3

An immediate application of Green's theorem may be found by letting $Q = x$ and $P = -y$, in which case (59.3) yields

$$\begin{aligned}\int_C x\,dy - y\,dx &= \iint_D (1+1)dx\,dy \\ &= 2\iint_D dx\,dy = 2A\end{aligned}$$

where A is the area of the domain D, so that

$$A = \frac{1}{2}\int_C x\,dy - y\,dx \tag{59.5}$$

Another way of computing the area would be to place $Q = 0$ and $P = -y$ so that

$$\int_C -y\,dx = \iint_D 1\,dx\,dy = A \tag{59.6}$$

and hence

$$A = -\int_C y\,dx$$

On the other hand, if $Q = x$ and $P = 0$, then

$$A = \int_C x\,dy \tag{59.7}$$

where C is always the bounding curve of the area under consideration and the line integral evaluated in the positive direction. This allows one to express the area enclosed by a curve in terms of a line integral over the closed curve.

As a simple example of the expression (59.5) we may compute the area of the circle $x^2 + y^2 = a^2$. From (59.5),

$$A = \frac{1}{2}\int_C x\,dy - y\,dx \quad \text{but} \quad x = a\cos t \quad \text{and} \quad y = a\sin t,$$

so that

$$A = \frac{1}{2}\int_0^{2\pi} (a^2\cos^2 t + a^2\sin^2 t)dt = \frac{a^2}{2}\int_0^{2\pi} dt = \pi a^2$$

60. Triple integrals. The definition of a triple integral is essentially the same as that for a double integral, the difference being that the fundamental domain V is a parallelepiped. We need only mention a few important facts concerning them. If $\int\int\int f(x,y,z)dx\,dy\,dz$ exists over the domain V which is a rectangular parallelepiped $a \leq x \leq b$, $\alpha \leq y \leq \beta$, $A \leq z \leq B$, then

(1) If $\int_a^b f(x,y,z)dx$ exists, the integral

$$\int_A^B \int_\alpha^\beta dz\,dy \int_a^b f(x,y,z)dx$$

also exists, and is equal to the triple integral; or

(2) If $\int_\alpha^\beta \int_a^b f(x,y,z)dx\,dy$ exists,

$$\int_A^B dz \int_\alpha^\beta \int_a^b f(x,y,z)dx\,dy$$

also exists and is equal to the triple integral.

The extension of the above facts to nonrectangular domains may be accomplished either by an auxiliary integrand or by a suitable choice of the limits of integration as in the case of double integrals.

61. The element of area on a transformed plane. When points in a certain domain in one coordinate plane are associated with a domain of points in another by means of transformation equations (see Sec. 47) it is important to establish the connection between corresponding areas in the two planes. A domain D_1 in the $\xi\eta$ plane transforms into a domain D in the image plane, and if C_1 is a closed curve bounding a certain area within D_1, there will be a closed curve C in D bounding the corresponding area in the image plane as shown in Fig. 61.1. It must be carefully noted that if the curve C_1 is traversed in the positive sense (see Sec. 59), the image point may traverse C in either the positive or negative sense, depending on the particular transformation. If the transformation is such that both curves are traversed in the same sense, it is said to be direct; otherwise it is termed an inverse transformation.

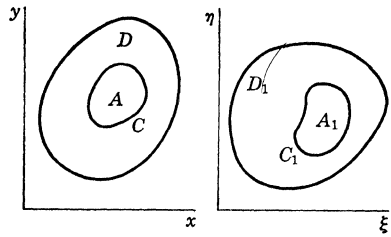


FIG. 61.1

As examples of a direct and inverse transformation, it is easily seen that if a rectangle in the $\xi\eta$ plane of Fig. 61.2a is traversed in

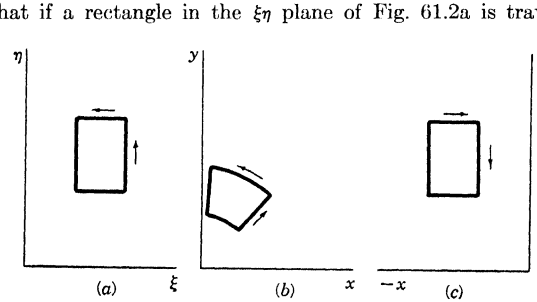


FIG. 61.2

the positive direction, the image curve Fig. 61.2b for the transformation $x = \xi \cos \eta$, $y = \xi \sin \eta$ is also traversed in the positive direction, making the transformation direct. On the other hand, the simple transformation $x = -\xi$, $y = \eta$ is inverse, as shown in Fig. 61.2c. The general criterion for directness of the transformation will be established below.

The area of the region A bounded by the curve C in Fig. 61.1 is by definition $\iint_A dx\, dy$. According to (59.7),

$$\iint_A dx\,dy = \int_C x\,dy$$

where C is traversed in the positive sense. When the integrand of the integral on the right is expressed in terms of ξ and η, one obtains

$$\int_C x\,dy = \int_{C_1} \left[x(\xi,\eta)\frac{\partial y}{\partial \xi}\right]d\xi + \left[x(\xi,\eta)\frac{\partial y}{\partial \eta}\right]d\eta$$

From Green's theorem (59.3), the line integral on the right is equal to

$$\pm\iint_{A_1}\left[\frac{\partial}{\partial \xi}\left(x\frac{\partial y}{\partial \eta}\right) - \frac{\partial}{\partial \eta}\left(x\frac{\partial y}{\partial \xi}\right)\right]d\xi\,d\eta$$

where the $+$ or $-$ sign must be used according as the curve C_1 is traversed in the positive or negative direction as C is traversed positively. The above integral can be simplified and we have the result

$$\iint_A dx\,dy = \pm\iint_{A_1}\left[\frac{\partial x}{\partial \xi}\frac{\partial y}{\partial \eta} - \frac{\partial x}{\partial \eta}\frac{\partial y}{\partial \xi}\right]d\xi\,d\eta$$

$$= \pm\iint_{A_1}\frac{\partial(x,y)}{\partial(\xi,\eta)}\,d\xi\,d\eta$$

It was pointed out in Sec. 19 that the transformation equations $x = f(\xi,\eta)$, $y = g(\xi,\eta)$ would not possess unique inverses if the Jacobian $\partial(x,y)/\partial(\xi,\eta)$ vanished in the domain considered. Consequently, the Jacobian must be either always positive or always negative within A_1. On the other hand, the area $\iint_A dx\,dy$ is essentially positive, so that the Jacobian must have the sign which makes the term on the right positive. The result then is that the Jacobian is positive for a direct transformation, and negative for an inverse transformation. An application of this test to the transformation $x = \xi\cos\eta$, $y = \xi\sin\eta$ considered above gives ξ for the value of the Jacobian, so that for $\xi > 0$ the transformation is direct. This is the useful case, since actually ξ plays the roll of the radius in plane polar coordinates. For the transformation $x = -\xi$, $y = \eta$ also considered above, the value of the Jacobian is -1, and the transformation is inverse for all regions in the $\xi\eta$ plane.

No matter whether the transformation is direct or not we have established the important result

$$\iint_A dx\,dy = \iint_{A'}\left|\frac{\partial(x,y)}{\partial(\xi,\eta)}\right|d\xi\,d\eta \tag{61.1}$$

where the absolute value of the Jacobian is used. Since the integrands are positive and the domain A is arbitrary, we have the direct result

$$dx\,dy = \left|\frac{\partial(x,y)}{\partial(\xi,\eta)}\right|d\xi\,d\eta \tag{61.2}$$

where the right-hand side is the element of area in the $\xi\eta$ plane corresponding to the element $dx\,dy$ in the xy plane.

62. Evaluation of definite integrals by double integration. It often happens that a single integral may be replaced by a double integral, after which a transformation of coordinates allows the integral to be evaluated in a simple manner. Two illustrations of this method will be given.

First, let it be required to evaluate

$$I = \int_0^\infty e^{-a^2x^2}\,dx$$

The integration variable can be changed and I equally well expressed as

$$I = \int_0^\infty e^{-a^2y^2}\,dy,$$

or I^2 may be computed instead of I, so that

$$I^2 = \int_0^\infty e^{-a^2x^2}\,dx \cdot \int_0^\infty e^{-a^2y^2}\,dy = \iint_D e^{-a^2(x^2+y^2)}\,dx\,dy$$

where the domain D of the double integral is all of the first quadrant in the xy plane. Transforming to polar coordinates, $x = r\cos\varphi$ and $y = r\sin\varphi$, the element of area $dx\,dy$ according to (61.2) is

$$dx\,dy = \left|\frac{\partial(x,y)}{\partial(r,\varphi)}\right| dr\,d\varphi = \begin{vmatrix} \cos\varphi & \sin\varphi \\ -r\sin\varphi & r\cos\varphi \end{vmatrix} dr\,d\varphi = r\,dr\,d\varphi$$

so that

$$\begin{aligned} I^2 &= \iint_D e^{-a^2r^2} r\,dr\,d\varphi \\ &= \int_0^\infty e^{-a^2r^2} r\,dr \int_0^{\pi/2} d\varphi = \frac{\pi}{2}\int_0^\infty e^{-a^2r^2} r\,dr \\ &= \frac{\pi}{4a^2} \end{aligned}$$

Hence

$$I = \frac{\sqrt{\pi}}{2a}$$

As a second example, let us attempt to obtain an expression for the product of two gamma functions. By definition,

$$\Gamma(q) = \int_0^\infty x^{q-1}e^{-x}\,dx \quad \text{and} \quad \Gamma(p) = \int_0^\infty y^{p-1}e^{-y}\,dy$$

where p and q are positive. Multiplying these expressions together, we have

$$\begin{aligned} \Gamma(p)\cdot\Gamma(q) &= \int_0^\infty x^{q-1}e^{-x}\,dx \cdot \int_0^\infty y^{p-1}e^{-y}\,dy \\ &= \iint_D x^{q-1}y^{p-1}e^{-(x+y)}\,dx\,dy \end{aligned}$$

where the domain D is the first quadrant in the xy plane. An initially convenient change of variables is $x = s^2$ and $y = t^2$ in which case

$$dx\,dy = \frac{\partial(x,y)}{\partial(s,t)}\,ds\,dt = 4st\,ds\,dt$$

and

$$\Gamma(p)\cdot\Gamma(q) = 4\iint_D s^{2q-1}t^{2p-1}e^{-(s^2+t^2)}\,ds\,dt$$

In this form it is again convenient to transform to the polar coordinates $s = r\cos\varphi$ and $t = r\sin\varphi$, where as in the previous example

$$ds\,dt = r\,dr\,d\varphi$$

and
$$\begin{aligned}\Gamma(p)\cdot\Gamma(q) &= 4\iint_D r^{2q+2p-1}\cos^{2q-1}\varphi\sin^{2p-1}\varphi\,e^{-r^2}\,dr\,d\varphi \\ &= 4\int_0^\infty r^{2p+2q-1}e^{-r^2}\,dr\int_0^{\pi/2}\cos^{2q-1}\varphi\sin^{2p-1}\varphi\,d\varphi\end{aligned}$$

In the integral over r we may substitute $r^2 = z$, where $2r\,dr = dz$ and

$$\begin{aligned}\int_0^\infty r^{2p+2q-1}e^{-r^2}\,dr &= \frac{1}{2}\int_0^\infty z^{p+q-1}e^{-z}\,dz \\ &= \frac{1}{2}\Gamma(p+q)\end{aligned}$$

Therefore

$$\Gamma(p)\cdot\Gamma(q) = \Gamma(p+q)\int_0^{\pi/2} 2\cos^{2q-1}\varphi\sin^{2p-1}\varphi\,d\varphi$$

The integral on the right is known as the beta function and designated $B(p,q)$. An important formula expressing the beta function in terms of gamma functions has thus been established, namely,

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \tag{62.1}$$

A more common integral expression for the beta function results if the transformation $x = \cos^2\varphi$ is made. Then

$$dx = -2\cos\varphi\sin\varphi\,d\varphi, \quad \text{and}$$

$$\begin{aligned}B(p,q) &= \int_0^{\pi/2} 2\cos^{2q-1}\varphi\sin^{2p-1}\varphi\,d\varphi = \int_0^1 \frac{x^q(1-x)^p\,dx}{x(1-x)} \\ &= \int_0^1 x^{q-1}(1-x)^{p-1}\,dx\end{aligned} \tag{62.2}$$

In passing it is worth while to note the following important characteristics of the beta function.

(1) $B(p,q) = B(q,p)$

(2) $B(p,q+1) = \dfrac{\Gamma(p)\Gamma(q+1)}{\Gamma(p+q+1)}$

$$= \frac{q\Gamma(p)\Gamma(q)}{(p+q)\Gamma(p+q)} = \frac{q}{p+q}B(p,q)$$

$$(3)\quad B(p,q) = \int_0^1 x^{q-1}(1-x)^{p-1}\,dx$$

$$= -\frac{x^q}{q}(1-x)^{p-1}\Big]_0^1 + \frac{p-1}{q}\int_0^1 x^q(1-x)^{p-2}\,dx$$

Replacing p by $p+1$ throughout, we have

$$B(p+1,q) = \frac{p}{q}B(p,q+1)$$

63. Elements of length and area on a surface. The parametric equations

$$x = f(u,v), \qquad y = g(u,v), \qquad z = h(u,v)$$

which define a surface in the xyz space may be regarded as establishing a correspondence between points in a given domain in the uv plane and a domain on the surface, so that we have to do with a transformation and may speak of the surface as being the image of a domain in the uv plane. The square of the element of length on the surface is

$$(ds)^2 = (dx)^2 + (dy)^2 + (dz)^2$$

where dx, dy, dz are of course not independent differentials, but expressed in terms of the variables u and v are

$$dx = \frac{\partial f}{\partial u}\,du + \frac{\partial f}{\partial v}\,dv$$

$$dy = \frac{\partial g}{\partial u}\,du + \frac{\partial g}{\partial v}\,dv$$

$$dz = \frac{\partial h}{\partial u}\,du + \frac{\partial h}{\partial v}\,dv$$

The square of the line element on the surface thus becomes

$$(ds)^2 = L_1(du)^2 + 2M_1\,du\,dv + L_2(dv)^2 \tag{63.1}$$

where L_1, L_2, and M_1 have the same meaning as in Sec. 51.

The element of surface area may be arrived at in the following manner. Let the surface S in Fig. 63.1 be the image of the domain D in the uv plane, and let ΔA_1 be any plane element of area enclosing the point P_1 in D be imaged by the domain ΔA enclosing the image point P on S. Under suitable conditions the ratio $\Delta A/\Delta A_1$ approaches a unique limit as the element ΔA_1 is diminished to zero, always enclosing P_1. This limit multiplied by $du\,dv$ is defined as the element of area on the surface. When the above limit exists its value is $\sqrt{L_1L_2 - M_1^2}$. The proof of the existence of this limit does not fall within the scope of this book and will be omitted. It can be shown without difficulty that

$$\sqrt{L_1L_2 - M_1^2} = \sqrt{A^2 + B^2 + C^2} \tag{63.2}$$

where A, B, and C are the Jacobians defined by (45.12). Consequently the element of area on the surface is taken to be

$$dS = \sqrt{L_1 L_2 - M_1^2}\, du\, dv$$
$$= \sqrt{A^2 + B^2 + C^2}\, du\, dv \tag{63.3}$$

The first of these definitions depends on the line element on the surface being $ds = \sqrt{(dx)^2 + (dy)^2 + (dz)^2}$ and hence is independent of the choice of the orthogonal axes x, y, z as one would naturally expect. The element of area dS is thus invariant with respect to a change from one set of Cartesian axes to another.

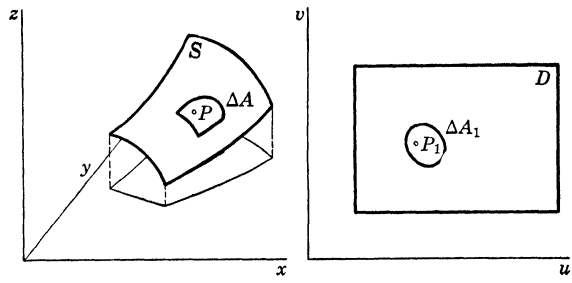


FIG. 63.1

64. Surface areas. Let the surface S be the image of a domain D in the uv plane, the transformation equations connecting S to D being

$$x = f(u,v), \qquad y = g(u,v), \qquad z = h(u,v)$$

If these functions, together with their first partial derivatives, are single-valued and continuous throughout the domain D, and if the three Jacobians A, B, C defined by (45.12) do not vanish simultaneously in D, the area of the surface is $\iint_D dS$, where dS is the element of surface area defined by (63.3).

The integral $\iint_D dS$ giving the surface area can be modified in certain cases so as to eliminate the coordinates u and v. From eq. (45.16) it follows that

$$\frac{\cos\alpha}{A} = \frac{\cos\beta}{B} = \frac{\cos\gamma}{C} = \frac{1}{\sqrt{A^2 + B^2 + C^2}} \tag{64.1}$$

where $\cos\alpha$, $\cos\beta$, $\cos\gamma$ are the direction cosines of the normal to the tangent plane to the surface. For definiteness it will be supposed that $C \neq 0$ in D for the surface under consideration. This means that the tangent plane is never perpendicular to the xy plane. Furthermore it

will be assumed that the normal to the tangent plane always makes an acute angle with the z axis, so that $\cos\gamma > 0$. Then

$$\cos\gamma = \frac{C}{\sqrt{A^2 + B^2 + C^2}} \tag{64.2}$$

The surface area can then be written

$$\iint_D dS = \iint_D \sqrt{A^2 + B^2 + C^2}\, du\, dv = \iint_D \frac{|C|}{\cos\gamma}\, du\, dv \tag{64.3}$$

Instead of integrating over the region D in the uv plane one can transform the integral so as to integrate over a region $\sum$ in the xy plane. In order to do this we would need to know u and v as functions of x and y, and to express the plane element of area in the xy plane corresponding to $du\, dv$. From (61.2) this would be

$$du\, dv = \left|\frac{\partial(u,v)}{\partial(x,y)}\right| dx\, dy = \frac{1}{|C|}\, dx\, dy$$

Consequently the surface area may be expressed

$$\iint_D dS = \iint_D \sqrt{A^2 + B^2 + C^2}\, du\, dv = \iint_{\Sigma_{xy}} \frac{dx\, dy}{\cos\gamma} \tag{64.4}$$

The domain $\sum_{xy}$ is just the projection of the surface S on the xy plane. When similar assumptions are made about A and B, a similar argument gives

$$\iint_D dS = \iint_{\Sigma_{yz}} \frac{dy\, dz}{\cos\alpha} = \iint_{\Sigma_{xz}} \frac{dx\, dz}{\cos\beta} \tag{64.5}$$

where $\sum_{yz}$ and $\sum_{xz}$ are the projections of the surface on the yz and xz planes, respectively.

When $C \neq 0$ in D, it is always possible to define the surface by the explicit relation $z = \varphi(x,y)$. From relations in Sec. 45, $\cos\gamma$ in this case is

$$\cos\gamma = \frac{1}{\sqrt{1 + (\partial z/\partial x)^2 + (\partial z/\partial y)^2}}$$

so that the area of the surface when defined by $z = \varphi(x,y)$ is

$$\iint_\Sigma \sqrt{1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}\, dx\, dy \tag{64.6}$$

The use of these formulas for determining the area of a surface will be illustrated by calculating the area of the hemispherical surface defined by the parametric equations

$$x = a\sin\theta\cos\varphi, \qquad y = a\sin\theta\sin\varphi, \qquad z = a\cos\theta$$

where θ and φ correspond to the uv coordinates used above. The domain D is the rectangle $0 \leq \theta \leq \pi/2$ and $0 \leq \varphi \leq 2\pi$. The values of

the Jacobians A, B, C are easily found to be

$$A = a^2 \sin^2 \theta \cos \varphi, \quad B = a^2 \sin^2 \theta \sin \varphi, \quad C = a^2 \sin \theta \cos \theta$$

The element of area

$$dS = \sqrt{A^2 + B^2 + C^2} = a^2 \sin \theta \, d\theta \, d\varphi$$

Hence the area of the hemisphere is

$$S = \int_0^{2\pi} \int_0^{\pi/2} a^2 \sin \theta \, d\theta \, d\varphi = 2\pi a^2$$

The area may be computed by formula (64.4) or (64.6), which gives rise to the same integral, but the integrals are not as convenient to evaluate as the one above. In order to use (64.4), which involves an integration over the domain $\sum$ which is the projection of the surface on the xy plane, which in this case is simply the circle of radius a with center at the origin, we need $\cos \gamma$. It is clear that

$$\cos \gamma = \cos \theta = \frac{z}{a} = \frac{\sqrt{a^2 - x^2 - y^2}}{a}$$

Thus

$$S = \int_{-a}^{+a} \int_{-\sqrt{a^2-y^2}}^{\sqrt{a^2-y^2}} \frac{a \, dx \, dy}{\sqrt{a^2 - x^2 - y^2}} = \pi a \int_{-a}^{+a} dy = 2\pi a^2$$

That this integral is exactly the same as the one found by substituting the values of $\partial z/\partial x$ and $\partial z/\partial y$ in (64.6) may easily be verified.

65. Surface integrals. The integrals considered in the last section are special cases of a more general class of integrals known as surface integrals, just as the integral for the length of a curve is a particular case of a line integral. Surface integrals are of two types: one in which the integration is with respect to an element of surface area dS; the other in which the integration is with respect to an element of area $dx \, dy$ in the xy plane or an element $dy \, dz$ in the yz plane, etc. The two types correspond to the two types (64.3) and (64.1) of the last section.

A surface integral of the first type is written

$$\iint_S F(x,y,z) dS \tag{65.1}$$

where $F(x,y,z)$ is assumed to be a single-valued continuous function for all values of x, y, z on the surface S, where S is defined by the usual parametric equations

$$x = f(u,v), \qquad y = g(u,v), \qquad z = h(u,v)$$

The value of (65.1) is defined to be

$$\iint_D G(u,v) \sqrt{L_1 L_2 - M_1^2} \, du \, dv \tag{65.2}$$

or
$$\iint_D G(u,v)\sqrt{A^2 + B^2 + C^2}\, du\, dv \tag{65.3}$$

where
$$G(u,v) = F[f(u,v),g(u,v),h(u,v)]$$

and D is the domain in the uv plane of which the surface S is the image. The remainder of the integrands are the elements of surface area defined in Sec. 63.

In defining a surface integral of the second type, only integrals involving the element of area on the xy plane will be considered, since those with elements on the other planes are entirely similar. It will be supposed that the surface S is regular (see Sec. 46) so that it can be divided into a finite number of subsurfaces S_r, where each subsurface is defined by the single-valued function $z = \varphi(x,y)$. The projections of the surfaces S_r on the xy plane will be denoted by $\sum_r$.

The surface integral of the second type over the subsurface S_r is written

$$\iint_{\Sigma_r} F(x,y,z)dx\, dy = \pm \iint_{\Sigma_r} F[x,y,\varphi(x,y)]dx\, dy$$

where the positive sign is taken for the upper side of S_r and the negative sign for the lower side (see Sec. 46). If any of the sections into which S has been divided are portions of a cylindrical surface with generators parallel to the z axis, they may be ignored, since they make no contribution to the surface integral.

The integral on the right may be transformed into an integral over the domain D_r in the uv plane [see eq. (61.2)], in which case

$$\begin{aligned}\iint_{\Sigma_r} F[x,y,\varphi(x,y)]dx\, dy &= \iint_{D_r} G(u,v)\left|\frac{\partial(x,y)}{\partial(u,v)}\right| du\, dv \\ &= \iint_{D_r} G(u,v)|\, C\,|du\, dv\end{aligned}$$

Adding the values of the integrals for all the sections $\sum_r$, one obtains finally

$$\iint_{\Sigma_{xy}} F(x,y,z)dx\, dy = \iint_D G(u,v)C\, du\, dv \tag{65.4}$$

where the side of S to be taken is determined by the sign of C. Similar results hold for integrals over domains in the yz and xz planes. In this type of surface integral it is to be understood that the domain of integration $\sum_{xy}$ is the projection of S on the appropriate plane, in this case the xy plane.

Just as in the case of the two types of integrals in Sec. 64, one type may be expressed in terms of the other. Consider the surface integral

$$\iint_S \cos\gamma\, F(x,y,z)dS,$$

where $\cos\gamma$ is the direction cosine of the normal to the tangent plane

at the point (x,y,z). According to (65.3), this integral is defined to be

$$\iint_D G(u,v)(\cos\gamma)\sqrt{A^2+B^2+C^2}\,du\,dv = \iint_D G(u,v)C\,du\,dv$$

in which the relation (64.2) has been used. The right side is the same as the right side of (65.4) so that

$$\iint_S (\cos\gamma)F(x,y,z)dS = \iint_{\Sigma_{xy}} F(x,y,z)dx\,dy \tag{65.5}$$

As an example of a surface integral, let it be required to find the total electric charge on the surface of the hemisphere considered in the example of Sec. 64, where the surface charge density is $\rho = \sqrt{1 - z^2/a^2}$. The total charge Q is given by the surface integral

$$Q = \iint_S \sqrt{1 - \frac{z^2}{a^2}}\,dS$$

Using the definition (65.3) and the value of $\sqrt{A^2+B^2+C^2}$ already found for this surface in Sec. 64, and noting that

$$G(u,v) = \sqrt{1 - \cos^2\theta} = \sin\theta$$

$$Q = \int_0^{2\pi}\int_0^{\pi/2} \sin\theta\cdot a^2\sin\theta\,d\theta\,d\varphi = \frac{\pi^2 a^2}{2}$$

$$= \frac{\pi}{4} \times \text{area of surface}$$

In terms of the second type of surface integrals the total charge Q is

$$Q = \int_{-a}^{+a}\int_{-\sqrt{a^2-y^2}}^{\sqrt{a^2-y^2}} \sqrt{1 - \frac{z^2}{a^2}\frac{a}{z}}\,dx\,dy$$

66. Transformation of coordinates in a surface integral. When a surface S in the xyz coordinate space is defined by the parametric equations $x = f(u,v)$, $y = g(u,v)$, and $z = h(u,v)$ and a transformation from the xyz space to a new space (ξ,η,ζ) is carried out by means of the transformation equations

$$x = x(\xi,\eta,\zeta), \qquad y = y(\xi,\eta,\zeta), \qquad z = z(\xi,\eta,\zeta) \tag{66.1}$$

the surface S goes over into a surface S_1 in the $\xi\eta\zeta$ space which we may call the image of S. In order to express the surface integral

$$\iint_{\Sigma_{xy}} F(x,y,z)dx\,dy$$

in terms of ξ, η, ζ and the surface S_1, it is to be noted in the first place according to (65.4) that the surface S_1 can also be defined by the parametric equations

$$\xi = f_1(u,v), \qquad \eta = g_1(u,v), \qquad \zeta = h_1(u,v) \tag{66.2}$$

Although these equations will not actually be required, they can be obtained by the elimination of x, y, z from the transformation eqs. (66.1) and the parametric equations of the surface S. From eqs. (66.1) and (66.2), the function $F(x,y,z)$ can be expressed in terms of ξ, η, ζ or u, v, so that

$$F(x,y,z) = F_1(\xi,\eta,\zeta) = G(u,v)$$

From (65.4),

$$\iint_{\Sigma_{xy}} F(x,y,z)\,dx\,dy = \iint_D G(u,v)\frac{\partial(x,y)}{\partial(u,v)}du\,dv$$

where D is the domain in the uv plane whose image is the surface S. If in the integral on the right x and y are regarded as functions of ξ,η,ζ, then according to (19.3),

$$\frac{\partial(x,y)}{\partial(u,v)} = \frac{\partial(x,y)}{\partial(\xi,\eta)}\frac{\partial(\xi,\eta)}{\partial(u,v)} + \frac{\partial(x,y)}{\partial(\eta,\zeta)}\frac{\partial(\eta,\zeta)}{\partial(u,v)} + \frac{\partial(x,y)}{\partial(\zeta,\xi)}\frac{\partial(\zeta,\xi)}{\partial(u,v)}$$

When this is substituted in the integral on the right there result integrals of the form

$$\iint_D G(u,v)\frac{\partial(x,y)}{\partial(\xi,\eta)}\frac{\partial(\xi,\eta)}{\partial(u,v)}\,du\,dv = \iint_{\Sigma_{\xi\eta}} F_1(\xi,\eta,\zeta)\frac{\partial(x,y)}{\partial(\xi,\eta)}d\xi\,d\eta$$

where $\dfrac{\partial(\xi,\eta)}{\partial(u,v)}du\,dv$ is just the element of area corresponding to the element $d\xi\,d\eta$ in the $\xi\eta$ plane [see (61.2)]. The domain $\sum_{\xi\eta}$ of integration of the integral on the right is the projection of S_1 on the $\xi\eta$ plane. The final result is

$$\iint_{\Sigma_{xy}} F(x,y,z)dx\,dy = \int\int_{\Sigma_{\xi\eta},\Sigma_{\eta\zeta},\Sigma_{\xi\zeta}} F_1(\xi,\eta,\zeta)\left[\frac{\partial(x,y)}{\partial(\xi,\eta)}d\xi\,d\eta + \frac{\partial(x,y)}{\partial(\eta,\zeta)}d\eta\,d\zeta + \frac{\partial(x,y)}{\partial(\zeta,\xi)}d\xi\,d\zeta\right] \qquad (66.3)$$

The cyclical order of ξ, η, ζ should be maintained, or the sign will be changed. It should be noted that whereas the domain of integration in the xyz space was the projection of S on the xy plane, the domain of integration in the $\xi\eta\zeta$ space in general includes the projections of S_1 on all three coordinate planes.

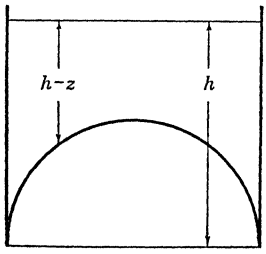


FIG. 66.1

As an example of the transformation considered above, let the spherical shell defined by the parametric equations

$$x = a\sin\theta\cos\varphi,\quad y = a\sin\theta\sin\varphi,$$
$$z = a\cos\theta,\quad 0 \le \theta \le \pi/2, 0 \le \varphi \le 2\pi$$

form the bottom of a cylindrical vessel which is filled to a height h with fluid of density ρ, as indicated in Fig. 66.1, and let it be required

to find the total force on the shell. This force is given by a surface integral of the type just considered, namely,

$$F = \rho \iint_{\Sigma} (h - z)dx\, dy$$

the integral being carried out over the projection of the shell on the xy plane. The integral may be evaluated directly. By definition

$$F = \iint_{\Sigma_1} \rho(h - z)dx\, dy = \iint_D G(u,v) \left| \frac{\partial(x,y)}{\partial(\theta,\varphi)} \right| d\theta\, d\varphi$$

$$= \rho \int_0^{2\pi} \int_0^{\pi/2} (h - a\cos\theta)a^2 \sin\theta\cos\theta\, d\theta\, d\varphi = \rho\pi a^2(h - \tfrac{2}{3}a)$$

A transformation of coordinates makes the integral a little simpler. Let the transformation be defined by

$$x = \xi\cos\eta, \qquad y = \xi\sin\eta, \qquad z = \zeta$$

The Jacobians required in the transformed integral [see (66.3)] are

$$\frac{\partial(x,y)}{\partial(\xi,\eta)} = \xi, \qquad \frac{\partial(x,u)}{\partial(\eta,\zeta)} = 0, \qquad \frac{\partial(x,y)}{\partial(\zeta,\xi)} = 0$$

Also $h - z(x,y) = h - \sqrt{a^2 - \xi^2}$, so that

$$F = \rho \int_0^{2\pi} \int_0^{a} (h - \sqrt{a^2 - \xi^2})\xi\, d\xi\, d\eta = \pi a^2 \rho(h - \tfrac{2}{3}a)$$

67. Green's theorem in three dimensions. This theorem establishes a relation between an integral over a three-dimensional domain or volume and an integral over the bounding closed surface S. If $P(x,y,z)$, $Q(x,y,z)$, and $R(x,y,z)$ are continuous, together with $\partial P/\partial x$, $\partial Q/\partial y$, and $\partial R/\partial z$, throughout the domain V, this relation is

$$\int\int_{\Sigma_{yz},\Sigma_{xz},\Sigma_{xy}} (P\, dy\, dz + Q\, dz\, dx + R\, dx\, dy)$$

$$= \iiint_V \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) dx\, dy\, dz \qquad (67.1)$$

This relation can be proved in parts. First consider the part

$$\iint_{\Sigma_{xy}} R\, dx\, dy = \iiint_V \frac{\partial R}{\partial z}\, dx\, dy\, dz$$

and divide the domain V by a plane so that the normal to the upper bounding surface makes an acute angle with the positive z axis, and the normal to the lower surface makes an obtuse angle with the positive z axis. Let the lower surface S_1 and the upper surface S_2 be specified by the explicit relations $z = z_1(x,y)$ and $z = z_2(x,y)$, respectively.

$$\iiint_V \frac{\partial R}{\partial z}\,dx\,dy\,dz = \iint dx\,dy \int_{z_1}^{z_2} \frac{\partial R}{\partial z}\,dz$$

$$= \iint_{\Sigma_{xy}} [R(x,y,z_2) - R(x,y,z_1)]dx\,dy$$

where Σ_{xy} is the domain obtained by projecting the surface S onto the xy plane. The separate integrals on the right may be written as one integral $\int\int R(x,y,z)dx\,dy$, where it is to be remembered that the integral changes sign in passing from the upper to the lower side of the surface (see Secs. 46 and 65). Equality between the other parts of (67.1) can easily be shown in the same manner, giving the complete result (67.1). Green's theorem may be conveniently written in terms of the other type of surface integral [see (65.5)], in which case

$$\iiint_V \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\,dx\,dy\,dz$$

$$= \iint_S (P\cos\alpha + Q\cos\beta + R\cos\gamma)dS \qquad (67.2)$$

As a simple example of the use of Green's theorem, let $P = x$, $Q = 0$, $R = 0$, whence (67.1) gives

$$\iiint_V dx\,dy\,dz = \iint_{\Sigma_{yz}} x\,dy\,dz$$

The left integral just gives the volume of the domain. Take the volume bounded by the right circular cone whose axis coincides with the z axis, whose height is h, and of which the radius of the base in the xy plane is a. The parametric equations of this surface are $x = r\cos\theta$, $y = r\sin\theta$, $z = h\left(1 - \frac{r}{a}\right)$. The integral on the right is then

$$\iint_{\Sigma_{yz}} x\,dy\,dz = \iint_D r\cos\theta \frac{\partial(y,z)}{\partial(r,\theta)}dr\,d\theta$$

$$= \int_0^a \int_0^{2\pi} \frac{h}{a} r^2 \cos^2\theta\,dr\,d\theta = \frac{h\pi a^2}{3}$$

In this case the contributions from the lower and upper side of the surface add; that is,

$$\text{upper side} = \int_{-\pi/a}^{\pi/2} \frac{ha^2}{3}\cos^2\theta\,d\theta, \quad \text{lower side} = \int_{\pi/2}^{-\pi/2} \frac{ha^2}{3}\cos^2\theta\,d\theta$$

and

$$\int_{-\pi/2}^{+\pi/2} \ldots d\theta - \int_{\pi/2}^{-\pi/2} \ldots d\theta = \int_0^{2\pi} \ldots d\theta$$

68. The volume element and its transformation. When points in a domain V of the three-dimensional xyz space are associated with points in another domain V_1 in the $\xi\eta\zeta$ space by means of the transformation equations

$$x = x(\xi,\eta,\zeta), \qquad y = y(\xi,\eta,\zeta), \qquad z = z(\xi,\eta,\zeta)$$

it is important to know how the corresponding volume elements in the two spaces are related. This relation is determined in a manner analogous to that used in Sec. 61 for finding out how the element of area transforms.

The right-hand side of eq. (67.1) gives three times the volume of the domain V when we let $P = x$, $Q = y$, $R = z$, so that

$$\iiint_V dx\,dy\,dz = \frac{1}{3}\int\int_{\Sigma_{yz},\Sigma_{zx},\Sigma_{xy}} (x\,dy\,dz + y\,dz\,dx + z\,dx\,dy)$$

The method of transforming the integral on the right to the $\xi\eta\zeta$ space has already been given in Sec. 66. Applying the transformation formula of the type (66.3), we have

$$\begin{aligned}
&\int\int_{\Sigma_{yz},\Sigma_{zx},\Sigma_{xy}} (x\,dy\,dz + y\,dz\,dx + z\,dx\,dy) \\
&\quad = \int\int_{\Sigma_{\xi\eta},\Sigma_{\eta\zeta},\Sigma_{\zeta\xi}} \left[x\frac{\partial(y,z)}{\partial(\xi,\eta)} + y\frac{\partial(z,x)}{\partial(\xi,\eta)} + z\frac{\partial(x,y)}{\partial(\xi,\eta)}\right] d\xi\,d\eta \\
&\qquad + \left[x\frac{\partial(y,z)}{\partial(\eta,\zeta)} + y\frac{\partial(z,x)}{\partial(\eta,\zeta)} + z\frac{\partial(x,y)}{\partial(\eta,\zeta)}\right] d\eta\,d\zeta \\
&\qquad + \left[x\frac{\partial(y,z)}{\partial(\zeta,\xi)} + y\frac{\partial(z,x)}{\partial(\zeta,\xi)} + z\frac{(x,y)}{(\zeta,\xi)}\right] d\zeta\,d\xi \\
&\quad = \int\int_{\Sigma_{\eta\zeta},\Sigma_{\zeta\xi},\Sigma_{\xi\eta}} [P_1(\xi,\eta,\zeta)d\eta\,d\zeta + Q_1(\xi,\eta,\zeta)d\zeta\,d\xi + R_1(\xi,\eta,\zeta)d\xi\,d\eta] \\
&\quad = \int\int\int_{V_1} \left(\frac{\partial P_1}{\partial \xi} + \frac{\partial Q_1}{\partial \eta} + \frac{\partial R_1}{\partial \zeta}\right) d\xi\,d\eta\,d\zeta
\end{aligned}$$

Green's theorem (67.1) has again been used in making the last step. When the indicated differentiations are carried out in the last integral, the integrand reduces to

$$3\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)}$$

Thus

$$\iiint_V dx\,dy\,dz = \iiint_{V_1} \left|\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)}\right| d\xi\,d\eta\,d\zeta \tag{68.1}$$

The absolute value of the Jacobian is used in order to avoid confusion in the sign, the volume always being positive. Since the integrands are positive, and V and therefore V_1 may be chosen arbitrarily small, it can be concluded that the volume element in the $\xi\eta\zeta$ space corresponding to $dx\,dy\,dz$ in the xyz space is

$$dx\,dy\,dz = \left|\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)}\right| d\xi\,d\eta\,d\zeta \tag{68.2}$$

Consequently the relations between a general volume integral in the xyz space and that over the corresponding volume in the transformed space is

$$\iiint_V F(x,y,z)dx\,dy\,dz = \iiint_{V_1} F_1(\xi,\eta,\zeta)\left|\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)}\right| d\xi\,d\eta\,d\zeta \qquad (68.3)$$

An example showing the convenience of making use of a space transformation or change of variable in a triple or volume integral is encountered in the evaluation of the Dirichlet integral

$$I = \iiint x^\alpha y^\beta z^\gamma (1 - x - y - z)^\delta dx\,dy\,dz$$

over the interior of the tetrahedron formed by the coordinate planes and the plane $x + y + z = 1$. One step in the simplification of the integrand is effected by making the substitution

$$x + y + z = \xi, \qquad y + z = \xi\eta, \qquad z = \xi\eta\zeta$$

By solving for x, y, and z, one obtains the transformation equations

$$x = \xi(1 - \eta), \qquad y = \xi\eta(1 - \zeta), \qquad z = \xi\eta\zeta$$

The aim in making this transformation was to make the tetrahedron in the xyz space correspond to a cube in the $\xi\eta\zeta$ space. It may readily be verified that the tetrahedron in the xyz space is the unique image of a unit cube with edges coinciding with the positive ξ, η, ζ axes.

According to (68.2), we require the value of the Jacobian $\partial(x,y,z)/\partial(\xi,\eta,\zeta)$. The evaluation of it can be facilitated by an artifice which is convenient to use in many cases. If an intermediate change of variable is made, then according to eq. (19.1) the Jacobian may be written

$$\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)} = \frac{\partial(x,y,z)}{\partial(u,v,w)}\,\frac{\partial(u,v,w)}{\partial(\xi,\eta,\zeta)}$$

It is convenient then to let $u = \xi$, $v = \xi\eta$, $w = \xi\eta\zeta$, whence $x = u - v$, $y = v - w$, $z = w$, and

$$\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)} = \begin{vmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{vmatrix} \cdot \begin{vmatrix} 1 & \eta & \eta\zeta \\ 0 & \xi & \xi\zeta \\ 0 & 0 & \xi\eta \end{vmatrix} = \xi^2\eta$$

The integral I in the $\xi\eta\zeta$ space is then

$$I = \int_0^1 \xi^{\alpha+\beta+\gamma+2}(1 - \xi)^\delta\,d\xi \int_0^1 \eta^{\beta+\gamma+1}(1 - \eta)^\alpha\,d\eta \int_0^1 \zeta^\gamma(1 - \zeta)^\beta\,d\zeta$$

According to (62.2) the above integrals are beta functions, so that

$$I = B(\delta+1,\alpha+\beta+\gamma+3)\cdot B(\alpha+1,\beta+\gamma+2)\cdot B(\beta+1,\gamma+1)$$

From relations given in Sec. 62 these beta functions can be expressed in terms of gamma functions.

69. Transformation of the Laplacian operator to orthogonal curvilinear coordinates. A method of expressing

$$\Delta\psi = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2}$$

in other coordinate systems has already been given in Sec. 13. Although the method given is fundamental, the transformations are usually tedious. The expression for $\Delta\psi$ in terms of another orthogonal coordinate system can be found more readily with the help of Green's theorem.

Let a domain of volume V be bounded by the surface S in the xyz space which is the image of the corresponding domain V_1 and bounding surface S_1 in the $\xi\eta\zeta$ space, where the spaces are associated by the transformation equations $x = x(\xi,\eta,\zeta)$, $y = y(\xi,\eta,\zeta)$, and $z = z(\xi,\eta,\zeta)$. If in the expression of Green's theorem (67.2), P, Q, and R are replaced by $\partial\psi/\partial x$, $\partial\psi/\partial y$, and $\partial\psi/\partial z$, respectively, one obtains

$$\iiint_V \Delta\psi \, dx\, dy\, dz = \iint_S \left(\frac{\partial\psi}{\partial x}\cos\alpha + \frac{\partial\psi}{\partial y}\cos\beta + \frac{\partial\psi}{\partial z}\cos\gamma\right) dS$$

where $\cos\alpha$, $\cos\beta$, $\cos\gamma$ are the direction cosines of the outer normal of the tangent plane to the surface S. The integrand on the left is the Laplacian, while the one on the right is the directional derivative of ψ with respect to the outer normal n of the surface (see Sec. 14). Hence the above equation may be written

$$\iiint_V \Delta\psi \, dx\, dy\, dz = \iint_S \frac{\partial\psi}{\partial n} \, dS$$

When one transforms the above volume integral to the $\xi\eta\zeta$ space according to (68.3) one obtains

$$\iiint_V \Delta\psi_{x,y,z} \, dx\, dy\, dz = \iiint_{V_1} \Delta\psi_{\xi,\eta,\zeta} \left|\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)}\right| d\xi\, d\eta\, d\zeta = \iint_S \frac{\partial\psi}{\partial n} \, dS \tag{69.1}$$

The next step is to compute the Jacobian determinant and then evaluate the surface integral. The Jacobian can be conveniently evaluated by using the square of the line element in the xyz space, which for an orthogonal transformation (see Sec. 52) is

$$(ds)^2 = L_1(d\xi)^2 + L_2(d\eta)^2 + L_3(d\zeta)^2$$

It is convenient to replace L_1, L_2, and L_3 by $1/h_1^2$, $1/h_2^2$, and $1/h_3^2$, whereupon

$$(ds)^2 = \frac{(d\xi)^2}{h_1^2} + \frac{(d\eta)^2}{h_2^2} + \frac{(d\zeta)^2}{h_3^2} \tag{69.2}$$

It may be shown that

$$\left|\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)}\right| = \sqrt{L_1L_2L_3} = \frac{1}{h_1h_2h_3}$$

so that the transformed volume integral becomes

$$\iiint_{V_1} \frac{(\Delta\psi)_{\xi,\eta,\zeta}}{h_1h_2h_3}\,d\xi\,d\eta\,d\zeta = \iint_S \frac{\partial\psi}{\partial n}\,dS \tag{69.3}$$

The surface integral on the right may be evaluated with most facility by taking the volume V_1, of which V is the image, to be a rectangular parallelepiped bounded by the planes

$$\begin{aligned} \xi &= \xi_1 & \eta &= \eta_1 & \zeta &= \zeta_1 \\ \xi &= \xi_2 & \eta &= \eta_2 & \zeta &= \zeta_2 \\ \xi_1 &< \xi_2 & \eta_1 &< \eta_2 & \zeta_1 &< \zeta_2 \end{aligned}$$

The surface S is then made up of the images of these plane surfaces in the $\xi\eta\zeta$ space. Furthermore, when ξ is placed equal to ξ_1, the transformation equations immediately constitute the definition, in terms of the parameters $\eta\zeta$, of that part of S, say S_{ξ_1} which is the image of the face D_{ξ_1} of the parallelepiped coincident with the plane $\xi = \xi_1$. From (65.2)

$$\iint_{S_{\xi_2}} \frac{\partial\psi}{\partial n}\,dS = \iint_{D_{\xi_2}} \frac{\partial\psi}{\partial n}\frac{d\eta\,d\zeta}{h_2h_3} = \iint_{D_{\xi_2}} h_1\frac{\partial\psi}{\partial\xi}\frac{d\eta\,d\zeta}{h_2h_3}$$

where from the line element (69.2) it is evident that $\partial/\partial n = h_1(\partial/\partial\xi)$. Also

$$\iint_{S_{\xi_1}} \frac{\partial\psi}{\partial n}\,dS = -\iint_{D_{\xi_1}} \frac{\partial\psi}{\partial n}\frac{d\eta\,d\zeta}{h_2h_3} = -\iint_{D_{\xi_1}} \left(-h_1\frac{\partial\psi}{\partial\xi}\right)\frac{d\eta\,d\zeta}{h_2h_3}$$

The negative sign before the integral arises because the integral is being taken over the lower side of the surface. The other negative sign comes because the normal to the surface D_{ξ_1} is in the direction of decreasing ξ. Similar results would have been obtained for the other plane surfaces bounding V_1, so that

$$\begin{aligned} &\iint_S \frac{\partial\psi}{\partial n}\,dS \\ &\quad= \iint \left(\frac{h_1}{h_2h_3}\frac{\partial\psi}{\partial\xi}\,d\eta\,d\zeta + \frac{h_2}{h_1h_3}\frac{\partial\psi}{\partial\eta}\,d\xi\,d\zeta + \frac{h_3}{h_1h_2}\frac{\partial\psi}{\partial\zeta}\,d\xi\,d\eta\right) \\ &\quad= \iiint_{V_1} \left(\frac{\partial}{\partial\xi}\frac{h_1}{h_2h_3}\frac{\partial\psi}{\partial\xi} + \frac{\partial}{\partial\eta}\frac{h_2}{h_1h_3}\frac{\partial\psi}{\partial\eta} + \frac{\partial}{\partial\zeta}\frac{h_3}{h_1h_2}\frac{\partial\psi}{\partial\zeta}\right)d\xi\,d\eta\,d\zeta \end{aligned} \tag{69.4}$$

The last integral results from a further application of Green's theorem. Since the volume is essentially positive, and the volume V_1 can be taken arbitrarily small, it is to be concluded from (69.3) and (69.4) that

$$\begin{aligned} &\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} \\ &\qquad= h_1h_2h_3\left(\frac{\partial}{\partial\xi}\frac{h_1}{h_2h_3}\frac{\partial\psi}{\partial\xi} + \frac{\partial}{\partial\eta}\frac{h_2}{h_1h_3}\frac{\partial\psi}{\partial\eta} + \frac{\partial}{\partial\zeta}\frac{h_3}{h_1h_2}\frac{\partial\psi}{\partial\zeta}\right) \end{aligned} \tag{69.5}$$

This formula makes it possible to express the Laplacian operator in terms of any orthogonal system of curvilinear coordinates when the line element is known. It is usually a simple matter to determine h_1, h_2, and h_3.

As an illustrative example, let us make a transformation to spherical coordinates. The square of the line element is easily found to be

$$(ds)^2 = (dr)^2 + r^2(d\theta)^2 + r^2 \sin^2 \theta \, (d\varphi)^2$$

Consequently $h_1 = 1$, $h_2 = 1/r$, $h_3 = 1/(r \sin \theta)$. Substituting these quantities in (69.5) gives

$$\begin{aligned}\Delta\psi &= \frac{1}{r^2 \sin \theta}\left(\frac{\partial}{\partial r} r^2 \sin \theta \frac{\partial \psi}{\partial r} + \frac{\partial}{\partial \theta} \sin \theta \frac{\partial \psi}{\partial \theta} + \frac{\partial}{\partial \varphi} \frac{1}{\sin \theta} \frac{\partial \psi}{\partial \varphi}\right) \\ &= \frac{1}{r^2 \sin \theta}\left(\sin \theta \frac{\partial}{\partial r} r^2 \frac{\partial \psi}{\partial r} + \frac{\partial}{\partial \theta} \sin \theta \frac{\partial \psi}{\partial \theta} + \frac{1}{\sin \theta} \frac{\partial^2 \psi}{\partial \varphi^2}\right)\end{aligned}$$

It is thus seen that with eq. (69.5) available, this transformation is a relatively simple matter compared with the tedious manipulations required when the method outlined in Sec. 13 is used.

70. Stokes' theorem. It was seen in Sec. 59 that Green's theorem in two dimensions allows one to express the integral over a plane domain in terms of a line integral over the closed bounding plane curve. Again in Sec. 67, Green's theorem in three dimensions allows one to express an integral over a three-dimensional domain in terms of a surface integral over the closed surface bounding this domain. Stokes' theorem makes it possible to express a surface integral over a given surface (not closed) in terms of a line integral over the closed space curve which bounds the surface.

If C is the closed space curve which bounds the two-sided surface S (see Sec. 46),

$$\begin{aligned}\int_C (P\,dx + Q\,dy + R\,dz) = \int\int_{\Sigma_{yz},\Sigma_{zx},\Sigma_{xy}} \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy\,dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dz\,dx \\ + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy \qquad (70.1)\end{aligned}$$

where P, Q, and R, together with the indicated partial derivatives, are continuous functions of x, y, z, and C is traversed in such a direction that the side of S over which the surface integral is taken is on the left of a person traversing C in that direction.

A more accurate way to determine the correct direction to traverse C is as follows. Suppose for definiteness the surface integral is to be taken over the upper side of the surface (the side for which the normal to the tangent plane makes an acute angle with the positive z axis). Erect a

positive normal n to S at any point and let the base of the normal move over the same side of the surface until it is coincident with a point on the bounding curve C. Then at this point draw an inner normal N to the curve C lying in the tangent plane to the surface at this point. Then draw the tangent T to C such that N, n, and T will bear the same orientation to each other as the x, y, and z axes. Then T will point in the direction in which C must be traversed.

The relation (70.1) may be established in parts just as in the case of Green's theorem. It may be shown that

$$\int_C P\,dx = \iint_{\Sigma_{xz},\Sigma_{xy}} \left(\frac{\partial P}{\partial z}\,dz\,dx - \frac{\partial P}{\partial y}\,dx\,dy\right) \tag{70.2}$$

It will be supposed that S is defined by the equation $z = \varphi(x,y)$ and that the projection of S on the xy plane is the region $\sum_{xy}$ bounded by the

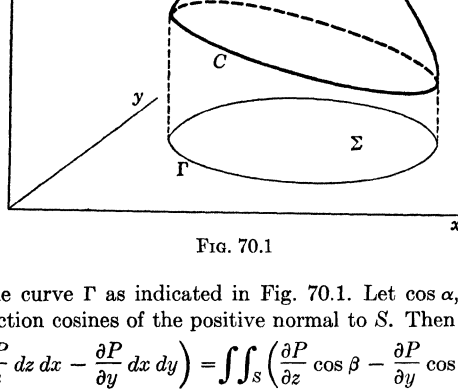


FIG. 70.1

closed plane curve Γ as indicated in Fig. 70.1. Let $\cos\alpha$, $\cos\beta$, $\cos\gamma$ be the direction cosines of the positive normal to S. Then from (65.5),

$$\int_{\Sigma_{xz}}\int_{\Sigma_{xy}} \left(\frac{\partial P}{\partial z}\,dz\,dx - \frac{\partial P}{\partial y}\,dx\,dy\right) = \iint_S \left(\frac{\partial P}{\partial z}\cos\beta - \frac{\partial P}{\partial y}\cos\gamma\right) dS$$

$$= \iint_{\Sigma_{xy}} \left(\frac{\partial P}{\partial z}\frac{\cos\beta}{\cos\gamma} - \frac{\partial P}{\partial y}\right) dx\,dy \tag{70.3}$$

Comparing eqs. (45.11) and (45.2), it is seen that $\cos\beta/\cos\gamma = -\partial z/\partial y$, so that

$$\frac{\partial P}{\partial z}\frac{\cos\beta}{\cos\gamma} - \frac{\partial P}{\partial y} = -\frac{\partial P}{\partial z}\frac{\partial z}{\partial y} - \frac{\partial P}{\partial y} = -\frac{\partial P_1}{\partial y}$$

where $P_1(x,y) = P[x,y,z(x,y)]$. Consequently (70.3) becomes

$$\iint_{\Sigma_{xz},\Sigma_{xy}} \left(\frac{\partial P}{\partial z}\,dz\,dx - \frac{\partial P}{\partial y}\,dx\,dy\right) = -\iint_{\Sigma_{xy}} \frac{\partial P_1}{\partial y}\,dx\,dy = \int_\Gamma P_1\,dx$$

The last step follows from Green's theorem in two dimensions by placing $C = 0$ in (59.3). Since P has the same value at a point on C as P_1 has for the projection of this point, it follows that $\int_\Gamma P_1\,dx = \int_C P\,dx$. This, together with the above relations, establishes (70.2). The other parts of (70.1) can be handled in a similar manner.

SUGGESTED REFERENCES

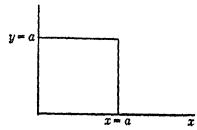
W. F. Osgood, *Advanced Calculus*, New York, The Macmillan Company, 1925.

R. Courant, *Differential and Integral Calculus*, New York, Nordemann Publishing Company, 1936.

PROBLEMS

53.1. Determine whether $f(x) = x \cos \pi/x$ has bounded variation in the region $-1 < x < 1$.

54.1. Evaluate the line integral $\int_C (x + y)dx + xy\,dy$, where: (1) the path C is the straight line connecting the points $(x = 0, y = 1)$ and $(x = 1, y = 0)$; (2) the path C is the boundary of the accompanying square.



54.2. Evaluate the line integral $\int_C (x\,dy - y\,dx)/(x^2 + y^2)$, where: (1) the path C is a circle of radius a with center at the origin: (2) the path C is the boundary of a ring bounded by circles of radii $a < b$.

54.3. Evaluate the line integral $\int_C (2x^2 + 3xy)ds$ where C is the part of the circle $x^2 + y^2 = 1$ lying in the first quadrant.

58.1. Evaluate $\iint xy\,dx\,dy$ over the domain bounded by the axes and the curves $y = a$, $x^2 = 4ay$.

Compute the masses of the following objects if ρ is the mass per unit area.

58.2. A right triangle, where ρ is proportional to the distance from an acute angle vertex.

58.3. A semicircle, where ρ is proportional to the distance from the bounding diameter.

Compute the moments of inertia of the following objects of uniform density.

58.4. A rectangular plate, sides of length a and b, about an axis through the center perpendicular to the plane of the rectangle.

58.5. A cylinder, radius r, length l, about its axis.

59.1. Show that

$$\iint_D \frac{\partial(u,v)}{\partial(x,y)}dx\,dy = \int_C u\,dv,$$

where C is the boundary curve of the domain D. *Hint*: Let $P = u\dfrac{\partial v}{\partial x}$, $Q = u\dfrac{\partial v}{\partial y}$ in Green's theorem.

59.2. Show that

$$\iint_D u\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) + \left(\frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y}\right)dx\,dy = -\int_C u\frac{\partial v}{\partial n}\,dS$$

where $\partial v/\partial n$ denotes the derivative in the normal direction to the curve C bounding the domain D.

59.3. If $\Delta w = \dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2}$, show that

$$\int_D \int (u\,\Delta v + v\,\Delta u)dx\,dy = -\int_C \left(u\frac{\partial v}{\partial n} - v\frac{\partial u}{\partial n}\right)ds$$

60.1. Find the moment of inertia of a sphere of uniform density about a diameter.

60.2. A hydrogen atom in a certain excited state may be regarded as a proton surrounded by a distribution of negative charge of density

$$-\rho = \left(\frac{1}{2a}\right)^{5/2}\frac{2r}{\sqrt{3}}e^{-r/2a}\cos\theta$$

compute (1) the total charge, and (2) the first moment of this charge distribution.

61.1. The electrons emitted from the cylindrical surface of radius r at a temperature T have the velocity distribution function

$$f(u,v)du\,dv = Ce^{(-m/2kT)(u^2+v^2)}\,du\,dv$$

where u and v are the components of velocity in the radial direction and tangential direction, respectively. Between the cylindrical surface of radius r and a coaxial cylindrical surface of radius $R > r$ there exists the difference of potential φ. Let the corresponding velocity components at the outer surface be denoted by U and V as shown. Compute the distribution function $F(U,V)dU\,dV$ giving the distribution of velocities of the electrons arriving at the outer surface. Show that if $0 < u$ and $-\infty < v < +\infty$ the possible values of U and V lie on or within the hyperbola

$$U^2 - V^2\left(\frac{R^2}{r^2} - 1\right) = 2\frac{e}{m}\varphi$$

Hint: Regard this as a transformation from the w plane to the UV plane.

Compute the element of area corresponding to the following coordinates.

61.2. Elliptic coordinates [see eq. (50.2)].

61.3. Parabolic coordinates.

61.4. Polar coordinates.

62.1. The integral of $x^p y^q$ over the triangle bounded by the coordinate axes and the line $x + y = 1$ is one form of Dirichlet's integral. Show that it is equal to

$$\frac{\Gamma(p+1)\Gamma(q+1)}{\Gamma(p+q+3)}$$

62.2. Evaluate $\iint_A x^p y^q\,dx\,dy$, where A is the area bounded by the coordinate axes and the curve $(x/\alpha)^a + (y/\beta)^b = 1$.

62.3. Evaluate $\iiint_V x^A y^B z^C\,dx\,dy\,dz$, over a sphere of radius R. Hence find the moment of inertia of a sphere about an axis through its center.

62.4. Find the moment of inertia of an ellipsoid about one of its principal axes.

62.5. Show that $\int_0^1 \frac{x^{2n}}{\sqrt{1-x^2}}\,dx = \frac{\sqrt{\pi}}{2}\cdot\frac{\Gamma(n+\frac{1}{2})}{\Gamma(n+1)}$.

62.6. Show that $B(p,q) = B(p+1,q) + B(p,q+1)$.

64.1. Compute the surface area of the surface $x = u\cos v$, $y = u\sin v$, $z = a^2 - u^2$ above the xy plane by two methods.

66.1. Given a cylinder of radius a and length l of moment of inertia about its axis of I, initially spinning about its axis with angular velocity w. If the air frictional force exerted on the cylindrical surface is proportional to the peripheral velocity, determine the length of time required for the cylinder to stop spinning.

66.2. A thrown ball is spinning about an axis perpendicular to its direction of flight. Assuming that Bernoulli's law ($\frac{1}{2}\rho v^2 + p = \text{const.}$) holds for a thin layer of air dragged with the surface of the ball, find the force on the ball. Here ρ is air density, v is surface velocity, and p is the pressure. *Hint*: neglect gravity and approximate the ball by a more convenient shape if too difficult. Find roughly the deviation of its path from a straight line.

67.1. Use Green's theorem to prove Archimedes' principle for a general gravitational field. *Hint*: Use Bernoulli's equation.

67.2. Show that the time rate of change of mass of fluid in a given volume is equal to

$$\iiint_V \left(\frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z}\right)dx\,dy\,dz$$

69.1. Show that for an orthogonal transformation

$$\left|\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)}\right|^2 = L_1L_2L_3$$

Hint: Transpose one of the determinants on the left and use the rule for multiplication of determinants.

69.2. Express $\partial^2x/\partial x^2 + \partial^2x/\partial y^2 + \partial^2x/\partial z^2$ in terms of parabolic coordinates [see eq. (49.2)].

70.1. The circulation of a fluid around a closed curve C is defined as

$$\Gamma = \int_C v_x\,dx + v_y\,dy + v_z\,dz,$$

where v_x, v_y, v_z are the Cartesian components of velocity of the fluid. Show under what conditions the flow is irrotational or vortex free, i.e., $\Gamma = 0$. Further show that under these conditions

$$\int_A^B v_x\,dx + v_y\,dy + v_z\,dz$$

is independent of the path connecting points A and B.

Chapter 6

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

71. Concept of a complex number. The introduction of complex numbers into mathematics had its origin in the circumstance that quadratic equations fell into two classes, one comprising all equations for which solutions existed, and the other containing those for which no solution existed. As an example of the first class may be cited the equation $x^2 - 1 = 0$, while the equation $x^2 + 1 = 0$ falls into the second class, since no ordinary number could be found which when squared gave -1. This lead to the introduction of a more general number system to supplement the system of real numbers so that when this system was used all quadratic equations had solutions. Numbers belonging to this more general class are called complex numbers. In defining the system of complex numbers it is necessary to consider the totality of all number pairs (a,b) where a and b are real numbers. Since any such pair determines a point in a plane, it is convenient, but by no means necessary, to think of the point specified by the rectangular coordinates a and b as shown in Fig. 71.1. Corresponding to each number pair (a,b) there is associated the combination $a + bi$ in which the letter i is to be regarded as a pure symbol. The combination $a + bi$ is called a complex number since it will have the character of a number when rules for performing additions, multiplications, etc., with two such combinations are formulated. It is convenient to denote the complex number $a + bi$ by a single letter α. If the domain of complex numbers is really to supplement the domain of real numbers, the rules of calculation with such numbers must reduce to those for calculating with real numbers whenever $b = 0$. For this reason, when $b = 0$ and the complex number $\alpha = a$, the numbers a are real. When representing

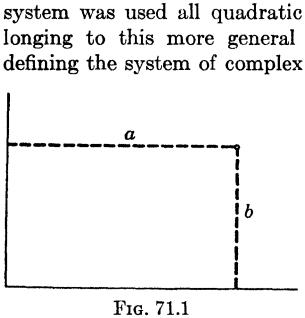


FIG. 71.1

complex numbers in a plane, the axis of a is called the real axis. When $a = 0$ the complex number $\alpha = bi$ is said to be pure imaginary, and the axis along which b is plotted is called the imaginary axis; i is then the unit of imaginary numbers.

Rules for Calculation. The laws for combining complex numbers will be given.

The *addition* of complex numbers is defined by the relation

$$\left.\begin{aligned} \alpha + \beta &= (a + bi) + (a' + b'i) \\ &= (a + a') + (b + b')i \\ \alpha - \beta &= (a + bi) - (a' + b'i) \\ &= (a - a') + (b - b')i \end{aligned}\right\} \quad (71.1)$$

The *multiplication* of complex numbers is defined by the relation

$$\begin{aligned} \alpha \cdot \beta &= (a + bi)(a' + b'i) \\ &= (aa' - bb') + (ab' + ba')i \end{aligned} \quad (71.2)$$

Two special cases of this definition should be noted.

(1) When $a = a' = 0$ and $b = b' = 1$; then

$$i^2 = -1 \quad (71.3)$$

This shows that the square of the unit of imaginary numbers yields a negative real number.

(2) When $a' = a$ and $b' = -b$; then

$$(a + bi)(a - bi) = a^2 + b^2 \quad (71.4)$$

The two numbers $a + bi$ and $a - bi$ are called conjugate complex numbers. If $a + bi$ is denoted by α, the conjugate is usually denoted by either α^* or $\bar{\alpha}$. In a complex number the conjugate can be found in general by replacing i by $-i$ wherever it appears.

The commutative and associative laws of multiplication hold for complex numbers.

A complex number is zero only when both a and b are zero.

Division of Complex Numbers. A complex number β can be divided by a complex number α if α is different from zero. To define the quotient γ of the two numbers α and β, we start from the relation $\alpha\gamma = \beta$ or $(a + bi)\gamma = a' + b'i$. Multiply both sides of the last expression by $(a - bi)$ and obtain

$$(a^2 + b^2)\gamma = (a - bi)(a' + b'i) = \alpha^*\beta$$

or

$$\gamma = \frac{\alpha^*\beta}{a^2 + b^2} = \frac{(aa' + bb')}{a^2 + b^2} + \frac{(ab' - ba')}{a^2 + b^2}\, i$$

72. Geometric representation of complex numbers. It is expedient to consider certain further geometrical aspects of complex numbers.

As already noticed, every point on a plane with rectangular coordinates (a,b) determines a complex number α. The plane whose points serve to represent the complex numbers is called the complex plane. The origin of the coordinate system coincides with the point $\alpha = 0$ and is called the zero point.

Coordinate systems other than rectangular are sometimes convenient in specifying complex numbers; in fact the polar system is very useful in this connection. The familiar transformation

$$a = \rho \cos \varphi \qquad b = \rho \sin \varphi$$

serves to relate the two systems. In the polar system

$$\alpha = a + bi = \rho(\cos \varphi + i \sin \varphi)$$

where ρ is the length of the radius vector to the point α. Here ρ is called the absolute value of α, which may be written $\rho = |\alpha|$. From the geometry of Fig. 72.1 it is evident that $|\alpha| = \rho = \sqrt{a^2 + b^2}$. From eq. (71.4) it follows that $|\alpha| = \rho = \sqrt{\alpha\alpha^*}$. It is therefore seen that all complex numbers having the same absolute value lie on a circle in the complex plane. The angle φ is called the argument of the complex number.

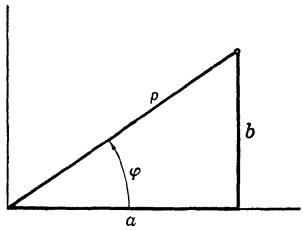


FIG. 72.1

Still another mode of representing complex numbers is obtained when one replaces $\cos \varphi$ by $(e^{i\varphi} + e^{-i\varphi})/2$ and $\sin \varphi$ by $(e^{i\varphi} - e^{-i\varphi})/2i$. The three modes of expression become

$$\alpha = a + bi = \rho(\cos \varphi + i \sin \varphi) = \rho e^{i\varphi}$$

where $$\rho = \sqrt{a^2 + b^2}, \qquad \varphi = \tan^{-1} b/a.$$

It should be pointed out that α is uniquely defined, represented in the form $\alpha = \rho e^{i\varphi}$, for $-\pi < \varphi \leq \pi$. This limitation on φ avoids the confusion as to whether $-1 = e^{i\pi}$ or $e^{-i\pi}$.

Stereographic Projection. Still another way of representing complex numbers is often of convenience, namely that in which the numbers are represented on the surface of a sphere. In order to obtain the connection between the representation on the complex plane and the sphere, let it be supposed that a sphere rests upon the complex plane so that one of its diameters passes through the zero point of the plane as shown in Fig. 72.2. Then draw a line from the opposite end N of this diameter to a point in the complex plane representing the arbitrary complex number α. This line will intersect the surface of the sphere in one point other than N. This point may be taken as the representation of the

complex number α. It is readily seen that numbers lying on circles in the complex plane also lie on circles on the number sphere. Furthermore, it can be shown that two intersecting curves in the plane go over into intersecting curves on the sphere such that the angle of intersection is the same as for the curves in the plane. In other words, the transformation from plane to sphere is conformal. One of the conveniences of the number sphere lies in the fact that the point N corresponds to any complex number having an infinite absolute value.

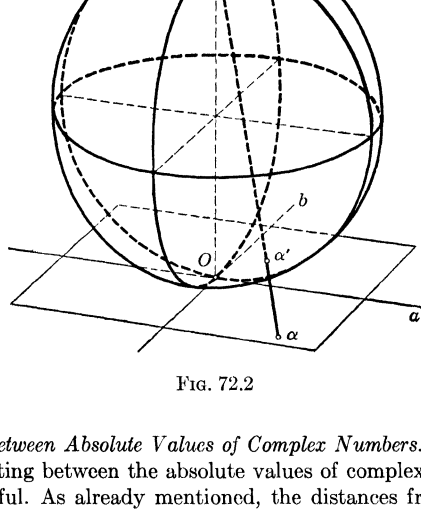


FIG. 72.2

Relations between Absolute Values of Complex Numbers. A few of the relations existing between the absolute values of complex numbers will be found useful. As already mentioned, the distances from the origin to the points in the complex plane representing the complex numbers α and β are equal to $|\alpha|$ and $|\beta|$. The distance between the points α and β is just $|\alpha - \beta|$ since

$$|\alpha - \beta| = \sqrt{(a - a')^2 + (b - b')^2},$$

where $\alpha = a + bi$ and $\beta = a' + b'i$. From the geometry of Fig. 72.3a and 72.3b, it is readily seen that the following inequalities are satisfied.

$$|\alpha - \beta| \geq |\alpha| - |\beta| \quad \text{and} \quad |\alpha + \beta| \leq |\alpha| + |\beta|$$

Given two complex numbers, then $|\alpha\beta| = |\alpha|\,|\beta|$, since

$$|\alpha\beta| = \sqrt{\alpha\beta\,\alpha^*\beta^*} = \sqrt{\alpha\alpha^*}\sqrt{\beta\beta^*} = |\alpha|\,|\beta|$$

In general

$$|\alpha\cdot\beta\cdot\ \ldots\ \lambda| = |\alpha|\,|\beta|\ \ldots\ |\lambda| \quad \text{and} \quad |\alpha^n| = |\alpha|^n$$

Also

$$\left|\frac{\alpha}{\beta}\right| = \frac{|\alpha|}{|\beta|}$$

since

$$\left|\frac{\alpha}{\beta}\beta\right| = \left|\frac{\alpha}{\beta}\right|\,|\beta| = |\alpha|$$

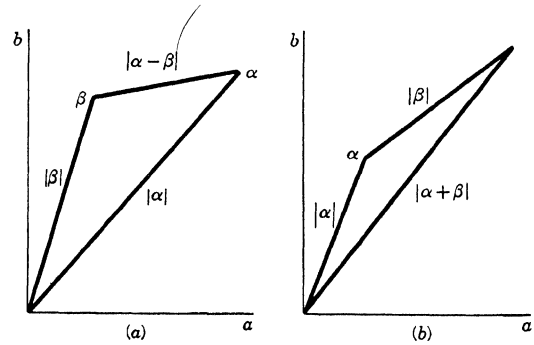


FIG. 72.3

73. Infinite series of complex numbers. The useful facts concerning infinite series can just as well be given when the terms of the series are complex numbers as when they are all real, for this becomes a special case. Consider the series of complex numbers

$$\alpha_1 + \alpha_2 + \alpha_3 + \ \ldots\ + \alpha_n + \ \ldots \tag{73.1}$$

where $\alpha_n = a_n + b_n i$. Such a sum of terms may or may not have a finite value. To determine whether it does or not and what the complex number is which is the sum, one considers the so-called partial sums

$$\begin{aligned}
\sigma_1 &= \alpha_1 \\
\sigma_2 &= \alpha_1 + \alpha_2 \\
&\ \ldots\ \ldots\ \ldots\ \ldots \\
\sigma_n &= \alpha_1 + \alpha_2 + \ \ldots\ + \alpha_n
\end{aligned}$$

If $\lim_{n\to\infty} \sigma_n$ exists and is equal to the complex number σ, the series (73.1) is said to converge, and its sum is σ. It is obvious that the series (73.1) can converge when and only when both of the series

$$\left.\begin{aligned} a &= a_1 + a_2 + \ \ldots\ a_n + \ \ldots \\ b &= b_1 + b_2 + \ \ldots\ b_n + \ \ldots \end{aligned}\right\} \tag{73.2}$$

converge, whence $\sigma = a + bi$.

Another criterion for the convergence of (73.1) can be given, which states that the series (73.1) converges when and only when, for a given real positive number ϵ, a value of the index n can be found for which

$$| \alpha_{n+1} + \alpha_{n+2} + \ldots + \alpha_{n+k} | < \epsilon$$

for any value of k whatever.

The series (73.1) is said to converge unconditionally if it converges to the same sum σ for an arbitrary rearrangement of the terms. It is therefore to be born in mind that unless it is known that a series converges unconditionally one is not at liberty to rearrange the terms. The necessary and sufficient condition for unconditional convergence is that the two series (73.2) converge absolutely, which means that the two series

$$\left.\begin{aligned} &| a_1 | + | a_2 | + \ldots + | a_n | + \ldots \\ &| b_1 | + | b_2 | + \ldots + | b_n | + \ldots \end{aligned}\right\} \quad (73.3)$$

converge. Since

$$\begin{aligned} | \alpha_n | &= \sqrt{a_n^2 + b_n^2} \\ &= | a_n + b_n i | \leq | a_n | + | b_n i | = | a_n | + | b_n | \end{aligned}$$

it is seen that the series

$$| \alpha_1 | + | \alpha_2 | + \ldots + | \alpha_n | + \ldots \quad (73.4)$$

converges whenever the series (73.3) converge, or in other words the series (73.1) converges absolutely. Furthermore, since $| a_n |$ or $| b_n |$ is not greater than $| \alpha_n |$, the series (73.1) is unconditionally convergent whenever the series (73.4) converges. Methods for testing the convergence of series are given in Sec. 76.

Combinations of Series of Complex Numbers. The sum of two convergent series

$$\sum_{k=1}^{\infty} \alpha_k \text{ and } \sum_{k=1}^{\infty} \beta_k \text{ may be written } \sum_{k=1}^{\infty} (\alpha_k + \beta_k)$$

and is likewise a convergent series having a sum equal to the sum of the individual series.

The product of two convergent series

$$\sum_{k=1}^{\infty} \alpha_k \quad \text{and} \quad \sum_{k=1}^{\infty} \beta_k$$

written in the form

$$\sum_{k=1}^{\infty} (\alpha_1\beta_k + \alpha_2\beta_k + \ldots + \alpha_n\beta_k + \ldots)$$

is convergent *provided* in addition at least one of the original series is

absolutely or unconditionally convergent, in which case the sum of the above series is equal to the product of the sums of the individual series.

74. Complex variables and functions of the same. Consider a set D of complex numbers and suppose that it is agreed that the complex number $z = x + iy$ is used to identify each number of the set D; then z is a complex variable and D its domain.

If to each value that z may assume, that is, every number of D, there is associated another complex number $w = u + iv$ according to a definite law, w is said to be a function of the complex variable z. This means that if $z = x + iy$, then u and v are real functions of the two real variables x and y. Geometrically speaking this means that to every point z in the domain D of the z complex plane or number sphere there corresponds a point w in the domain D' of the w complex plane or number sphere. It may happen that the domain D' may contain a value of w several times if the same value of w is associated with several different values of z. If the number sphere representation is used the point at infinity need not be excluded from the domain.

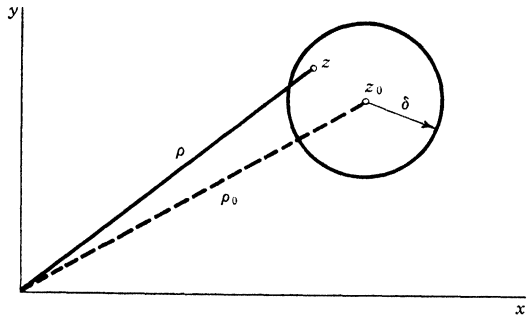


FIG. 74.1

The function w, which assumes the value w_0 for $z = z_0$ is continuous at $z = z_0$ if for each arbitrarily small positive number ϵ, it is possible to find a positive number δ such that, $| w - w_0 | < \epsilon$ when $| z - z_0 | < \delta$.

In order to obtain a clear idea of the meaning of the above concept it is instructive to examine the continuity of the function $w = z^n$ where n is an integer. Let z_0 be any point other than $z = \infty$. Then, if $w_0 = z_0^n$, the value of $w - w_0$ may be written

$$\begin{aligned} w - w_0 &= z^n - z_0^n \\ &= (z - z_0)(z^{n-1} + z^{n-2} z_0 + z^{n-3} z_0^2 + \ldots + z_0^{n-1}) \end{aligned}$$

Using the inequalities satisfied by the absolute values given in Sec. 72, the following inequality holds.

$$| w - w_0 | \leq | z - z_0 | (| z |^{n-1} + | z |^{n-2} | z_0 | + \ldots + | z_0 |^{n-1})$$

Using z_0 as a center, describe a circle of radius δ so as to include the point z within it as shown in Fig. 74.1. Then

$$| z - z_0 | < \delta \quad \text{and} \quad \rho = | z |, \ \rho_0 = | z_0 | \quad \text{and} \quad \rho < \rho_0 + \delta$$

so that

$$\begin{aligned} | w - w_0 | &\leq \delta[(\rho_0 + \delta)^{n-1} + (\rho_0 + \delta)^{n-2}(\rho_0 + \delta) + \ldots + (\rho_0 + \delta)^{n-1}] \\ &\leq \delta n(\rho_0 + \delta)^{n-1} \end{aligned}$$

From this inequality, it is evident that δ can be chosen so small that $| w - w_0 |$ can be made smaller than any previously assigned positive number ϵ, thus establishing the continuity of z^n.

It can likewise be shown that the entire function

$$w = a_0 z^n + a_1 z^{n-1} + \ldots + a_n$$

is a continuous function.

75. Uniform convergence. In the preceding sections, infinite series of complex numbers were considered. It is extremely important to consider the more general series in which each term is itself a function of the complex variable z. Thus we have to do with the series

$$w = w_1 + w_2 + w_3 + \ldots + w_n + \ldots \tag{75.1}$$

First of all, the functions w_1, w_2, etc. must have a domain of definition which is common. As regards convergence, the series (75.1) may converge for *all, some, one,* or *no* values of z in the common domain D. The domain of z for which (75.1) converges is called the domain of convergence.

If the series (75.1) has the domain of convergence D, then for values of z in this domain, $| r_k | < \epsilon$ for $k > n$, where ϵ is an arbitrarily small selected positive number and

$$r_k = w - (w_1 + w_2 + \ldots + w_k)$$

In general the value of n which must be chosen so that $r_k < \epsilon$ will be different for every value of z in D. If, however, there exists one fixed index N such that $n < N$ for all z in D, the series (75.1) is said to converge uniformly.

In order to understand the implications of the above restrictions to secure uniform convergence let us examine the series

$$\sum_{\nu=1}^{\infty} \frac{x^2}{(1 + x^2)^\nu}$$

Using methods given in the following sections, it can be established that this series converges for all values of x except $x = 1$. Restricting the domain of x to that for which $-1 \leq x \leq 1$, we inquire whether the above series converges uniformly in this domain. Proceeding as outlined above

$$r_n = w - \left[\frac{x^2}{(1+x^2)} + \frac{x^2}{(1+x^2)^2} + \ldots + \frac{x^2}{(1+x^2)^n}\right]$$

$$= \frac{x^2}{(1+x^2)^{n+1}} + \frac{x^2}{(1+x^2)^{n+2}} + \ldots + \frac{x^2}{(1+x^2)^{n+k+1}} + \ldots$$

$$= \frac{x^2}{(1+x^2)^{n+1}}\left[1 + \frac{1}{1+x^2} + \frac{1}{(1+x^2)^2} + \ldots + \frac{1}{(1+x^2)^k} + \ldots\right]$$

$$= \frac{x^2}{(1+x^2)^{n+1}} \frac{1}{1 - 1/(1+x^2)} = \frac{1}{(1+x^2)^n}$$

Choosing a small positive number ϵ, the problem is to see if there exists a fixed n for which $r_n = 1/(1+x^2)^n < \epsilon$ for all values of x in the interval $-1 < x < 1$. It is evident that if $x \to 0$, then $n \to \infty$ in order that $r_n < \epsilon$, so that no fixed n exists for x in the domain considered, and the series is not uniformly convergent in this interval.

The importance of the concept of uniform convergence rests on the theorem that if the individual terms of the series

$$w = w_1 + w_2 + \ldots + w_n + \ldots$$

are continuous functions of z in the domain D, and if the series converges *uniformly* in this domain, the sum w is a *continuous* function of z in D.

76. Tests for convergence of a series. Given the series

$$\alpha_1 + \alpha_2 + \alpha_3 + \ldots + \alpha_n + \ldots$$

of complex numbers to test for convergence. First let

$$\sigma_n = \alpha_1 + \alpha_2 + \ldots + \alpha_n$$

$$\sigma_{n,k} = \alpha_{n+1} + \alpha_{n+2} + \ldots + \alpha_{n+k}$$

The fundamental definition for convergence given in Sec. 73 requires that an n exist such that $|\sigma_{n,k}| < \epsilon$ for all positive k. Thus if a value of k can be found for which this inequality fails, the series certainly diverges. To illustrate the use of this test, consider the series

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots + \frac{1}{n} + \ldots \tag{76.1}$$

Then

$$\sigma_{n,n} = \frac{1}{n+1} + \frac{1}{n+2} + \ldots + \frac{1}{n+n} > \frac{1}{2n} + \frac{1}{2n} + \ldots + \frac{1}{2n} = \frac{n}{2n} = \frac{1}{2}$$

so that $|\sigma_{n,n}| > \frac{1}{2}$; hence in particular for $k = n$, the term $|\sigma_{n,n}|$ is not less than ϵ, and the series diverges. This may be checked by calculating the sum of 2^{n+1} terms. This sum is

$$\begin{aligned}\sigma_{2^{n+1}} &= 1 + \sigma_{1,1} + \sigma_{2,2} + \sigma_{4,4} + \sigma_{8,8} + \ldots + \sigma_{2^n 2^n} \\ &> 1 + \frac{1}{2} + \frac{1}{2} + \ldots + \frac{1}{2} = \frac{3}{2} + \frac{n}{2} = \frac{1}{2}(n+3)\end{aligned}$$

As pointed out in Sec. 73, the series would converge if the $\lim_{n\to\infty} \sigma_{2^{n+1}}$ exists, but the

$$\lim_{n\to\infty} \sigma_{2^{n+1}} = \infty$$

and the series diverges.

As another example, consider the series

$$\sum_{n=1}^{\infty} \frac{1}{n^x} \quad \text{when} \quad x > 1 \tag{76.2}$$

This series, when considered as a function of the complex variable z instead of x, is known as the Riemann zeta function. The sum of n terms is

$$\sigma_n = \sum_{m=1}^{n} \frac{1}{m^x}$$

It is to be noticed that

$$\frac{1}{2^x} + \frac{1}{3^x} < \frac{2}{2^x} = \frac{1}{2^{x-1}}$$

$$\frac{1}{4^x} + \frac{1}{5^x} + \frac{1}{6^x} + \frac{1}{7^x} < \frac{4}{4^x} = \frac{1}{4^{x-1}}$$

Then

$$\begin{aligned}\sigma_{2^p-1} &= \sum_{m=1}^{2^p-1} \frac{1}{m^x} < \frac{1}{1^{x-1}} + \frac{1}{2^{x-1}} + \frac{1}{4^{x-1}} + \frac{1}{8^{x-1}} + \ldots + \frac{1}{[2^{(p-1)}]^{x-1}} \\ &< \frac{1}{1 - 2^{1-x}}\end{aligned}$$

Since the right-hand side is independent of p, then $\lim_{p\to\infty} \sigma_{2^p-1}$ is less than $(1 - 2^{1-x})^{-1}$, and the series converges for $x > 1$. The series (76.1) and (76.2) are of use in determining the convergence of other series by using the comparison test.

Comparison Test. Given two series

$$W_1 + W_2 + W_3 + \ldots + W_n + \ldots \tag{76.3}$$

$$w_1 + w_2 + w_3 + \ldots + w_n + \ldots \tag{76.4}$$

each of whose terms may be a function of z in a certain domain D. If for every z in D, $|W_n| \geq |w_n|$ and the series

$$|W_1| + |W_2| + \ldots + |W_n| + \ldots$$

converges, the series

$$w_1 + w_2 + \ldots + w_n + \ldots$$

converges absolutely, or if the series

$$|W_1| + |W_2| + \ldots + |W_n| + \ldots$$

converges uniformly in D, the series (76.4) converges absolutely and uniformly in D.

On the other hand, if it is known that

$$|W_1| + |W_2| + \ldots + |W_n| + \ldots$$

diverges and $|W_n| \leq |w_n|$, the series

$$|w_1| + |w_2| + \ldots + |w_n| + \ldots$$

certainly diverges. As an example of the use of the comparison test, consider the series

$$\sum_{n=1}^{\infty} \frac{1}{n^n}$$

Since the terms are positive and all less than the corresponding terms in the series

$$\sum_{n=1}^{\infty} \frac{1}{n^x} \quad \text{for} \quad 1 < x < 2$$

which is known to converge, the series under consideration converges absolutely.

Cauchy's Ratio Test for Absolute Convergence. The series

$$w_1 + w_2 + w_3 + \ldots + w_n + \ldots$$

converges absolutely provided

$$\lim_{n \to \infty} \left| \frac{w_{n+1}}{w_n} \right| = r < 1$$

If the limit r happens to be unity, it cannot be concluded that the series converges or diverges, but a special test to be given below must be applied. On the other hand, if the ratio

$$\left| \frac{w_{n+1}}{w_n} \right|$$

remains greater than unity for all values of n greater than some fixed N, the series diverges, so that the questionable case is that for which $r = 1$.

As an example of the use of this test, consider the series

$$\sum_{n=1}^{\infty} \frac{n}{2^{n-1}}$$

The ratio of the $(n+1)$st term to the nth term is

$$\frac{(n+1)/2^n}{n/(2^{n-1})} = \frac{n+1}{n}\frac{1}{2} \quad \text{and} \quad \lim_{n\to\infty} \frac{n+1}{n}\frac{1}{2} = \frac{1}{2}$$

and the series converges absolutely. As a further example consider the series

$$\frac{1}{1\cdot 2} + \frac{1}{3\cdot 4} + \frac{1}{5\cdot 6} + \cdots + \frac{1}{(2n-1)(2n)} + \cdots$$

In this case the ratio is

$$\frac{1/[(2n+1)2n+2]}{1/(2n-1)(2n)} = \frac{n(2n-1)}{(2n+1)(n+1)} = \frac{2n^2-n}{2n^2+3n+1}$$

$$= \frac{1-1/2n}{1+3/2n+1/2n^2}$$

and

$$\lim_{n\to\infty} \left| \frac{1-1/2n}{1+3/2n+1/2n^2} \right| = 1$$

Since the limit is exactly 1, no conclusion can be drawn from this test.

Test when $r = 1$. In the case where

$$\lim_{n\to\infty} \left| \frac{w_{n+1}}{w_n} \right| = 1$$

the series will converge absolutely provided a positive number l exists for which

$$\lim_{n\to\infty} n\left[\left| \frac{w_{n+1}}{w_n} \right| - 1 \right] = -1 - l$$

Applying this test to the series given above it is found that

$$n\left[\left| \frac{w_{n+1}}{w_n} \right| - 1 \right] = n\,\frac{-2/n - 1/2n^2}{1+3/2n+1/2n^2} = \frac{-1-1/2n}{1+3/2n+1/2n^2}$$

and

$$\lim_{n\to\infty} n\left[\left| \frac{w_{n+1}}{w_n} \right| - 1 \right] = -2 = -1 - 1$$

Thus $l = 1$ and the original series converges absolutely.

77. Power series and their radii of convergence. A series of the type

$$c_0 + c_1 z + c_2 z^2 + c_3 z^3 + \cdots \tag{77.1}$$

arranged in ascending powers of z is called a power series. These series are of particular importance in the theory of functions of a complex variable because of their functional properties. The first problem of importance is to determine the domain of values of z for which (77.1) converges. For convenience a series of the form (77.1) will be denoted henceforth by $P(z)$. To begin with, it should be noted that the domain

of convergence always includes the point $z = 0$. It may happen, of course, that the domain of convergence is restricted to this one point alone as in the case of the following series.

$$1 + z + 2!z^2 + 3!z^3 + \dots$$

On the other hand, if there exists another point $z = z_0 \neq 0$ for which $P(z)$ converges, it can be shown that $P(z)$ converges absolutely and uniformly for all values of z for which $|z| < |z_0|$; this means for all values of z lying within the circle of radius $|z_0|$ with center at the origin. It may furthermore be shown that for every power series $P(z)$ there exists a circle of radius r with its center at the origin, such that $P(z)$ converges absolutely and uniformly for every point within it, and diverges for every value of z outside. Whether or not $P(z)$ converges for some or no points on the circle of convergence has not been specified, and depends on circumstances.

The radius r of the circle of convergence is called the radius of convergence of $P(z)$. For the particular case when $P(z)$ converges only at $z = 0$, the radius of convergence is said to be zero.

The above theorems insure that every power series $P(z)$ represents a continuous function for values of z within the circle of convergence, since each term is a continuous function and the series converges uniformly.

The problem met with in practice is that of determining the radius of convergence. If $P(z)$ denotes the power series

$$P(z) = c_0 + c_1z + c_2z^2 + c_3z^3 + \dots$$

the method of determining r due to Cauchy and Hadamard makes use of the sequence of nonnegative numbers

$$|c_1|, \quad |c_2|^{1/2}, |c_3|^{1/3}, \dots, |c_n|^{1/n}, \dots$$

They have shown that if the greater limit (often called superior limit) of the above sequence is l, then $r = 1/l$, or in limit notation

$$\overline{\lim_{n\to\infty}} |c_n|^{1/n} = l, \quad \text{then} \quad r = 1/l$$

It sometimes happens that the sequence in question has more than one limit point, and in this case the largest value of l is taken. An example illustrating this will be given later. If l happens to be zero, $r = \infty$ and the series converges for every finite value of z.

Example. Determine the radius of convergence of the series

$$P(z) = \log(1 - z) = z + \frac{z^2}{2} + \frac{z^3}{3} + \dots$$

Here $c_n = 1/n$ and the $\overline{\lim_{n\to\infty}} \left|\frac{1}{n}\right|^{1/n}$ must be computed. Writing

$$\left(\frac{1}{n}\right)^{1/n} = e^{(1/n)\log(1/n)} \quad \text{then} \quad \lim_{n\to\infty}\left(\frac{1}{n}\right)^{1/n} = 1$$

Thus $l = 1$ and $r = 1$, so that the circle of convergence is the unit circle.

Example. Find the radius of convergence of the series

$$P(z) = 1 + z^2 + z^4 + z^9 + \ldots$$

Here we notice that the coefficients c_n belong to two separate sets, one being those for which $n = k^2$ where $k = 0, 1, 2, \ldots$ and those for which $n \neq k^2$. In the first set $c_{k^2} = 1$ for all k, and in the second $c_n = 0$. Thus the sequence $|\, c_n \,|^{1/n}$ has two limit points, since

$$\lim_{k\to 0} |\, 1 \,|^{1/k^2} = 1 \quad \text{and} \quad \lim_{n\to\infty} |\, 0 \,|^{1/n} = 0$$

Since the superior limit is 1, the radius of convergence is unity.

Example. Determine the radius of convergence of the power series

$$P(z) = e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \ldots$$

In this case $|\, c_n \,| = 1/n!$. In order to find the $\lim_{n\to\infty} (1/n!)^{1/n}$ we may write

$$(n!)^2 = (1\cdot n)[2(n-1)][3(n-2)] \ldots (n\cdot 1)$$

No factor on the right-hand side is smaller than n, since

$$k(n - k + 1) - n = (k - 1)(n - k) \geq 0$$
$$k, = 1, 2, 3, \ldots, n$$

Consequently $(n!)^2 \geq n^n$ or $n! \geq n^{n/2}$

Thus $$\left|\frac{1}{n!}\right|^{1/n} \leq \left|\frac{1}{n^{n/2}}\right|^{1/n} = \frac{1}{n^{1/2}}$$

and $\lim_{n\to\infty} 1/n^{1/2} = 0$, making $l = 0$ and $r = \infty$, so that the circle includes all finite values of z.

78. Calculations with power series. Among the several power series

$$P_1(z),\ P_2(z),\ \ldots,\ P_k(z)$$

there will be one with the least radius of convergence. The circle with this radius incloses the common domain of convergence of all the series. An important theorem concerning the combination of power series states that if

$$G[P_1(z), P_2(z), \ldots, P_k(z)]$$

is an entire rational function of the power series $P_1(z), \ldots, P_k(z)$, i.e., an expression formed exclusively from the processes of addition, subtraction, and multiplication of these series, together with constants, G can be brought into the form of another power series. The resulting power series $P(z)$ will converge for all z within the common convergence circle of the series $P_1(z), \ldots, P_k(z)$ and for every such z

$$G[P_1(z), P_2(z), \ldots, P_k(z)] = P(z)$$

The reciprocal of a power series $P(z)$ having a radius of convergence different from zero and which does not vanish for $z = 0$ can be represented as a power series which will possess a radius of convergence not zero. Consequently

$$\frac{1}{P(z)} = \frac{1}{c_0 + c_1 z + c_2 z^2 + \ldots} = k_0 + k_1 z + k_2 z^2 + \ldots$$

In order to determine the constants $k_0, k_1, \ldots$, use is often made of the fact that two power series having a common domain of convergence are equal everywhere in this domain only when the coefficients of like powers of z are equal. Consequently the coefficients k_n may be determined either by formal division or by writing

$$(c_0 + c_1 z + c_2 z^2 + \ldots)(k_0 + k_1 z + k_2 z^2 + \ldots) = 1$$

and equating like powers of z on both sides. Thus

$$c_0 k_0 + (c_0 k_1 + c_1 k_0) z + \ldots = 1$$

and

$$c_0 k_0 = 1 \qquad k_0 = \frac{1}{c_0}$$

$$c_0 k_1 + c_1 k_0 = 0 \qquad k_1 = -\frac{c_1}{c_0^2} \quad \text{etc.}$$

79. Transformations of a power series. Thus far only power series having circles of convergence with centers at the origin have been considered, but the facts already given for these series apply to the power series

$$P(z - a) = c_0 + c_1(z - a) + c_2(z - a)^2 + \ldots \tag{79.1}$$

If $t = z - a$ be substituted in the above series,

$$P(t) = c_0 + c_1 t + c_2 t^2 + \ldots \tag{79.2}$$

which is of the type already considered. If the radius of convergence of the latter series is r, both series converge for values of t and z for which

$$|t| = |z - a| < r$$

This means that the series (79.1) converges for values of z within the circle of radius r with center at a.

All of the theorems concerning $P(z)$ also hold for $P(z - a)$ in its domain of convergence.

It is well to note that the point $a = \infty$ is not excluded if the symbol $(z - \infty)$ be replaced by $1/z$, so that

$$P\left(\frac{1}{z}\right) = c_0 + \frac{c_1}{z} + \frac{c_2}{z^2} + \dots$$

This series converges for all values of z for which $\left|\frac{1}{z}\right| < r$, where r is again the radius of convergence of (79.2). But the condition $\left|\frac{1}{z}\right| < r$ is equivalent to $|z| > 1/r$. This means that $P(1/z)$ converges for all values of z lying outside a circle of radius $1/r$ with center at the origin.

It is often desirable to be able to express a power series $P(z - a)$ in terms of another series $P_1(z - b)$. This may be done as follows. Suppose that the radius of convergence of $P(z - a)$ is r, and b is a point inside the convergence circle (see Fig. 79.1). Then write

$$z - a = (z - b) + (b - a)$$

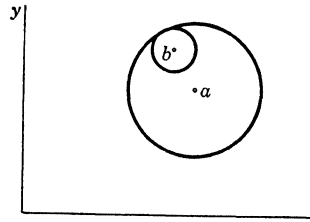


FIG. 79.1

If this be substituted in $P(z - a)$, can the terms be grouped or arranged according to ascending powers of $(z - b)$? Since $P(z - a)$ converges absolutely for $|z - a| < r$, the desired arrangement can be made if

$$|(z - b) + (b - a)| < r, \quad \text{or since}$$
$$|(z - b) + (b - a)| \leq |z - b| + |b - a|$$

the conditions are still fulfilled if

$$|z - b| + |b - a| < r \quad \text{or} \quad |z - b| < r - |b - a|.$$

Therefore the new series $P_1(z - b)$ converges for all z within a circle with center at b which is tangent to the circle of convergence of $P(z - a)$ and

$$P(z - a) = P_1(z - b)$$

The series $P_1(z - b)$ is called the transformation of the series $P(z - a)$.

80. Laurent's series. In order to represent certain functions in the form of an infinite series it is necessary to extend the power series so as to include negative powers of z. First consider the series of complex numbers

$$\sum_{n=-\infty}^{+\infty} \alpha_n \tag{80.1}$$

By this series is meant the sum of the two series

$$\left.\begin{aligned} \sigma_1 &= \alpha_0 + \alpha_1 + \alpha_2 + \dots \\ \sigma_2 &= \alpha_{-1} + \alpha_{-2} + \alpha_{-3} + \dots \end{aligned}\right\} \tag{80.2}$$

The series (80.1) converges only when both series (80.2) converge simultaneously, in which case the series (80.1) converges to the sum $\sigma_1 + \sigma_2$.

In the light of the above definition, the series

$$L(z) = \sum_{n=-\infty}^{\infty} c_n z^n \tag{80.3}$$

converges only when both series

$$P(z) = c_0 + c_1 z + c_2 z^2 + \dots$$

$$P_1\left(\frac{1}{z}\right) = \frac{c_{-1}}{z} + \frac{c_{-2}}{z^2} + \frac{c_{-3}}{z^3} + \dots$$

converge. From the previous section, the series $P_1(1/z)$ will converge in a domain outside a circle of radius r_1 with center at the origin. On the other hand, the domain of convergence of $P(z)$ will be on the inside of a circle with radius r. Consequently if $r > r_1$ the two series will have a common domain of convergence, and the Laurent series (80.3) will converge for values of z in the ring-shaped region shown in Fig. 80.1. The value of $L(z) = P(z) + P_1(1/z)$. The method for determining the coefficients in the Laurent expansion of a function is given in Sec. 91.

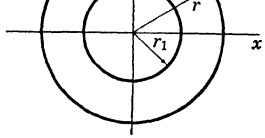


FIG. 80.1

The Laurent series has the following important property. If a circle of radius ρ is described such that it lies within the ring of convergence of $L(z)$, i.e., if $r_1 < \rho < r$, then $|L(z)| \leq M$ for all values of z on this circle, where M is some positive number. This means that $L(z)$ has a finite maximum as z traverses the circle, and furthermore

$$|c_n|\rho^n \leq M \qquad \text{for all } n$$

81. Derivatives of complex functions. The derivative of a function $f(z)$ of the complex variable z is defined in much the same way as the derivative of a function of a real variable, namely, the derivative of $f(z)$ at the point $z = z_0$ is defined to be

$$\lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h}$$

when this limit exists. This differs from the definition of the derivative of a function of a real variable in the respect that the point z_0 may be approached in any number of ways, because h is a complex number. Thus h may be regarded as a vector beginning at $z = z_0$, whose end point describes a curve as $h \to 0$; see Fig. 81.1. The shape of the curve depends upon the manner in which the real and imaginary parts of h

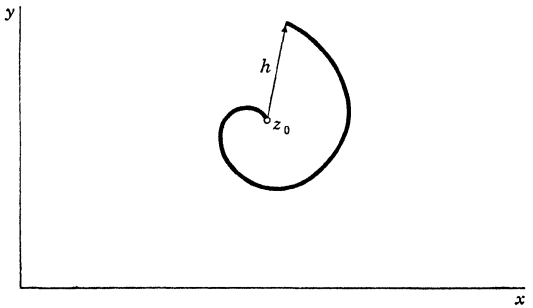


FIG. 81.1

approach zero. It may happen that the value of the above limit depends on the mode of approach to z_0, and in fact this is the case even for some of the elementary functions at certain values of z_0. In general there will be certain points of the complex plane where the limit either does not exist or depends on the manner in which $h \to 0$. As examples of these let $f(z) = 1/(z - a)$. What is the derivative at $z_0 = a$? Forming the above limit

$$\lim_{h\to 0} \frac{1}{h}\left(\frac{1}{z - a + h} - \frac{1}{z - a}\right)$$

it is readily seen that the limit fails to exist for $z_0 = a$, no matter what path the end point of the vector h follows as $h \to 0$.

Again let us examine the derivative of $f(z) = |z|$. If the point z_0 is approached along radial lines,

$$\lim_{h\to 0} \frac{|z_0 + h| - |z_0|}{h}$$

may be written in the convenient form

$$\lim_{\delta\to 0} \frac{|\rho_0 e^{i\theta} + \delta e^{i\varphi}| - \rho_0}{\delta e^{i\varphi}}$$

where the particular radial line of approach is specified by the angle φ.

The above expression may also be written

$$\lim_{\delta \to 0} \frac{\sqrt{\rho_0^2 + 2\rho_0\delta \cos(\theta - \varphi) + \delta^2} - \rho_0}{\delta e^{i\varphi}} \quad \text{or}$$

$$\lim_{\delta \to 0} \frac{\rho_0\left(1 + \dfrac{\delta}{\rho_0}\cos(\theta - \varphi) + \text{higher powers in } \delta\right) - \rho_0}{\delta e^{i\varphi}}$$

$$= e^{-i\varphi}\cos(\theta - \varphi)$$

The limit depends on the angle φ, so that the limit is different for every radial approach for any value of z_0 in the complex plane except $z_0 = 0$.

Nevertheless there exists a most important class of functions for which, in a certain domain D of the complex plane, the above limit exists and is *independent* of the manner in which the point z_0 is approached. Such functions are said to be *analytic* over the domain D. Points at which $f(z)$ is not analytic are called singular points or merely singularities.

82. Cauchy-Riemann differential equations. For practical use, it is desirable to have a method for determining whether a function is analytic or not, which is easier to apply than that given above. If the function $w = f(z)$ is analytic and is written in the form

$$w = u + iv = u(x,y) + iv(x,y), \quad \text{then}$$

$$\frac{dw}{dz} = \lim_{h \to 0} \frac{f(z_0+h) - f(z_0)}{h}$$

$$= \lim_{a,b \to 0,0} \frac{u(x_0+a,y_0+b) + iv(x_0+a,y_0+b) - u(x_0,y_0) - iv(x_0,y_0)}{a + ib}$$

where $h = a + ib$, and the limit is independent of the manner in which $h \to 0$. Let us first find the limit as z_0 is approached along the imaginary axis, so that $a = 0$ and $b \to 0$. Then

$$\frac{dw}{dz} = \lim_{b \to 0} \frac{1}{i}\frac{u(x_0,y_0+b) - u(x_0,y_0)}{b} + \frac{v(x_0,y_0+b) - v(x_0,y_0)}{b}$$

$$= \frac{1}{i}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} = \frac{1}{i}\frac{\partial w}{\partial y} \tag{82.1}$$

If on the other hand, z_0 is approached along the real axis, then $b = 0$ and $a \to 0$, so that

$$\frac{dw}{dz} = \lim_{a \to 0} \frac{u(x_0+a,y_0) - u(x_0,y_0)}{a} + i\frac{v(x_0+a,y_0) - v(x_0,y_0)}{a}$$

$$= \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial w}{\partial x} \tag{82.2}$$

Since the function has been assumed analytic, the two values of the

derivative must be the same, so that

$$\frac{1}{i}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}$$

Equating real and imaginary terms, this requires that the two differential equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{82.3}$$

must be satisfied simultaneously. These equations are called the Cauchy-Riemann differential equations.

It may also be shown that if in a certain domain D the functions u and v possess continuous partial derivatives of the first order and the Cauchy-Riemann differential equations are satisfied throughout D, then $f(z) = u + iv$ is analytic in D.

If in addition, the functions u and v possess continuous partial derivatives of the second order, eqs. (82.3) may be differentiated with respect to x and y, the result being

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x\,\partial y}, \qquad \frac{\partial^2 u}{\partial y^2} = -\frac{\partial^2 v}{\partial y\,\partial x}$$

Therefore
$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = \Delta u = 0$$

Likewise
$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = \Delta v = 0 \tag{82.4}$$

Consequently, under the above conditions, it is apparent that the real and imaginary parts of any analytic function satisfy one and the same partial differential equation. Since this equation is identical with that which must be satisfied by the electric potential in a charge-free two-dimensional space, and is also satisfied by the velocity potential in the steady-state flow of an ideal fluid, the use of analytic functions of a complex variable in practical problems is enormously important. Certain instances of their use will be given in subsequent sections.

In connection with the above remarks, it must not be concluded that if u and v are *any* two solutions of Laplace's equation, the function $u + iv$ will be analytic, because they will not in general satisfy the Cauchy-Riemann equations. However, if one solution u of (82.4) is known, the proper solution v can be found so that $w = u + iv$ will be analytic. To find the proper v, it is supposed that v possesses first-order partial derivatives, so that

$$dv = \frac{\partial v}{\partial x}\,dx + \frac{\partial v}{\partial y}\,dy$$

and
$$v = \int \frac{\partial v}{\partial x}\,dx + \frac{\partial v}{\partial y}\,dy + \text{constant}$$

Making use of the Cauchy-Riemann equations the above expression becomes

$$v = \int -\frac{\partial u}{\partial y}\,dx + \frac{\partial u}{\partial x}\,dy + \text{constant}$$

The additive constant is quite arbitrary. This function v is defined in terms of u and its derivatives, and $w = u + iv$ will be analytic.

83. Complex integration. Integration of a function of a complex variable may be defined in terms of line integrals of functions of two real variables. In Sec. 54 the line integral of a function $P(x,y)$ along a curve C was written as $\int_C P(x,y)dx$ and defined to be

$$\int_{t_1}^{t_2} P[\varphi,(t),\psi(t)]\varphi'(t)dt$$

where $x = \varphi(t)$, $y = \psi(t)$, $t_1 \leqslant t \leqslant t_2$ are the parametric equations of the curve C. Similarly the integral

$$\int_C Q(x,y)dy = \int_{t_1}^{t_2} Q[\varphi(t),\psi(t)]\psi'(t)dt$$

If $w = f(z) = u(x,y) + iv(x,y)$ is a function of the complex variable $z = x + iy$, the integral of $f(z)$ along the curve C in the complex plane is defined to be

$$\begin{aligned}\int_C f(z)dz &= \int_{t_1}^{t_2} \{u[\varphi(t),\psi(t)] + iv[\varphi(t),\psi(t)]\}[\varphi'(t) + i\psi'(t)]dt \\ &= \int_{t_1}^{t_2} u\varphi'\,dt - \int_{t_1}^{t_2} v\psi'\,dt + i\int_{t_1}^{t_2} v\varphi'\,dt + i\int_{t_1}^{t_2} u\psi'\,dt \\ &= \int_C u\,dx - v\,dy + i\int_C v\,dx + u\,dy \qquad (83.1)\end{aligned}$$

The existence of integrals of a function of a complex variable therefore depends on the existence of line integrals of real functions (see Sec. 54).

An upper bound to the value of an integral of a complex function can readily be established, since

$$\left|\int_C f(z)dz\right| \leq \int_C \left|f(z)dz\right| \leq \int_{t_1}^{t_2} \sqrt{u^2 + v^2}\sqrt{\varphi'^2 + \psi'^2}\,dt$$

Thus if M be such that in the domain of integration $|f(z)| < M$, then

$$\left|\int_C f(z)dz\right| \leq M\int_{t_1}^{t_2} \sqrt{\varphi'^2 + \psi'^2}\,dt = Ml \qquad (83.2)$$

where l is the length of the curve C.

84. Integration and differentiation of infinite series. Some care has to be exercised in integrating or differentiating an infinite series of functions term by term. It may happen that the series obtained by term

by term differentiation of an infinite series does not yield the derivative of the function represented by the original series.

Given the infinite series

$$w = w_1(z) + w_2(z) + \dots \tag{84.1}$$

it may be *integrated* term by term, and the resulting series will converge to $\int_C w\,dz$ *provided* the individual terms of the above series are continuous functions of z and the series converges uniformly within a certain domain.

With regard to differentiating (84.1), the derivative dw/dz may be found by differentiating the series term by term, *provided* the individual terms are continuous functions and the series (84.1) converges uniformly, and in addition the series

$$\frac{dw_1}{dz} + \frac{dw_2}{dz} + \frac{dw_3}{dz} + \dots$$

converges within a certain domain.

In case all the functions $w_1, w_2, \dots$, etc. are *analytic* functions of z in a domain D, and the series

$$w = w_1 + w_2 + w_3 + \dots \tag{84.2}$$

converges uniformly in any domain d lying inside D, the sum w is an analytic function of z, and the series

$$\frac{dw}{dz} = \frac{dw_1}{dz} + \frac{dw_2}{dz} + \dots$$

converges uniformly in d. Then since the last series satisfies all the conditions of the first series, the process of differentiation may be repeated, so that

$$w^{(k)} = w_1^{(k)} + w_2^{(k)} + w_3^{(k)} + \dots$$

A power series furnishes an example of a series satisfying the conditions of (84.2).

85. Cauchy's integral theorem. This theorem has to do with the integral of a complex function over a closed path. A simple closed curve C in the z plane is usually called a contour. If $f(z)$ is a one-valued continuous but not necessarily analytic function for all z on the curve C, the integral $\int_C f(z)dz$ [sometimes written $\oint_C f(z)dz$] taken around the contour in a counterclockwise direction is referred to as a contour integral.

In case $f(z)$ is analytic at all points *on* and *within* the contour C, then $\int_C f(z)dz = 0$. Thus if $f(z)$ is analytic in a domain D, the integral of

$f(z)$ taken over *any* contour lying completely within D is zero. This is known as Cauchy's integral theorem.

The theorem may be proved very easily by making use of the relation (83.1), namely,

$$\oint f(z)dz = \oint (u\,dx - v\,dy) + i\oint (v\,dx + u\,dy) \qquad (85.1)$$

The line integrals on the right may be expressed in terms of surface integrals over the surface bounded by the closed contour by making use of Green's theorem (59.3).

$$\oint (u\,dx - v\,dy) = \iint_S \left(-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)dx\,dy$$

$$\oint (v\,dx + u\,dy) = \iint_S \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)dx\,dy$$

so that eq. (85.1) becomes

$$\oint f(z)dz = -\iint_S \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)dx\,dy + i\iint_S \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)dx\,dy$$

Since $f(z)$ is analytic on and within C, the functions u and v satisfy the Cauchy-Riemann differential equations (82.3). Making use of these, the above expression becomes

$$\oint f(z)dz = -\iint_S \left(-\frac{\partial u}{\partial y} + \frac{\partial u}{\partial y}\right)dx\,dy + i\iint_S \left(\frac{\partial u}{\partial x} - \frac{\partial u}{\partial x}\right)dx\,dy = 0$$

The converse of this theorem also holds, namely, if $f(z)$ is continuous throughout a simply connected region D of the complex plane and $\oint f(z)dz = 0$ for all contours within D, then $f(z)$ is analytic in the region D.

By a simply connected region D, is meant one in which any closed curve may be drawn and then shrunk down to a point without any portion of the curve extending outside of D. An example of a region not simply connected is the region between two closed curves, as shown in Fig. 85.1a. The dotted curve is a closed curve which cannot be shrunk to a point and still have it remain in D. Such a region can be converted into a simply connected region by connecting the inner and outer boundaries of D by any curve. If only one such connecting curve is required, the region is said to be doubly connected, etc.

Cauchy's integral theorem can also be used in the case of a multiply connected region when suitably modified. For instance, take the doubly connected region D shown in Fig. 85.1b, where the inner and outer

boundaries C_1 and C_2 have been connected by the two dotted lines. These divide the region D into two simply connected regions within which the function $f(z)$ is analytic. Consequently applying Cauchy's theorem to both regions

$$\oint_{abfcd} f(z)dz = 0 \quad \text{and} \quad \oint_{adcgb} f(z)dz = 0$$

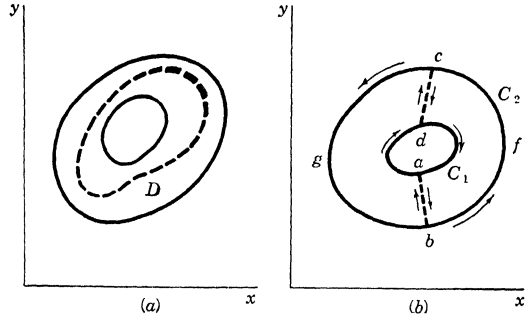


FIG. 85.1

In these two contours the paths ab and cd are common and traversed in opposite directions, so that if the two contour integrals are added, the result is the same as

$$\int_{C_2} f(z)dz - \int_{C_1} f(z)dz = 0 \tag{85.2}$$

This means that in a multiply connected region the integral must be taken over all the boundaries. The minus sign arises because the curve C_1 is traversed in the clockwise direction.

86. Cauchy's integral formula. This formula allows one to express the value of an analytic function at a point in terms of the values of $f(z)$ lying on a contour surrounding the point. More precisely, let $f(z)$ be an analytic function everywhere in a domain D, and let C be a contour entirely within D, and let $z = a$ be an arbitrary point inside the contour C. If $f(z)$ is divided by $z - a$, the resulting function remains analytic at all points on and within C except at $z = a$. But if a is surrounded by a circle γ of radius δ (see Fig. 86.1), then $f(z)/(z - a)$ will be analytic in the region bounded by γ and C, and Cauchy's integral theorem will hold in this region, so that

$$\int_C \frac{f(z)}{z - a} dz - \int_\gamma \frac{f(z)}{z - a} dz = 0 \tag{86.1}$$

The second integral may be evaluated by writing $z - a = \delta e^{i\theta}$, so that

$$\int_\gamma \frac{f(z)}{z-a}\,dz = \int_0^{2\pi} \frac{f(a+\delta e^{i\theta})}{\delta e^{i\theta}}\,\delta i e^{i\theta}\,d\theta$$

$$= i\int_0^{2\pi} f(a+\delta e^{i\theta})d\theta$$

Since this integral is a continuous function of δ, we seek to evaluate the integral for a circle of vanishing radius; thus

$$\lim_{\delta\to 0} i\int_0^{2\pi} f(a+\delta e^{i\theta})d\theta = 2\pi i f(a)$$

Then from (86.1),

$$f(a) = \frac{1}{2\pi i}\oint_C \frac{f(z)dz}{z-a} \tag{86.2}$$

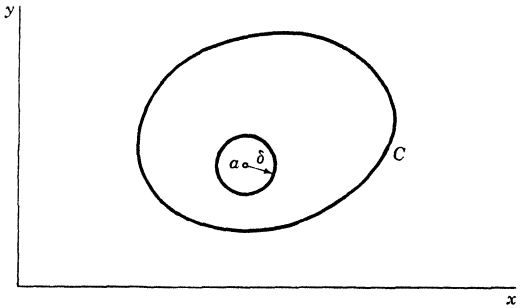


FIG. 86.1

As mentioned at the beginning of this section, this formula expresses the value of $f(z)$ at a point $z = a$ in terms of values of $f(z)$ on a contour C surrounding a.

If the original region is doubly connected and a lies in the region where $f(z)$ is analytic, it is easy to show that

$$f(a) = \frac{1}{2\pi i}\int_{C_2} \frac{f(z)}{z-a}\,dz - \frac{1}{2\pi i}\int_{C_1} \frac{f(z)}{z-a}\,dz \tag{86.3}$$

87. The derivative of an analytic function defined by a Cauchy integral. If $f(z)$ is an analytic function of z in the domain D, the derivative of $f(z)$ at the point $z = a$ exists if a is within D and is defined as

$$f'(a) = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}$$

For values of h for which $a + h$ lies within D, the above quotient may be expressed as a Cauchy integral.

$$\frac{f(a+h) - f(a)}{h} = \frac{1}{2\pi i h}\left[\int_C \frac{f(z)dz}{z-a-h} - \int_C \frac{f(z)dz}{z-a}\right]$$

where C is a contour containing the points a and $a + h$ and lying within D. Making use of the above equation, the derivative becomes

$$\begin{aligned} f'(a) &= \lim_{h\to 0} \frac{1}{2\pi i h} \int_C \frac{hf(z)dz}{(z-a-h)(z-a)} \\ &= \lim_{h\to 0} \frac{1}{2\pi i} \int_C \frac{f(z)dz}{(z-a)^2[1-h/(z-a)]} \\ &= \lim_{h\to 0} \frac{1}{2\pi i} \int_C \frac{f(z)dz}{(z-a)^2} \left[1 + \frac{h/(z-a)}{1-h/(z-a)}\right] \\ &= \lim_{h\to 0} \frac{1}{2\pi i} \int_C \frac{f(z)dz}{(z-a)^2} + \frac{h}{2\pi i} \int_C \frac{f(z)dz}{(z-a)^2(z-a-h)} \end{aligned}$$

Since $|h|$ can always be taken less than $\frac{1}{2}|z-a|$, the integrand of the second integral is bounded for all values of z on C, and in the limit $h \to 0$ the second term vanishes, so that

$$f'(a) = \frac{1}{2\pi i} \int_C \frac{f(z)}{(z-a)^2} dz \tag{87.1}$$

This result is the same as that obtained by differentiating the expression

$$f(a) = \frac{1}{2\pi i} \int_C \frac{f(z)}{(z-a)} dz$$

with respect to a. The same procedure could be continued, showing that the nth derivative of $f(a)$ exists and is expressible by the Cauchy integral

$$f^{(n)}(a) = \frac{n!}{2\pi i} \int_C \frac{f(z)}{(z-a)^{n+1}} dz$$

Somewhat greater symmetry is obtained in the above expression if $n!$ is replaced by the gamma function $\Gamma(n+1)$, so that

$$f^{(n)}(a) = \frac{\Gamma(n+1)}{2\pi i} \int_C \frac{f(z)}{(z-a)^{n+1}} dz \tag{87.2}$$

With this formulation, the function on the right is well defined even when n is not an integer, and may be taken as the definition of fractional differentiation.

88. Taylor's theorem. Let $f(z)$ be an analytic function in a domain including the point $z = a$ and let C be a circle with center at a such that $f(z)$ has no singularity on or within it. In other words, C lies within D. If $a + h$ be a point within C, then

$$\begin{aligned} f(a+h) &= \frac{1}{2\pi i} \oint_C \frac{f(z)dz}{(z-a-h)} \\ &= \frac{1}{2\pi i} \oint \frac{f(z)dz}{(z-a)} \frac{1}{1-h/(z-a)} \end{aligned}$$

The power series $1 - h/(z-a)$ satisfies the conditions in Sec. 78, admitting a power series representation of its reciprocal.

$$\frac{1}{1-h/(z-a)} = 1 + \frac{h}{z-a} + \left(\frac{h}{z-a}\right)^2 + \cdots$$

Since this series converges uniformly when $\left|\frac{h}{z-a}\right| < 1 - \epsilon$, where $\epsilon > 0$, the series may be safely integrated term by term so that

$$\begin{aligned} f(a+h) &= \frac{1}{2\pi i}\oint_C \frac{f(z)dz}{z-a} + \frac{h}{2\pi i}\oint_C \frac{f(z)dz}{(z-a)^2} + \frac{h^2}{2\pi i}\oint_C \frac{f(z)dz}{(z-a)^3} + \cdots \\ &= f(a) + f'(a)h + \frac{f''(a)h^2}{2!} + \cdots \\ &\quad + \frac{f^{(n)}(a)}{n!}h^n + \frac{h^{n+1}}{2\pi i}\oint_C \frac{f(z)dz}{(z-a)^{n+1}(z-a-h)} + \cdots \end{aligned}$$

Making use of the inequality (83.2), a bound may be imposed on the last integral.

$$\left|\frac{h^{n+1}}{2\pi i}\oint \frac{f(z)dz}{(z-a)^{n+1}(z-a-h)}\right| \leq \frac{M2\pi R|\ h\ |^{n+1}}{2\pi R^{n+1}}$$

where R is the radius of the circle C and $|\ f(z)\ | \leq M$ on C. Since this tends to zero as $n \to \infty$, the analytic function $f(z)$ may be expressed as a power series.

$$f(a+h) = f(a) + f'(a)h + \frac{f''(a)h^2}{2!} + \cdots + \frac{f^{(n)}(a)h^n}{n!} + \cdots$$

or writing $h = z - a$, then

$$f(z) = f(a) + f'(a)(z-a) + \frac{f''}{2}(z-a)^2 + \cdots$$

This important result is known as Taylor's theorem. In the course of its derivation it has been shown that any function $f(z)$ which is analytic in the neighborhood of a point a can be represented by the power series $P(z-a)$. It also follows that the radius of convergence of $P(z-a)$ is at least as great as that of a circle with center at a, which just excludes the nearest singular point of $f(z)$. This provides another powerful method of determining the radius of convergence of a power series expansion of a function $f(z)$.

Two examples of the method will be given. First consider the function $1/(1+x^2)$. Suppose that we desire its power series expansion in the neighborhood of the point $x = 0$, and wish to know its radius of convergence. Since x is real, this function has no singularity, but if instead the function $1/(1+z^2)$ of the complex variable z is considered, then since

$$\frac{1}{1+z^2} = \frac{1}{(z+i)(z-i)}$$

the function is seen to have two singularities, one at $z = i$ and the other at $z = -i$. The circle with center at $z = 0$ passing through the nearest singularity has a radius unity, so that the power series representation of $1/(1 + z^2)$ has a radius of convergence equal to unity, and consequently the original function converges for $| x | < 1$.

As a second example consider the expansion of $(1 - 2xh + h^2)^{-1/2}$ as a power series in h

$$(1 - 2xh + h^2)^{-1/2} = 1 + P_1(x)h + P_2(x)h^2 + P_3(x)h^3 + \ldots$$

Here x and h will be considered real, and we wish to know how the radius of convergence depends on x. Again replacing h by the complex variable z, we look for the singularities of $(1 - 2xz + z^2)^{-1/2}$ which amounts to locating the zero points of the denominator, and consequently the values of z for which $z^2 - 2xz + 1 = 0$. These values are

$$z = \frac{2x \pm \sqrt{4x^2 - 4}}{2} = x \pm \sqrt{x^2 - 1}$$

For $x^2 > 1$ the singular points lie on the real axis, and the nearest one to $z = 0$ is $x - \sqrt{x^2 - 1}$, so that the radius of convergence is $x - \sqrt{x^2 - 1}$. Consequently $| h | < x - \sqrt{x^2 - 1}$ guarantees convergence. When $x^2 < 1$, the singular points are complex, being $x \pm i\sqrt{1 - x^2}$. These singularities lie at equal distances from the origin, and the radius of the circle passing through them is

$$\sqrt{x^2 + (1 - x^2)} = 1$$

Consequently, when $x^2 < 1$ the series under consideration converges for all values of h for which $| h | < 1$.

89. Continuation of a power series. The method for determining the radius of convergence of the power series representation of a function given in the preceding section also indicates a method for extending the domain of representation. Let $f(z)$ be analytic in the neighborhood of a point $z = a$. Then in a region near a, the function $f(z)$ can be represented by the Taylor's series

$$f(z) = f(a) + f'(a)(z-a) + \ldots = P(z-a)$$

If s_1 is the nearest singularity to $z = a$ as shown in Fig. 89.1, the above series certainly converges within a circle C_1 with center at a and passing through $z = s_1$. However, $f(z)$ can also be expanded in a Taylor's series about another point $z = b$, where b lies within C_1. Then $f(z) = P(z-b)$ in some neighborhood about b, but it is already known that $P(z-b)$ converges inside a circle C_2 with center b and passing through the nearest singularity to b. With the singularities as shown in Fig. 89.1 this would still be s_1. This process of extending the domain of conver-

gence by transforming to new power series is called continuation. This allows one to start from a power series representation of a function and find therefrom any number of power series representing the same function over different domains. The aggregate of all such power series constitutes the analytical expression of the function.

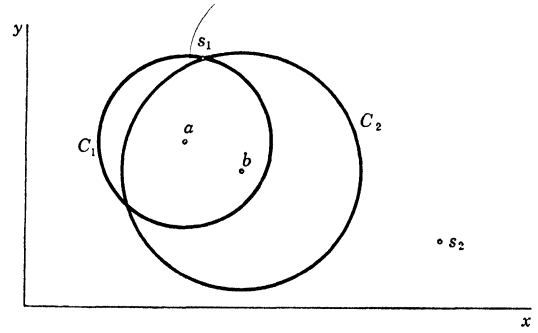


FIG. 89.1

90. Nature of singularities. Consider a function $f(z)$ which is analytic throughout a domain D except at a single point $z = a$ therein. It will be supposed that $f(z)$ can be expanded in a Laurent series.

$$f(z) = P_1(z-a) + P_2\left(\frac{1}{z-a}\right)$$

Since $P_1(z-a)$ is a Taylor's series, it is an analytic function of z everywhere in the domain of convergence. On the other hand,

$$P_2\left(\frac{1}{z-a}\right) = \frac{c_{-1}}{z-a} + \frac{c_{-2}}{(z-a)^2} + \cdots + \frac{c_{-n}}{(z-a)^n} + \cdots$$

If it should happen that all the coefficients c_{-n-k} vanish for all $k > 0$, then $f(z)$ is said to have a pole of order n at $z = a$. The part of $f(z)$ represented by $P_2[1/(z-a)]$ is called the principal part of $f(z)$ near $z = a$. If all the coefficients c_{-n} are zero except c_{-1}, then $f(z)$ is said to have a simple pole at $z = a$. If, on the other hand, none of the coefficients vanishes, i.e., $P_2[1/(z-a)]$ is an infinite series then $f(z)$ is said to have an essential singularity at the point $z = a$.

Most functions are not analytic at every point of the complex plane but usually have singular points somewhere, often at the point $z = \infty$. In fact, the only function which remains analytic at all points of the complex plane including $z = \infty$ is $f(z) =$ constant. Furthermore, the only single-valued functions which have no singularities except poles at any point including $z = \infty$ are rational functions.

There is an important class of functions which are analytic at all points except $z = \infty$. These are called integral or entire functions. With the exception of polynomials, these functions all have an essential singularity at $z = \infty$. An example of an entire function is the exponential e^z.

91. Determination of coefficients in a Laurent series. Let $f(z)$ be an analytic function within the domain D except at a single point $z = a$ where it has an isolated singularity. If two concentric circles C_1 and C_2 are drawn with centers at a, then $f(z)$ will be analytic on C_1 and C_2 and in the ring-shaped region between C_1 and C_2. This allows one to express the value of $f(z)$ at some point, say $z = a + h$, between the circles as shown in Fig. 91.1, by the Cauchy integral [see eq. (86.3)].

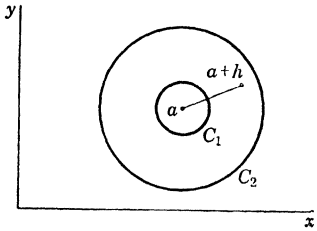


FIG. 91.1

$$f(a+h) = \frac{1}{2\pi i}\oint_{C_2}\frac{f(z)dz}{(z-a-h)} - \frac{1}{2\pi i}\oint_{C_1}\frac{f(z)dz}{(z-a-h)}$$

For z on C_2

$$\frac{1}{z-a-h} = \frac{1}{(z-a)[1-h/(z-a)]}$$
$$= \frac{1}{(z-a)}\left[1 + \frac{h}{z-a} + \left(\frac{h}{z-a}\right)^2 + \ldots\right]$$

This series converges uniformly provided $\left|\frac{h}{z-a}\right| < 1$. This condition is always fulfilled for values of z on C_2. Consequently, this series may be substituted in the first of the above integrals and integrated term by term.

On the other hand, when z is on C_1,

$$\frac{1}{z-a-h} = -\frac{1}{h[1-(z-a)/h]}$$
$$= -\frac{1}{h}\left[1 + \frac{z-a}{h} + \left(\frac{z-a}{h}\right)^2 + \ldots\right]$$

the series converges uniformly for $\left|\frac{z-a}{h}\right| < 1$. This condition is always fulfilled for values of z on C_1, so that this series may be substituted in the second integral and integrated term by term. The result is

$$f(a+h) = c_0 + c_1h + c_2h^2 + \ldots + \frac{c_{-1}}{h} + \frac{c_{-2}}{h^2} + \ldots$$

where
$$c_n = \frac{1}{2\pi i}\int_{C_2}\frac{f(z)dz}{(z-a)^{n+1}} \tag{91.1}$$
$$c_{-n} = \frac{1}{2\pi i}\int_{C_1}\frac{f(z)dz}{(z-a)^{-n+1}} \tag{91.2}$$

Examples of the practical use of these formulas are given in Sec. 103.

92. Many-valued functions. In addition to the singularities of a function $f(z)$ discussed in Sec. 90, still another important type presents itself in connection with many-valued functions. The nature of this singularity as well as the characteristics of some common many-valued functions can be clearly seen by considering a few such functions.

The Function $w = z^{1/2}$. If z is represented in the z plane by a point whose polar coordinates are ρ and φ, then

$$w = \rho^{1/2}e^{i\varphi/2} = \rho^{1/2}(\cos\varphi/2 + i\sin\varphi/2)$$

Written in this form it is easy to follow the behavior of w as z is allowed to traverse a circle of radius ρ with center at $z = 0$. It is to be noted that, starting with $\varphi = 0$, $w = \rho^{1/2}$, and after φ has steadily increased to 2π thereby returning z to the initial point, $w = -\rho^{1/2}$, so that for this value of z, the function w has two values. In fact, if z is allowed to traverse the circle a second time $2\pi \leq \varphi \leq 4\pi$, the function w will assume a different set of values than it did for the first circuit, and return to the value $\rho^{1/2}$ at the end of the second circuit. Consequently it is seen that two values of w correspond to one point in the z plane.

For definiteness let w_1 denote those values assumed by w when z lies in the domain $0 < \rho < \infty$; $0 \leq \varphi < 2\pi$, and w_2 the values taken on when z lies in the region $0 < \rho < \infty$; $2\pi \leq \varphi < 4\pi$. For the domains in which they are defined w_1 and w_2 are single-valued analytic functions of z. The two single-valued functions

$$w_1 = \rho^{1/2}(\cos\varphi/2 + i\sin\varphi/2)$$

and
$$w_2 = \rho^{1/2}[\cos 1/2(\varphi + 2\pi) + i\sin 1/2(\varphi + 2\pi)]$$

for $0 < \varphi \leq 2\pi$ are called branches of the function w. The two branches coincide at $z = 0$, there being only one value of w at this point. Such a point is called a branch point of the function, and constitutes the singularity referred to above. The derivative $dw/dz = 1/z^{1/2}$ does not exist at $z = 0$, and the function fails to be analytic at the branch point.

In order to distinguish which values of z belong to a given branch, Riemann chose to associate separate z planes with each of the two branches w_1 and w_2. He imagined these two planes located one above the other with the branch points coincident, the upper plane carrying the values of z for w_1, and the lower plane carrying the values of z for w_2. Since after one circuit about $z = 0$ in the upper plane we have trans-

ferred to the branch w_2, we must arrange to have the value of z be on the lower plane. This is accomplished by imagining the two planes cut and joined together along any line from $|z| = 0$ to $|z| = \infty$. After circling $z = 0$ in the lower plane, it again passes through the cut and up into the upper plane. This surface which carries the complex variable z is called a Riemann surface.

The Logarithm. The logarithm of a complex variable z is defined by means of the equation

$$e^w = z \tag{92.1}$$

The function $w = u + iv$ is written $w = \log z$. The real and imaginary parts u and v may easily be found when z is expressed in polar form, whence

$$e^{u+iv} = \rho e^{i\varphi}$$

or

$$e^u(\cos v + i \sin v) = \rho(\cos \varphi + i \sin \varphi)$$

When the real and imaginary parts are equated, the two resulting equations involving only real quantities

$$e^u \cos v = \rho \cos \varphi, \qquad e^u \sin v = \rho \sin \varphi$$

must be solved for u and v. The solution gives

$$u = \log \rho \quad \text{and} \quad v = \varphi$$

Thus the logarithm of a complex function is

$$w = \log z = \log |z| + i\varphi \tag{92.2}$$

It must be noticed, however, that eq. (92.1) does not serve to define w uniquely, because $e^{w \pm 2k\pi i}$ is indistinguishable from e^w provided k is an integer. The general solution of (92.1) is therefore

$$w = \log z = \log |z| + i\varphi + 2k\pi i$$

where $k = 0, \pm 1, \pm 2, \pm \ldots$ (92.3)

This function furnishes an example of an infinitely many-valued function, and the Riemann surface would be composed of an infinite number of sheets with the branch point $z = 0$ in common.

The value of the logarithm when $k = 0$, i.e., that given by eq. (92.2) is called the principal value of the logarithm. The principal value is an analytic function of z in the domain $0 < \rho < \infty$, $-\pi < \varphi \leq \pi$, and is often written $\ln z$.

The General Power $w = z^m$. If m is a positive integer, z^m is the value found by multiplying z by itself m times. When m is an arbitrary complex number, however, the value of w is defined by means of the logarithm to be

$$w = z^m = e^{m \log z} = e^{m \log |z| + mi\varphi + 2km\pi i} \tag{92.4}$$

It is seen at once that in general z^m is an infinitely many-valued function since $k = 0, \pm 1, \pm 2, \pm \ldots$.

Under certain conditions, however, there is only a finite number of distinct values of z^m, the remainder being merely repetitions of these values. This means that there must be certain values of k for which z^m has the same value. Two such values of k must satisfy the relation

$$e^{m2\pi k_1 i} = e^{m2\pi k_2 i} \quad \text{or} \quad e^{m2\pi i(k_1 - k_2)} = 1; \quad k_1 \neq k_2$$

The latter relation is true only if $m(k_1 - k_2)$ is an integer. Since k_1 and k_2 are integers, the difference is always an integer, and m must be an integer or a rational number $m = p/q$, where p and q are integers. If m is an integer, (92.4) shows that z^m is single-valued, but if $m = p/q$, then

$$z^m = e^{(p/q)\ln z + (p/q)2k\pi i}$$

Starting with $k = 0$, this yields different values until $k = q$, whereupon it assumes the value for $k = 0$. For still greater values of k, the same set of values of z^m are repeated, so that it suffices to take $k = 0, 1, 2, \ldots, q - 1$. Consequently, when $m = p/q$, the function z^m is q-valued.

The principal value of z^m is considered to be that value which corresponds to the principal value of the logarithm; hence the principal value of

$$z^m = e^{m \ln z} = e^{m \log |z| + im\varphi} \tag{92.5}$$

Since $\ln z$ is a single-valued analytic function of z in the domain $-\pi < \arg z \leq \pi$, so also is $w = z^m$.

SUGGESTED REFERENCES

A. Hurwitz and R. Courant, *Allgemeine Funktionentheorie und elliptische Funktionen*, Berlin, Julius Springer, 1925.

W. F. Osgood, *Lehrbuch der Funktionentheorie*, Vol. I, Berlin, B. G. Teubner, 1912.

A. R. Forsyth, *Theory of Functions of a Complex Variable*, London, Cambridge University Press, 1893.

T. M. MacRobert, *Functions of a Complex Variable*, London, Macmillan & Co., Ltd., 1917.

E. T. Copson, *Theory of Functions of a Complex Variable*, New York, Oxford University Press, 1935.

PROBLEMS

73.1. For the series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots$ show that the partial sums are always less than 1 and greater than $\frac{1}{2}$. Then prove that the series converges.

73.2. Rearrange the series of the preceding problem by taking the first and third terms, then the second, then the fifth and seventh, then the fourth, etc., so that the series $1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \ldots$ results. Show that the sum of this series is greater than the sum of the preceding series.

73.3. Test the convergence of the geometric series $1 + z + z^2 + \ldots$. *Hint.* Show that the nth partial sum is equal to $(1 - z^n)/(1 - z)$.

76.1. Test for uniform convergence

$$I = \frac{x}{x+1} + \frac{x}{(x+1)(2x+1)} + \frac{x}{(2x+1)(3x+1)} + \ldots.$$

76.2. Show that

$$z + \frac{a-b}{2!}z^2 + \frac{(a-b)(a-2b)}{3!}z^3 + \frac{(a-b)(a-2b)(a-3b)}{4!}z^4 + \ldots$$

converges absolutely when $|z| < \frac{1}{|b|}$.

76.3. Find radius of convergence of the hypergeometric series

$$F(\alpha,\beta,\gamma;z) = 1 + \frac{\alpha\beta}{1!\gamma}z + \frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\gamma(\gamma+1)}z^2 + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+2)}{3!\gamma(\gamma+1)(\gamma+2)}z^3 + \ldots$$

for all α, β, γ, provided, of course, that γ is not a negative integer.

76.4. Under certain conditions

$$\frac{1}{2\sinh x} = \sum_{n=0}^{\infty} e^{-x(2n+1)}.$$

For what values of x does this series converge?

76.5. For what x does

$$\int_0^x e^{-x^2}\,dx = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{n!(2n+1)} \quad \text{converge?}$$

76.6. For what values of a and b does

$$\int_0^1 \frac{x^{a-1}}{1+x^b}\,dx = \sum_{n=0}^{\infty} \frac{(-1)^n}{a+nb} \quad \text{converge?}$$

76.7. For what values of x and r does

$$\frac{r\sin x}{1 - 2r\cos x + r^2} = \sum_{n=1}^{\infty} r^n \sin nx \quad \text{converge?}$$

77.1. For what values of a and x does the binomial expansion of $(1+x)^a$ converge?

77.2. For what values of x does

$$K(x) = \frac{\pi}{2}\left(1 + \frac{1^2}{2^2}x^2 + \frac{1^2 3^2}{2^2 4^2}x^4 + \frac{1^2\cdot 3^2\cdot 5^2}{2^2\cdot 4^2\cdot 6^2}x^6 + \ldots\right)$$

converge?

77.3. Determine the radius of convergence of the series

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} + \ldots$$

$$\tan^{-1} z = z - \frac{z^3}{3} + \frac{z^5}{5} + \ldots.$$

77.4. For what values of x does

$$\int_0^x e^{-t^2}\,dt = \frac{\sqrt{\pi}}{2} - \frac{e^{-x}}{2x}\left(1 - \frac{1}{2x^2} + \frac{1\cdot 3}{2^2 x^4} - \frac{1\cdot 3\cdot 5}{2^3 x^6} + \ldots\right)$$

converge?

77.5. For what values of z does the series $\sum_{n=0}^{\infty} \frac{z^n}{n^2}$ converge?

81.1. For what domain of z is the function $(1 + z)z^*$ analytic?

81.2. For what domain of z is the function $\frac{z + z^*}{1 + |z|}$ analytic?

81.3. For what domain of z is the function $\frac{z^2 - z^{*2}}{|z|^2}$ analytic?

81.4. For what domain of z is the function

$$f(z) = \frac{(z + z^*)(z - z^*)^2 z}{(z + z^*)^2 + \frac{1}{4}(z - z^*)^4} \qquad \text{for } z \neq 0$$
$$= 0 \qquad \text{for } z = 0$$

analytic?

81.5. For what values of z is $e|z|$ analytic?

82.1. Show that if $w = u + iv$ is an analytic function, the curves $u =$ constant, and $v =$ constant are orthogonal.

82.2. Find the analytic function $w = u + iv$ of which $u = e^x \cos y$ is the real part.

82.3. Find the analytic function $w = u + iv$ of which $u = \frac{1}{2} \ln (x^2 + y^2)$ is the real part. Sketch the surfaces of constant u and v.

82.4. Find the analytic function of which the parabolic coordinates ξ, η are the real and imaginary parts [see eq. (49.1)]. Express the result in the form $w = f(z)$.

83.1. Compute the integral of $f(z) = 1/z$ around a unit circle with center at the origin.

83.2. Compute the integral of $1/z$ around a square of side $2a$ with center at the origin.

84.1. By expanding $1/(1 + x^2)$ according to the binomial theorem and integrating term by term, show that

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \ldots .$$

Express the following series in closed form by operating on both sides until a recognizable series is found.

84.2. $f(x) = x + \frac{x^3}{3} + \frac{x^5}{5} + \frac{x^7}{7}$.

84.3. $f(x) = \log x + \frac{x}{1!} + \frac{x^2}{2 \cdot 2!} + \frac{x^3}{3 \cdot 3!}$.

86.1. Using the inequality

$$\left| \oint f(z)dz \right| \leq \oint |f(z)dz|$$

and Cauchy's formula, show that

$$|f(a)| \leq \frac{1}{2\pi} \frac{Ml}{\rho}$$

where M is the greatest value of $|f(z)|$ on a curve C enclosing a, and l is the length of C, and ρ is the smallest distance from C to a. Hence show that a function cannot have a greatest magnitude in a domain in which it is analytic.

87.1. Show that $|f^{(n)}(a)| \leq \frac{\Gamma(n + 1)M}{\rho^n}$. (See Problem 86.1).

88.1. Use Taylor's theorem and the uniqueness of power series to show

$$\frac{(n/2)!\,d^n}{n!\,dx^n}(1-x^2)^n\Big\}_{x=0} = \begin{cases} 0 \text{ if } n \text{ is odd} \\ (-1)^{n/2} \text{ if } n \text{ is even.} \end{cases}$$

88.2. For $x^2 < 1$ the expansion

$$(1 - 2xh + h^2)^{-1/2} = 1 + P_1(x)h + P_2(x)h^2 + \ldots$$

defines the Legendre polynomials $P_l(x)$. By integrating both sides evaluate

$$\int_{-1}^{1} P_e(x)dx.$$

88.3. Show that the (real) function e^{-1/x^2} has derivatives of all orders at the origin. Why can it not be expanded in a Taylor's series about the point $x = 0$? For what values of x will the power series expansions of the following functions about the point $x = 0$ converge?

88.4. $\dfrac{1}{(a + bx^2)^2}$, where $0 < a < b$.

88.5. $\dfrac{1}{1 + e^x}$.

88.6. $\dfrac{1}{(a + b\cos x)^2}$, where $0 < b < a$.

91.1. The Bessel's functions of integral order are given by the Laurent expansion of

$$e^{(x/2)(z-1/2)} = \sum_{n=0}^{\infty} J_n(x)z^n + (-1)^{n+1}\sum_{n=1}^{\infty} J_n(x)\frac{1}{z^n}.$$

Show that $J(x)_n = \dfrac{1}{\pi}\displaystyle\int_0^{\pi} \cos(n\theta - x\sin\theta)d\theta$.

91.2. By differentiating both sides show that

$$\frac{d}{dx}J_n(x) = \frac{1}{2}(J_{n-1} - J_{n+1}).$$

91.3. Expand $1/(1 - e^z)$ in the neighborhood of $z = 0$.

91.4. Expand $\cot z$ in the neighborhood of $z = 0$.

Chapter 7

RESIDUES AND COMPLEX INTEGRATION

93. Theory of residues. The theory of residues in connection with the integral of a complex function taken around a suitable contour provides one of the most powerful methods of evaluating various types of definite integrals.

Let $f(z)$ be a single-valued analytic function throughout a certain domain D of the complex plane with the exception of the point $z = a$, where it possesses a pole of order m. In accordance with Sec. 90, this means that $f(z)$ may be represented by the Laurent series

$$f(z) = P(z-a) + \frac{c_{-1}}{z-a} + \frac{c_{-2}}{(z-a)^2} + \frac{c_{-3}}{(z-a)^3} + \cdots + \frac{c_{-m}}{(z-a)^m} \tag{93.1}$$

where $P(z-a)$ is an ordinary power series in $(z-a)$. The coefficient c_{-1} in this expansion is called the residue of $f(z)$ at the pole a, for a reason that will presently be apparent. Surround the point $z = a$ with a circle γ of radius ρ which lies in D and consider the integral $\oint_\gamma f(z)dz$. From (93.1),

$$\oint_\gamma f(z)dz = \oint_\gamma \sum_{r=1}^{m} \frac{c_{-r}}{(z-a)^r}\,dz + \oint_\gamma P(z-a)dz \tag{93.2}$$

$P(z-a)$ is an analytic function and Cauchy's theorem shows that the integral taken over the circle γ vanishes. In evaluating the remaining integral let us consider the general term

$$\oint_\gamma \frac{c_{-r}}{(z-a)^r}\,dz \qquad \text{when } r \neq 1$$

The value of the integral is found most conveniently by writing $z - a = \rho e^{i\varphi}$, so that z traverses the circle γ when φ traverses the range 0 to 2π. Hence

$$\oint_\gamma \frac{dz}{(z-a)^r} = \oint_0^{2\pi} \frac{i\rho e^{i\varphi}\,d\varphi}{\rho^r e^{ri\varphi}} = \frac{i}{\rho^{r-1}} \int_0^{2\pi} e^{-(r-1)\varphi i}\,d\varphi = 0$$

and we are left with the one term $\oint_\gamma \frac{c_{-1}\,dz}{z-a}$. Proceeding in the same way for this term,

$$\int_\gamma f(z)dz = c_{-1}\oint_\gamma \frac{dz}{z-a} = c_{-1}\int_0^{2\pi} \frac{i\rho e^{i\varphi}\,d\varphi}{\rho e^{i\varphi}} = 2\pi i c_{-1} = 2\pi i R_a(f)$$

This indicates why c_{-1} is called the residue of $f(z)$ at $z = a$. Here $R_a(f) = c_{-1}$ and means the residue of f at $z = a$.

The contour surrounding a need not be a circle for the above result to hold. This may be seen by enclosing the circle γ with any contour C lying in the domain D. In the region bounded by γ and C the function $f(z)$ is analytic, and Cauchy's theorem gives [see (85.2)]

$$\oint_C f(z)dz - \oint_\gamma f(z)dz = 0$$

so that

$$\oint_C f(z)dz = \oint_\gamma f(z)dz = 2\pi i R_a(f) \tag{93.3}$$

The path C can be altered or deformed in any way so long as it remains in the domain D and does not pass through or include any further singularities of $f(z)$. If a contour C encloses poles at the points $z = a$, $z = b$, $z = c$, etc., it is easily shown that

$$\oint_C f(z)dz = 2\pi i[R_a(f) + R_b(f) + R_c(f) + \ldots] \tag{93.4}$$

94. Determination of the residue. To find the residue of $f(z)$ at a given point it is not always convenient to determine the coefficient c_{-1} by actually obtaining the Laurent expansion for $f(z)$. When $f(z)$ has a simple pole at $z = a$, it is sometimes convenient to proceed as follows. Since

$$f(z) = \frac{c_{-1}}{z-a} + c_0 + c_1(z-a) + \ldots$$

it is necessary only to multiply $f(z)$ by $(z - a)$ and take the limit as $z \to a$ in order to select the coefficient c_{-1}. Consequently

$$c_{-1} = \lim_{z\to a}(z-a)f(z) = R_a(f) \tag{94.1}$$

When $f(z)$ has a pole of order m, then

$$f(z) = \frac{c_{-m}}{(z-a)^m} + \frac{c_{-m+1}}{(z-a)^{m-1}} + \ldots + \frac{c_{-1}}{z-a} + c_0 + c_1(z-a) + \ldots$$

First multiply $f(z)$ by $(z - a)^m$, so that

$$(z-a)^m f(z) = c_{-m} + c_{-m+1}(z-a) + \ldots + c_{-1}(z-a)^{m-1} + c_0(z-a)^m + \ldots$$

Further, if $(z - a)^m f(z)$ be differentiated $m - 1$ times, all terms up to c_{-1} vanish, and all others except c_{-1} vanish in the limit as $z \to a$, so that

$$c_{-1} = \frac{1}{(m-1)!} \lim_{z \to a} \frac{d^{m-1}}{dz^{m-1}} [(z - a)^m f(z)] = R_a(f) \qquad (94.2)$$

95. The definite integral $\int_0^{2\pi} f(\theta)d\theta$. Here $f(\theta)$ is a real rational function of $\sin\theta$ and $\cos\theta$. In order to apply the theory of residues, $f(\theta)$ must be replaced by a suitable function of a complex variable such that when integrated over a contour in the z plane the above integral results. The procedure may best be illustrated by means of the example

$$I = \int_0^{2\pi} \frac{d\theta}{a + b\cos\theta}$$

Let $z = e^{i\theta}$, from which it follows that

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} = \frac{1}{2}\left(z + \frac{1}{z}\right)$$

and

$$dz = ie^{i\theta}\, d\theta = iz\, d\theta$$

When θ traverses the interval 0 to 2π, the complex variable z traverses the unit circle C with center at $z = 0$. Hence

$$I = \oint_C \frac{-i\, dz}{z[a + (b/2)(z + 1/z)]} = \frac{-2i}{b} \oint_C \frac{dz}{z^2 + (2a/b)z + 1}$$

The denominator of the integrand may be written in the form $(z - \alpha)(z - \beta)$, where α and β are the roots of $z^2 + \frac{2a}{b}z + 1$, namely,

$$\alpha = \frac{-a + \sqrt{a^2 - b^2}}{b}, \qquad \beta = \frac{-a - \sqrt{a^2 - b^2}}{b}$$

The values of α and β may of course be complex. In any case the integrand has simple poles at $z = \alpha$ and $z = \beta$. In order to apply the theory of residues we need to know which if any of these points lie within the unit circle C.

Since $|\alpha\beta| = 1$, then of the two numbers $|\alpha|$ and $|\beta|$ one is greater than unity and the other is less than unity, or both are equal to unity. In the latter case, α and β can be written $\alpha = e^{i\gamma}$ and $\beta = e^{-i\gamma}$, whence $\alpha + \beta = 2\cos\gamma = -2a/b$, which requires that $-1 \leq a/b \leq 1$. For this situation α and β lie on the unit circle and the residue theory cannot be applied. This case can be excluded, however, because $b > a$ makes the original integral singular. When $|a - \sqrt{a^2 - b^2}| < |b|$, then $|\alpha| < 1$ and $|\beta| > 1$, so that the circle C encloses one pole at $z = \alpha$ and

$$I = \left(\frac{-2i}{b}\right) 2\pi i R_a \left[\frac{1}{(z - \alpha)(z - \beta)}\right]$$

From Sec. 94,

$$R\left[\frac{1}{(z-\alpha)(z-\beta)}\right] = \lim_{z\to\alpha}\frac{1}{z-\beta} = \frac{1}{\alpha-\beta} = \frac{b}{2\sqrt{a^2-b^2}}$$

giving the final result

$$\int_0^{2\pi}\frac{d\theta}{a+b\cos\theta} = \frac{2\pi}{\sqrt{a^2-b^2}} \qquad |a|>|b| \tag{95.1}$$

This result is an example of the general theorem

$$\int_0^{2\pi} f(\cos\theta,\sin\theta)d\theta = \int_C f\left(\frac{z+z^{-1}}{2},\frac{z-z^{-1}}{2i}\right)\frac{dz}{iz}$$
$$= 2\pi\sum_r R_{a_r}\left[\frac{1}{z}f\left(\frac{z+z^{-1}}{2},\frac{z-z^{-1}}{2i}\right)\right]$$

Where a_r are the poles of f/z enclosed in the unit circle, f must be a rational function of $\cos\theta$ and $\sin\theta$.

96. The integral $\int_{-\infty}^{\infty} F(x)dx$. Integrals of this type can be evaluated by means of the theory of residues, if when the real variable x is replaced by the complex variable z, the function $F(z)$ satisfies the following conditions.

(1) $F(z)$ is analytic except at a finite number of poles when the imaginary part of z is positive or zero.

(2) $F(z)$ has no poles on the real axis.

(3) $\lim_{|z|\to\infty} zF(z) = 0$ for all values $0 \le \arg z \le \pi$.

When these conditions are satisfied, the integral can be evaluated successfully by choosing the path of integration as shown in Fig. 96.1. The contour C extends along the real axis from $z = -\rho$ to $z = +\rho$, and then over the semicircle Γ of radius ρ in the upper half plane. The value of ρ is chosen so that all the poles of $F(z)$ in the upper half plane are enclosed in the resulting contour. The theory of residues gives

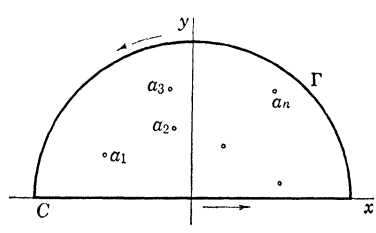


FIG. 96.1

$$\int_C F(z)dz = \int_{-\rho}^{+\rho} F(x)dx + \int_\Gamma F(z)dz$$
$$= 2\pi i\sum_{p=1}^{n} R_{a_p}[F(z)] \tag{96.1}$$

the poles of $F(z)$ for this positive half plane being at $a_1, a_2, \ldots, a_n$. The original integral

$$\int_{-\infty}^{+\infty} F(x)dx = \lim_{\rho\to\infty} \int_{-\rho}^{+\rho} F(x)dx$$

The expression (96.1) would yield the value of the original integral if $\lim_{\rho\to\infty} \int_\Gamma F(z)dz = 0$. That this is the case is easily shown since

$$\left| \int_\Gamma F(z)dz \right| \leq \int_\Gamma \left| F(z) \right| \left| dz \right| = \int_0^\pi \left| F(\rho e^{i\theta}) \right| \rho \, d\theta$$

and by condition 3 above,

$$\lim_{\rho\to\infty} \int_0^\pi \rho \left| F(\rho e^{i\theta}) \right| d\theta = 0$$

Consequently

$$\int_{-\infty}^{+\infty} F(x)dx = 2\pi i \sum_{p=1}^{n} R_{a_p}[F(z)] \tag{96.2}$$

As will be shown in subsequent sections, condition 3 can be made less restrictive when $F(z)$ is multiplied by $\sin z$ or $\cos z$. Also the theory can sometimes be applied successfully, if condition 3 is not satisfied, by selecting a suitable contour of integration. Examples of this are given in Sec. 97. It should be noticed that if $F(x)$ is an even function the method will handle integrals of the form $\int_0^\infty F(x)dx$.

To illustrate the above method, let us evaluate the integral

$$I = \int_{-\infty}^{+\infty} \frac{dx}{(x^2 + a^2)^2}$$

Here $F(z) = 1/(z^2 + a^2)^2$ and satisfies all the conditions given above, so that it is only a question of determining the residues at the poles in the positive half plane. Since

$$\frac{1}{(z^2 + a^2)^2} = \frac{1}{(z + ia)^2(z - ia)^2}$$

it is evident that the function has a pole in the positive half plane of order 2 at $z = ia$. From (94.2),

$$R_{ia}[F(z)] = \lim_{z\to ia} \frac{d}{dz} \frac{1}{(z + ia)^2} = \lim_{z\to ia} \frac{-2}{(z + ia)^3} = \frac{1}{i4a^3}$$

Thus

$$\int_{-\infty}^{+\infty} \frac{dx}{(x^2 + a^2)^2} = \frac{\pi}{2a^3}$$

97. Infinite integrals continued. Sometimes integrals over an infinite range can be evaluated by means of residues by cleverly choosing a suitable complex function other than the one obtained by replacing x by the complex variable z and integrating over the proper path. Since

there are no rules for determining the right complex function or path of integration completely, success is largely a matter of experience. The situation can be made clear by attempting to evaluate the integral

$$I = \int_{-\infty}^{+\infty} e^{-x^2} \cos 2ax \, dx$$

It is evident to begin with that a path of the type used in the preceding section cannot be used, because condition 3 of that section is not satisfied, since in particular

$$\lim_{|z|\to\infty} e^{-z^2} \cos 2az = \infty \quad \text{for} \quad \arg z = \pi/2$$

that is $|z|$ directed along the positive imaginary axis.

It is nevertheless possible to evaluate the above integral by integrating the function e^{-z^2} around the rectangular contour shown in Fig. 97.1. The reason underlying this choice will become apparent during the process of evaluation. Consider the integral $\int_C e^{-z^2} dz$. Since

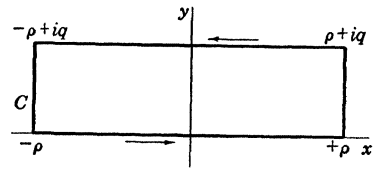


FIG. 97.1

there are no poles of e^{-z^2}, Cauchy's theorem gives $\int_C e^{-z^2} dz = 0$. Hence

$$\int_C e^{-z^2} dz = \int_{-\rho}^{+\rho} e^{-x^2} dx + \int_0^a e^{-(\rho+iy)^2} i \, dy + \int_{\rho}^{-\rho} e^{-(x+ia)^2} dx + \int_a^0 e^{-(-\rho+iy)^2} i \, dy = 0$$

$$= \int_{-\rho}^{+\rho} e^{-x^2} dx - e^{a^2} \int_{-\rho}^{+\rho} e^{-x^2-2aix} dx - 2e^{-\rho^2} \int_0^a e^{y^2} \sin 2\rho y \, dy = 0$$

The second integral on the right will yield the integral we seek. It may be written

$$e^{a^2} \int_{-\rho}^{+\rho} e^{-x^2}(\cos 2ax - i \sin 2ax) dx = e^{a^2} \int_{-\rho}^{+\rho} e^{-x^2} \cos 2ax \, dx$$

Since $e^{-x^2} \sin 2ax$ is an odd function its integral vanishes. Therefore

$$I = \lim_{\rho\to\infty} \int_{-\rho}^{+\rho} e^{-x^2} \cos 2ax \, dx$$

$$= e^{-a^2} \lim_{\rho\to\infty} \int_{-\rho}^{+\rho} e^{-x^2} dx$$

$$= 2e^{-a^2} \int_0^{\infty} e^{-x^2} dx = e^{-a^2} \int_0^{\infty} e^{-t} t^{-1/2} dt$$

$$= e^{-a^2} \Gamma\left(\frac{1}{2}\right) = e^{-a^2} \sqrt{\pi} \quad .$$

the integral $\int_0^{\infty} e^{-x^2} dx$ having been evaluated in Sec. 62.

98. Integrals of the type $\int_{-\infty}^{+\infty} F(x)e^{i\mu x}\,dx$. This integral is included in the class discussed in Sec. 96, and could be evaluated by the method given there if $F(x)$ satisfies the conditions prescribed in Sec. 96, since then $F(x)e^{i\mu x}$ would also satisfy the conditions. The above integral warrants some additional discussion, since condition 3 of Sec. 96 can be replaced by the condition that

$$\lim_{|z|\to\infty} F(z) = 0 \quad \text{when} \quad 0 \leq \arg z \leq \pi$$

First it may be noted that if the method of residues of Sec. 96 can be applied to the above integral,

$$\int_{-\infty}^{+\infty} F(x)e^{i\mu x}\,dx = 2\pi i \sum_{p=1}^{n} R_{a_p}[F(z)e^{i\mu z}] \qquad \mu > 0 \qquad (98.1)$$

On the other hand,

$$\int_{-\infty}^{+\infty} F(x)e^{i\mu x}\,dx = \int_0^{\infty} [F(x)e^{i\mu x} + F(-x)e^{-i\mu x}]dx \qquad \mu > 0$$

whence it is readily seen that *if* $F(x)$ is an even function, i.e., $F(x) = F(-x)$, then

$$\int_{-\infty}^{+\infty} F(x)e^{i\mu x}\,dx = 2\int_0^{\infty} F(x)\cos \mu x\,dx \qquad \mu > 0 \qquad (98.2)$$

If $F(x)$ is an odd function, i.e., $F(-x) = -F(x)$, then

$$\int F(x)e^{i\mu x}\,dx = +2i\int_0^{\infty} F(x)\sin \mu x\,dx \qquad \mu > 0 \qquad (98.3)$$

When these results are combined with eq. (98.1), we find that the method of residues may be used to evaluate integrals involving sines and cosines, so that

(1) if $F(x)$ is an even function,

$$\int_0^{\infty} F(x)\cos \mu x\,dx = \pi i \sum_{\rho=1}^{n} R_{a_\rho}[F(z)e^{i\mu z}] \qquad \mu > 0 \qquad (98.4)$$

(2) if $F(x)$ is an odd function,

$$\int_0^{\infty} F(x)\sin \mu x\,dx = \pi \sum_{\rho=1}^{n} R_{a_\rho}[F(z)e^{i\mu z}] \qquad \mu > 0 \qquad (98.5)$$

where the sum is taken over the residues of the function $F(z)e^{i\mu z}$ over all poles in the positive half of the complex plane.

It may be readily shown that if $F(z)$ satisfies conditions (1) and (2) of Sec. 96, and in addition if $\lim_{|z|\to\infty} F(z) = 0$ for all $0 \leq \arg z \leq \pi$, the method of residues may be used. Using the same path as in Fig. 96.1, it is necessary only to show that

$$\lim_{\rho\to\infty}\int_\Gamma F(z)e^{i\mu z}\,dz = 0 \qquad \mu > 0$$

Over the semicircular path Γ one may write

$$\left|\int_\Gamma F(z)e^{i\mu z}\,dz\right| \le \int_0^\pi \left|F(\rho e^{i\theta})e^{i\mu\rho(\cos\theta+i\sin\theta)}\rho i e^{i\theta}\right| d\theta$$
$$\le \int_0^\pi \left|F(\rho e^{i\theta})\right| \left|\rho e^{-\mu\rho\sin\theta}\right| d\theta$$

From the above condition on $F(z)$ it is possible to choose ρ such that $|F(\rho e^{i\theta})| < \epsilon/\pi$, where ϵ is an arbitrarily small quantity, so that

$$\left|\int_\Gamma F(z)e^{i\mu z}\,dz\right| \le \frac{\epsilon}{\pi}\int_0^\pi \rho e^{-\mu\rho\sin\theta}\,d\theta$$
$$\le \frac{2\epsilon}{\pi}\int_0^{\pi/2} \rho e^{-\mu\rho\sin\theta}\,d\theta$$

Referring to Fig. 98.1, it is obvious that $\sin\theta \ge 2\theta/\pi$ for $0 \le \theta \le \pi/2$. Thus

$$\int_0^{\pi/2} \rho e^{-\mu\rho\sin\theta}\,d\theta \le \int_0^{\pi/2} \rho e^{-2\mu\rho\theta/\pi}\,d\theta = \frac{\pi}{2\mu}(1 - 2e^{-\mu\rho})$$

and the final result is

$$\left|\int_\Gamma F(z)e^{i\mu z}\,dz\right| \le \frac{\epsilon}{\mu}(1 - 2e^{-\mu\rho})$$

and the integral over Γ vanishes in the limit $\rho \to \infty$ if $\mu > 0$. This establishes the validity of eqs. (98.1), (98.4), and (98.5) for a less stringent condition on $F(z)$.

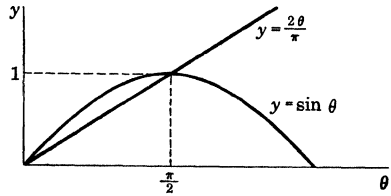


FIG. 98.1

The method will be illustrated by evaluating the integral

$$I = \int_0^\infty \frac{x\sin\mu x}{x^2 + k^2}\,dx, \qquad \mu > 0, \quad k > 0$$

The function $z/(z^2 + k^2)$ satisfies all the required conditions and $F(x)$ is an odd function as required. To evaluate I it is necessary only to find the residues of $ze^{i\mu z}/(z^2 + k^2)$ in the upper half of the complex plane. The function has only one pole in the upper half plane at the point $z = ik$; hence

$$R_{ik}\left(\frac{ze^{i\mu z}}{z^2 + k^2}\right) = \lim_{z\to ik}\frac{ze^{i\mu z}}{z + ik} = \frac{e^{-\mu k}}{2}$$

Consequently (98.5) gives

$$\int_0^\infty \frac{x\sin\mu x\,dx}{x^2 + k^2} = \frac{\pi}{2}e^{-\mu k} \qquad \mu > 0, \quad k > 0 \tag{98.6}$$

This example shows the ease with which certain definite integrals may be evaluated by this method.

99. Integrals involving hyperbolic functions. The application of the theory of residues to integrals involving hyperbolic functions usually presents a somewhat more complicated problem and at the same time indicates the possibilities of the method of contour integration. Here the path of integration and even the complex function must be chosen to suit the particular integral. The situation and procedure may be best illustrated by considering an example. Suppose the value of the integral

$$I = \int_0^\infty \frac{\cosh ax}{\cosh \pi x}\, dx$$

is desired, where $-\pi < a < \pi$. The integrand does not satisfy condition 3 in Sec. 96, so that the standard path used there is not applicable here.

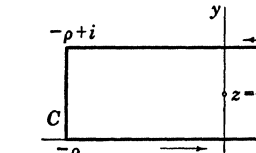


FIG. 99.1

The integral can be evaluated however by integrating the function $e^{az}/\cosh \pi z$ around the rectangular path C shown in Fig. 99.1. The function $e^{az}/\cosh \pi z$ has poles at the points where $\cosh \pi z$ has zeros. Since $\cosh \pi x > 0$, there are no zeros on the real axis.

It is first necessary to see whether

$$\int_C \frac{e^{az}\, dz}{\cosh \pi z}$$

can be made to yield the value of I. If this integral is written as the sum of the integrals over the separate sides of the rectangle, the result is

$$\begin{aligned}\int_C \frac{e^{az}}{\cosh \pi z}\, dz &= \int_{-\rho}^{+\rho} \frac{e^{ax}}{\cosh \pi x}\, dx + \int_0^1 \frac{e^{a(\rho+iy)} i\, dy}{\cosh \pi(\rho + iy)} \\ &\quad + \int_{\rho}^{-\rho} \frac{e^{a(x+i)}\, dx}{\cosh \pi(x + i)} + \int_1^0 \frac{e^{a(-\rho+iy)} i\, dy}{\cosh \pi(-\rho + iy)} \\ &= (1 + e^{ai}) \int_{-\rho}^{+\rho} \frac{e^{ax}}{\cosh \pi x}\, dx + A \qquad (99.1)\end{aligned}$$

where the relation $\cosh \pi(x + i) = -\cosh \pi x$ has been used. This furnishes the reason for taking the upper side of the rectangle along the line $z = x + i$. Here A denotes the sum of the remaining two integrals over the ends of the rectangle. Next it is to be noted that

$$\int_{-\rho}^{+\rho} \frac{e^{ax}}{\cosh \pi x}\, dx = \int_{-\rho}^{0} \frac{e^{ax}\, dx}{\cosh \pi x} + \int_{0}^{\rho} \frac{e^{ax}\, dx}{\cosh \pi x}$$

$$= \int_{0}^{\rho} \frac{e^{ax} + e^{-ax}}{\cosh \pi x}\, dx = 2\int_{0}^{\rho} \frac{\cosh ax}{\cosh \pi x}\, dx \qquad (99.2)$$

This shows that the function $e^{az}/\cosh \pi z$ integrated over the chosen path will yield the original integral, but it is still necessary to evaluate A, and in fact when the path was chosen it was hoped either that $\lim_{\rho \to \infty} A = 0$ or that the resulting integral could be evaluated. We shall first see if A vanishes, by writing

$$|A| \leq \left| \int_0^1 \frac{e^{a(\rho+iy)} i\, dy}{\cosh \pi(\rho + iy)} \right| + \left| \int_0^1 \frac{e^{a(-\rho+iy)} i\, dy}{\cosh \pi(-\rho + iy)} \right|$$

$$\leq \int_0^1 \frac{|e^{a(\rho+iy)}| dy}{|\cosh \pi(\rho + iy)|} + \int_0^1 \frac{|e^{a(-\rho+iy)}|dy}{|\cosh \pi(-\rho + iy)|}$$

Since

$$|\cosh \pi(\rho \pm iy)| = \frac{1}{2}\left| e^{\pi(\rho \pm iy)} + e^{-(\rho \pm iy)} \right| \geq \left| \frac{e^{\pi(\rho \pm iy)}}{2} \right| - \left| \frac{e^{-\pi(\rho \pm iy)}}{2} \right|$$

$$\geq \frac{e^{\pi\rho} - e^{-\pi\rho}}{2} = \sinh \pi\rho$$

Therefore

$$\left.\begin{aligned} |A| \leq \int_0^1 \frac{e^{a\rho}\, dy}{\sinh \pi\rho} + \int_0^1 \frac{e^{-a\rho}\, dy}{\sinh \pi\rho} &= 2\frac{e^{a\rho} + e^{-a\rho}}{e^{\pi\rho} - e^{-\pi\rho}} \\ &= 2\frac{e^{(a-\pi)\rho} + e^{-(a+\pi)\rho}}{1 - e^{-2\pi\rho}} \end{aligned}\right\} \qquad (99.3)$$

and

$$\lim_{\rho \to \infty} |A| = 0$$

provided $|a| < \pi$.

It is now necessary only to locate the poles of the function $e^{az}/\cosh \pi z$ and determine the residues of those located within the contour C. The zeros of $\cosh \pi z$ may easily be found by writing

$$\cosh \pi z = \cosh \pi x \cos \pi y + i \sinh \pi x \sin \pi y$$

The right-hand side vanishes only for $x = 0$, $y = (n - \frac{1}{2})$, so there is only one pole within our chosen contour at $z = i/2$.

When (99.1), (99.2), and (99.3) are combined, we have the result

$$\int_C \frac{e^{az}}{\cosh \pi z}\, dz = 2(1 + e^{ai}) \int_0^{+\infty} \frac{\cosh ax}{\cosh \pi x}\, dx$$

$$= 2\pi i R_{i/2}\left(\frac{e^{az}}{\cosh \pi z}\right) \qquad (99.4)$$

A convenient way of determining the residues at $z = i/2$ is to let $t = z - i/2$, so that

$$\frac{e^{az}}{\cosh \pi z} = \frac{e^{ai/2+at}}{\cosh \pi(i/2 + t)} = \frac{e^{ai/2+at}}{i \sinh \pi t}$$
$$= \frac{e^{ai/2}(1 + at + \text{high powers of } t)}{i\pi t(1 + \frac{1}{6}\pi^2 t^2 + \ldots)}$$

Since by definition the residue is the coefficient of the first inverse power of $(z - i/2)$, it is therefore $e^{ai/2}/i\pi$. Substituting this in (99.4), we have the final result

$$\int_0^\infty \frac{\cosh ax}{\cosh \pi x}\, dx = \frac{e^{ai/2}}{1 + e^{ai}} = \frac{1}{2 \cos a/2} = \frac{1}{2} \sec \frac{a}{2} \qquad (99.5)$$

A rectangular contour is often successful when the integrand does not satisfy condition 3 in Sec. 96.

100. The principal value of integrals. In applying the theory of residues to the evaluation of definite integrals thus far, it has been supposed that the integrand has no pole on the real axis. When the integrand has a finite number of poles on the real axis one may still use the theory of residues to evaluate the integral, but the value is of necessity (see Sec. 41) the principal value of the integral. If, for example, the integrand satisfies the conditions in Secs. 96 or 98 except for a finite number of poles on the real axis,

$$\int_{-\infty}^{\infty} F(x)dx = 2\pi i \left\{ \sum_{p=1}^{n} R_{a_p}[F(z)] + \frac{1}{2} \sum_{p=1}^{m} R_{b_p}[F(z)] \right\} \qquad (100.1)$$

where the term on the right denotes one-half the sum of the residues of $F(z)$ at the poles on the real axis.

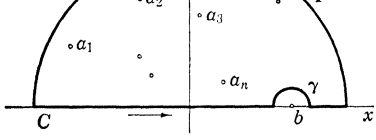


FIG. 100.1

In order to see how this comes about, suppose that $F(z)$ satisfies the conditions of Sec. 96 except that it has a pole at $z = b$ on the real axis. A contour C similar to that used in Sec. 96 is used except that the contour is indented at $x = b$ as shown in Fig. 100.1. The indentation is a semicircle γ of radius δ. Then as in Sec. 96,

$$\int_C F(z)dz = \int_{-\rho}^{b-\delta} F(x)dx + \int_\gamma F(z)dz + \int_{b+\delta}^{\rho} F(x)dx + \int_\Gamma F(z)dz$$
$$= 2\pi i \sum_{p=1}^{n} R_{a_p}[F(z)]$$

Passing to the limit $\rho \to \infty$, the integral over Γ vanishes as it did in Sec. 96, so that

$$\int_{-\infty}^{b-\delta} F(x)dx + \int_{b+\delta}^{\infty} F(x)dx$$

$$= 2\pi i \sum_{p=1}^{n} R_{a_p}[F(z)] - \int_{\gamma} F(z)dz$$

By definition the principal value of the integral is

$$\lim_{\delta \to 0} \int_{-\infty}^{b-\delta} F(x)dx + \int_{b+\delta}^{\infty} F(x)dx$$

$$= 2\pi i \sum_{p=1}^{n} R_{a_p}[F(z)] - \lim_{\delta \to 0} \int_{\gamma} F(z)dz$$

The limit of the integral on the right may easily be found, since it is known that $F(z)$ can be written in the form

$$F(z) = \frac{c_{-1}}{z - b} + c_0 + c_1(z - b) + \ldots$$

from which

$$\int_{\gamma} F(z)dz = \int_{\pi}^{0} \frac{c_{-1}}{\delta e^{i\theta}} \delta i e^{i\theta} d\theta + c_0 \delta i \int_{\pi}^{0} e^{i\theta} d\theta + \text{higher powers of } \delta$$

and $$\lim_{\delta \to 0} \int_{\gamma} F(z)dz = -\pi i c_{-1} = -2\pi i \left(\frac{c_{-1}}{2}\right)$$

This indicates clearly why only one-half the residue is required in (100.1) for the poles on the real axis.

The artifice of indenting the contour is often useful in rectangular paths. In those cases where the indentation is made when a pole lies at the corner of a rectangle, only one-quarter of the residue at that point is required.

An interesting example of the above method of finding the principal value of an integral is furnished by the integral

$$I = \frac{1}{2\pi i} \int_{-\infty}^{+\infty} \frac{e^{tix}}{x} dx \tag{100.2}$$

If $t > 0$ we may use (100.1) directly, the result being

$$I = \frac{1}{2} R_0[F(z)] = \frac{1}{2}$$

For the case $t = 0$, the integrand does not satisfy the condition 3, Sec. 96, namely, $\lim_{|z| \to \infty} zF(z) = 0$ for $0 \leq \arg z \leq \pi$, so that the contribution over circle Γ, Fig. 100.1, does not vanish. However

$$\int_{-\rho}^{+\rho} \frac{dz}{z} + \int_{\Gamma} \frac{dz}{z} = 2\pi i \frac{1}{2} R_0\left(\frac{1}{z}\right) = \pi i$$

and $$\int_\Gamma \frac{dz}{z} = \int_0^\pi \frac{\rho i e^{i\theta}\, d\theta}{\rho e^{i\theta}} = \pi i$$

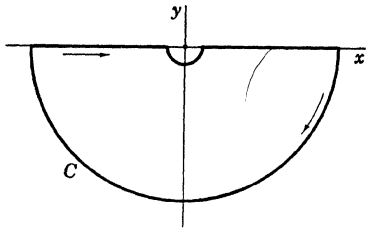


FIG. 100.2

Consequently

$$\lim_{\rho\to\infty} \int_{-\rho}^{+\rho} \frac{dz}{z} = 0$$

For $t < 0$ one may use the contour of Fig. 100.2, whereupon the result is

$$\frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{e^{tix}\, dx}{x} = -\frac{1}{2}$$

Hence I represents the discontinuous function

$$I = \frac{1}{2\pi i} \int_{-\infty}^{+\infty} \frac{e^{tix}}{x}\, dx = \begin{cases} -\frac{1}{2} & \text{for } t < 0 \\ 0 & \text{for } t = 0 \\ +\frac{1}{2} & \text{for } t > 0 \end{cases} \qquad (100.3)$$

The function I is closely related to the so-called impulse function used in the Heaviside operational analysis of electric circuits.

101. Integrals of the form $\int_0^\infty x^{m-1}F(x)dx$. This integral will serve as an illustration of the method of treatment when the integrand has one or more branch points. It will be supposed that $F(z)$ is analytic except for isolated poles not lying on the real axis (otherwise we would be dealing with the principal value of the integral, see last section) and such that $\lim_{|z|\to\infty} z^m F(z) = 0$ for $0 \leq \arg z \leq 2\pi$ and $\lim_{|z|\to 0} z^m F(z) = 0$. In order to evaluate the above integral it is convenient to consider the contour integral $\int_C (-z)^{m-1}F(z)dz$ over a suitable contour C. If m is not an integer but either a rational or irrational number, $(-z)^{m-1}$ is a finite or infinitely many-valued function (see Sec. 92) and $z = 0$ is a branch point. In applying the theory of residues this point must not lie within the closed contour, since on going completely around the branch point the function would not have the same value as at the starting point, i.e., we must keep the contour on one Riemann sheet. In order to specify $(-z)^{m-1}$ uniquely we write

$$(-z)^{m-1} = e^{(m-1)[\log|z| + i \arg(-z) + 2k\pi i]}$$

and select the principal value of the logarithm, i.e., $k = 0$. Then from Sec. 92, it is known that the principal value of the logarithm and consequently $(-z)^{m-1}$, are single-valued for $-\pi < \arg(-z) \leq \pi$. With this definition of $(-z)^{m-1}$ it is readily seen that the integrand is analytic except at the poles, and any path C' as shown in Fig. 101.1a which

includes the poles of $F(z)$ and excludes $z = 0$ is a legitimate one. Here C' may be deformed in any way without changing the value of $\int_{C'} (-z)^{m-1}F(z)dz$, provided no new poles are included within it or the real axis ($\arg z = 0$) is not crossed. In particular the contour can

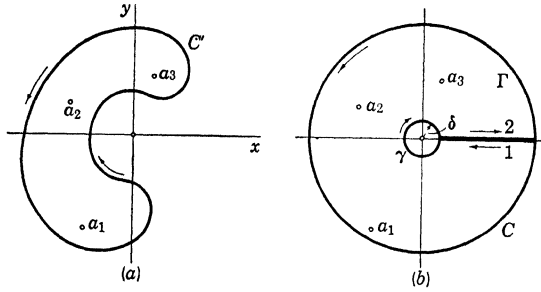


FIG. 101.1

be deformed to the one shown in Fig. 101.1b, where γ and Γ are the circles indicated with radii δ and ρ, respectively. Then

$$\int_C (-z)^{m-1}F(z)dz = 2\pi i \sum_{p=1}^{n} R_{a_p}[(-z)^{m-1}F(z)]$$
$$= \int_\Gamma + \int_1 + \int_\gamma + \int_2 (-z)^{m-1}F(z)dz \quad (101.1)$$

Over the circle Γ, $-z = \rho e^{i\varphi}$, where φ is measured with respect to the negative real axis and hence goes from $-\pi$ to $+\pi$. Similarly on the circle δ, $-z = \delta e^{i\varphi}$, and φ goes from $+\pi$ to $-\pi$. On path 1, $-z = xe^{i\pi}$, and on path two $-z = xe^{-i\pi}$. When these expressions for $(-z)$ are used in the appropriate integrals eq. (101.1) becomes

$$\int_{-\pi}^{+\pi} \rho^{m-1}e^{(m-1)i\varphi}F(-\rho e^{i\varphi})(-\rho i)e^{i\varphi}\,d\varphi$$
$$+\int_{\rho}^{\delta} x^{m-1}e^{(m-1)i\pi}F(-xe^{i\pi})(-e^{i\pi}\,dx)$$
$$+\int_{\pi}^{-\pi} \delta^{m-1}e^{(m-1)i\varphi}F(-\delta e^{i\varphi})(-\delta i e^{i\varphi})d\varphi$$
$$+\int_{\delta}^{\rho} x^{m-1}e^{-(m-1)i\pi}F(-xe^{-i\pi})(-e^{-i\pi}\,dx)$$
$$= 2\pi i \sum_{p=1}^{n} R_{a_p}[(-z)^{m-1}F(z)]$$

Because of the restriction placed upon m and $F(x)$ at the beginning of this section, it is readily seen that in the limit $\rho \to \infty$ and $\delta \to 0$ the

integrals over Γ and γ vanish, so that the above equation becomes

$$e^{m\pi i}\int_0^\infty x^{m-1}F(x)dx - e^{-m\pi i}\int_0^\infty x^{m-1}F(x)dx$$
$$= 2\pi i \sum_{p=1}^{n} R_{a_p}[(-z)^{m-1}F(z)]$$

or

$$\int_0^\infty x^{m-1}F(x)dx = \frac{\pi \sum_{p=1}^{n} R_{a_p}[(-z)^{m-1}F(z)]}{\sin m\pi} \qquad (101.2)$$

This formula obviously does not hold when m is an integer. As an example, consider the integral

$$\int_0^\infty \frac{dx}{x^{2/3}(x^2+a^2)}$$

The poles of the integrand lie at $z = ia$ and $z = -ia$; the residues of $(-z)^{1/3-1}/(z^2+a^2)$ are

$$\frac{a^{1/3-1}e^{-(1/3-1)\pi i/2}}{2ai} \quad \text{and} \quad \frac{a^{1/3-1}e^{(1/3-1)\pi i/2}}{-2ai}$$

When these are added together,

$$\sum R\left[(-z)^{1/3-1}\frac{1}{z^2+a^2}\right] = \frac{e^{+2i\pi/3} - e^{-2i\pi/3}}{a^{2/3+1}2i} = \frac{\sin 2\pi/3}{a^{5/3}}$$

From 101.2,
$$\int_0^\infty \frac{dx}{x^{2/3}(x^2+a^2)} = \frac{\pi \sin 2\pi/3}{a^{5/3}\sin \pi/3}$$

102. Contour integral representation of functions. In the evaluation of definite integrals, it is often very helpful to represent one of the functions in the integrand by a contour integral, after which it may be possible to invert the order of integration, evaluate the definite integral, and then compute the value of the resulting contour integration by the theory of residues. The representation of functions by contour integrals is useful in the solution of differential equations and in obtaining asymptotic expansions (see Sec. 114).

The use of a contour integral representation of a function in evaluating definite integrals may be illustrated by considering the example

$$I = \int_0^\infty e^{-ax}J_0(bx)dx$$

where $J_0(y)$ is the Bessels function of order zero (see Sec. 169). A contour integral representation of $J_0(y)$ is

$$J_0(y) = \frac{1}{2\pi i}\int_C e^{-(y/2)(z-1/z)}\frac{dz}{z}$$

where the contour C is any closed curve including the point $z = 0$.

If this expression is used in I and the order of integration is reversed,

$$I = \frac{1}{2\pi i}\int_C \frac{dz}{z}\int_0^\infty e^{-ax-(bx/2)(z-1/z)}\,dx$$

It should be noticed that the path C can be chosen so that the real part $(b/2)(z - 1/z) > 0$. If $b > 0$, such a path is shown in Fig. 102.1. Then if $a > 0$, the value of the definite integral is

$$\int_0^\infty e^{-[a+(b/2)(z-1/z)]x}\,dx = \frac{1}{a + (b/2)(z - 1/z)}$$

Thus $$I = \frac{1}{b\pi i}\int_C \frac{dz}{z^2 + 2az/b - 1}$$

The poles of the integrand lie at

$$z_1 = \frac{-a + \sqrt{a^2 + b^2}}{b}$$

and $$z_2 = \frac{-a - \sqrt{a^2 + b^2}}{b}$$

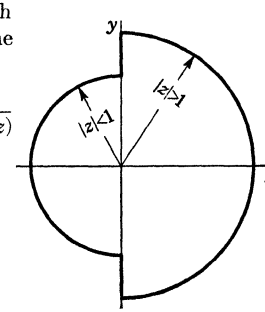


FIG. 102.1

and of these z_1 lies inside of the path C, and z_2 lies outside. The residue of $z^2 + (2az/b) - 1$ at $z = z_1$ is just $1/(z_1 - z_2)$, so that

$$I = \int_0^\infty e^{-ax}J_0(bx)dx = \frac{1}{\sqrt{a^2 + b^2}} \qquad a > 0 \qquad (102.1)$$

The way in which contour integral representations of functions are found is largely characteristic of the particular function under consideration. Such representations will be developed when required in subsequent sections, but an example given here will serve to illustrate something of the general problem. Let it be required to find a contour integral representation of the gamma function $\Gamma(t)$, where t may be complex.

The gamma function of a positive real variable has already been defined in terms of the definite integral [see eq. (40.3)].

$$\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\,dx \qquad (102.2)$$

so we shall start from this point. The considerations of Sec. 101 lead one to try the integral

$$\int_C (-z)^{t-1}e^{-z}\,dz$$

where C is the contour shown in Fig. 102.2. When t is not an integer, $z = 0$ is a branch point and in order to keep $(-z)^{t-1}$ single-valued $-\pi \leq \arg(-z) < \pi$; consequently the positive real axis should not

be crossed by the path C. Following the procedure in Sec. 101 and letting $\delta \to 0$ and $\rho \to \infty$,

$$\int_C (-z)^{t-1}e^{-z}\,dz = -2i \sin \pi t \int_0^\infty x^{t-1}e^{-x}\,dx$$

The path C can be deformed in any way so long as it starts at infinity on the positive real axis and goes around the point $z = 0$ in the positive direction and then to infinity on the under side of the real axis. This type of path is sometimes referred to as a Hankel contour and indicated by $\int_\infty^{(0+)}$. The contour integral representation

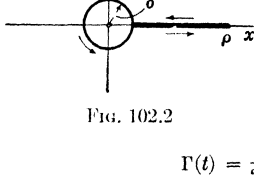


FIG. 102.2

$$\Gamma(t) = \frac{-1}{2i \sin \pi t} \int_\infty^{(0+)} (-z)^{t-1}e^{-z}\,dz \qquad (102.3)$$

has been established for $t > 0$ and not an integer. This expression, however, serves to define $\Gamma(t)$ for all real and complex values of t except $t = 0, \pm 1, \pm 2, \pm 3, \pm \ldots$

SUGGESTED REFERENCES

G. N. Watson, *Complex Integration and Cauchy's Theorem*, London, Cambridge University Press, 1914.

E. T. Whittaker and G. N. Watson, *Modern Analysis*, London, Cambridge University Press, 1920.

A. Hurwitz and R. Courant, *Allgemeine Funktionentheorie und elliptische Funktionen*, Berlin, Julius Springer, 1925.

H. Jeffreys and Bertha S. Jeffreys, *Methods of Mathematical Physics*, London, Cambridge University Press, 1946.

PROBLEMS

Evaluate the following integrals.

95.1. $\displaystyle\int_0^{2\pi} \frac{d\theta}{1 - 2a\cos\theta + a^2}$.

95.2. $\displaystyle\int_0^{2\pi} \frac{\cos^2 3\theta\,d\theta}{1 - 2a\cos\theta + a^2}$.

95.3. $\displaystyle\int_0^{2\pi} e^{\cos\theta}\begin{pmatrix}\cos\\ \sin\end{pmatrix}(n\theta - \sin\theta)d\theta$.

95.4. $\displaystyle\int_0^{2\pi} \frac{d\theta}{(a + b\cos^2\theta)^2}$.

96.1. $\displaystyle\int_{-\infty}^{\infty} \frac{dx}{(x^2 + 1)^3}$.

96.2. $\displaystyle\int_{-\infty}^{\infty} \frac{x^4\,dx}{(a + bx^2)^4}$.

96.3. $\int_{-\infty}^{\infty} \frac{dx}{(x^2+b^2)(x^2+a^2)^2}$.

97.1. $\int_0^{\infty} \frac{x \sin \lambda x}{x^2+k^2}\, dx$.

97.2. $\int_0^{\infty} \frac{x \sin \lambda x}{x^4+k^4}\, dx$.

97.3. $\int_0^{\infty} e^{a \cos \lambda x} \sin (a \sin \lambda x) \frac{x\, dx}{x^2+b^2}$.

99.1. $\int_0^{\infty} \frac{\sin ax}{e^{2\pi x}-1}\, dx$.

99.2. $\int_0^{\infty} \frac{\sin x}{\sinh x}\, dx$.

99.3. $\int_0^{\infty} \frac{x \cos ax}{\sinh x}\, dx$.

100.1. If an integral exists, its value is the same as its Cauchy principal value. Hence evaluate

$$\int_0^{\infty} \frac{\sin tx}{x}\, dx.$$

Evaluate the following integrals.

101.1. $\int_0^{\infty} \frac{dx}{x^p(1+x)}$.

Hint. Use contour of Fig. 100.1.

101.2. $P\int_0^{\infty} \frac{dx}{x^p(1-x)}$.

101.3. $\int_0^1 \frac{\log x\, dx}{1+x}$.

Hint. Expand $e^{p \ln x}$ in Taylor's series.

101.4. $\int_a^b \frac{dx}{(x-c)\sqrt{(x-a)(b-x)}}$

for a, b real $b > a$, c not on axis between a and b.

Chapter 8

REPRESENTATION OF FUNCTIONS BY INFINITE SERIES OF FUNCTIONS

103. General remarks. In Chapter 6, some of the general properties of infinite series of functions were considered, as well as the determination of the coefficients in power series and Laurent series, and tests to establish absolute and uniform convergence were given. In this chapter, certain practical aspects connected with the representation of functions in power series and series whose terms are powers of a given function are discussed. The problem of inverting such series is treated in some detail. Methods of determining the asymptotic expansion of functions, of so much practical importance, are given. The representation of functions by series of orthonormal functions is deferred until Chapter 12.

104. Laurent expansion of cot z and Bernoulli numbers. A typical example of a simple Laurent expansion is furnished by that for $\frac{1}{2}\cot z/2$. The coefficients in the Laurent series may be found by the application of formulas (91.1) and (91.2), which for the function under consideration become

$$c_{-n} = \frac{1}{2\pi i}\int_{C_1} \frac{\frac{1}{2}\cot(z/2)dz}{z^{-n+1}} \tag{104.1}$$

$$c_{+n} = \frac{1}{2\pi i}\int_{C_2} \frac{\frac{1}{2}\cot(z/2)dz}{z^{n+1}} \tag{104.2}$$

where the contours C_1 and C_2 encircle the point $z = 0$. Since for $n = 2, 3, \ldots$ the integrand of (104.1) is analytic, the contour integral is zero from Cauchy's theorem. Consequently, one term having an inverse power of z occurs in the Laurent expansion. The coefficient c_{-1} is easily found by the method of residues, so that

$$c_{-1} = \frac{1}{2\pi i}\int_C \frac{1}{2}\cot\frac{z}{2}\,dz = \frac{1}{2}\lim_{z\to 0} z\cot\frac{z}{2} = 1$$

The integrand for c_0 has a pole of order 2, so that the method of residue gives

$$c_0 = \frac{1}{2\pi i}\int_{C_2} \frac{\frac{1}{2}\cot(z/2)dz}{z} = \frac{1}{2}\left[\frac{d}{dz}\left(z^2\cot\frac{z}{2}\right)\right]_{z=0} = 0$$

The coefficients for positive values of n can be found most easily when it is noted that the integral (104.2), in accordance with (87.2), is

$$c_n = \frac{1}{2\pi i}\int_{C_2} \frac{\frac{1}{2}z\cot(z/2)dz}{z^{n+1}} = \frac{1}{(n+1)!}\left[\frac{d^{n+1}}{dz^{n+1}}\left(\frac{1}{2}z\cot\frac{z}{2}\right)\right]_{z=0} \tag{104.3}$$

This is just the coefficient in the $(n+1)$st term of the Taylor's expansion of $\frac{1}{2}z\cot z/2$. Since this is an even function, only those derivatives for which $n+1$ is even give any value at $z=0$. Hence $c_{2n-1} = [-1/(2n)!]B_n$, and

$$\frac{1}{2}\cot\frac{z}{2} = \frac{1}{z} - \frac{B_1}{2!}z - \frac{B_2}{4!}z^3 - \frac{B_3 z^5}{6!} - \ldots \tag{104.4}$$

where

$$B_n = -\left[\frac{d^{2n}}{dz^{2n}}\left(\frac{1}{2}z\cot\frac{z}{2}\right)\right]_{z=0} \tag{104.5}$$

The numerical coefficients B_n are called Bernoulli numbers. They occur in many such expansions. The numerical values of the first few are $B_1 = \frac{1}{6}$, $B_2 = \frac{1}{30}$, $B_3 = \frac{1}{42}$, $B_4 = \frac{1}{30}$, $B_5 = \frac{5}{66}$, $B_6 = \frac{691}{2730}$, $B_7 = \frac{7}{6}$, etc. These numbers were tabulated by Adams in the *British Association Reports* for 1877.

105. Expansion of a function in powers of another function. It is often advantageous to expand an analytic function $f(z)$ in powers of another analytic function $w(z)$ so that

$$f(z) = k_0 + k_1 w + k_2 w^2 + k_3 w^3 + \ldots \tag{105.1}$$

The practical problem at hand is the determination of the coefficients $k_1, k_2, \ldots$. It will be supposed for the present that the function $w(z)$ may be represented by the power series

$$w(z) = c_1 z + c_2 z^2 + c_3 z^3 + \ldots \tag{105.2}$$

whose circle of convergence in the z plane is C'. Let C be a circle within C' so that $w(z)$ and $f(z)$ are analytic for all points on C as well as within it. It will further be supposed that $w(z)$ has no zeros on C.

The function $f(z)$ may be expressed in terms of the contour integral

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(t)w'(t)}{w(t) - w(z)}\,dt \tag{105.3}$$

where $w'(t) = dw/dt$. This may be verified by applying the theory of residues (see Sec. 93). The residue at $t = z$ is

$$\lim_{t\to z}\frac{(t-z)f(t)w'(t)}{w(t)-w(z)}$$
$$= \lim_{t\to z}\frac{(t-z)f(t)w'(t)}{w(z) + w'(z)(t-z) + [w''(z)(t-z)^2]/2! + \ldots - w(z)}$$
$$= f(z)$$

which justifies (105.3).

Since $w(t)$ has no zeros on C, a value of z inside C can always be chosen so that $\left|\frac{w(z)}{w(t)}\right| < 1$, when t is on C, and we may write

$$\int_C \frac{f(t)w'(t)}{w(t) - w(z)}\,dt = \int_C \frac{f(t)w'(t)}{w(t)}\left[1 + \frac{w(z)}{w(t)} + \frac{w^2(z)}{w^2(t)} + \ldots\right]dt$$

so that

$$f(z) = k_0 + k_1 w(z) + k_2 w^2(z) + k_3 w^3(z) + \ldots$$

where

$$k_n = \frac{1}{2\pi i}\int_C \frac{f(t)w'(t)dt}{w^{n+1}(t)} \tag{105.4}$$

Referring to (105.2), it is readily seen that the integrand has a pole of order $n + 1$ at $t = 0$. The integral may be evaluated by the method of residues (see Sec. 94). Formula (94.2) gives for the residue of the integrand at $t = 0$, the value

$$\lim_{t\to 0} \frac{1}{n!}\cdot\frac{d^n}{dt^n}\left[\frac{f(t)w'(t)t^{n+1}}{w^{n+1}(t)}\right]$$

so that

$$k_n = \frac{1}{n!}\left[\frac{d^n}{dz^n}\frac{f(z)w'(z)z^{n+1}}{w^{n+1}(z)}\right]_{z=0} \tag{105.5}$$

For $n > 1$ a simpler expression can be obtained by integrating (105.4) once by parts. Then

$$\begin{aligned} k_n &= \frac{1}{2\pi i}\int_C \frac{f(t)w'(t)}{w^{n+1}(t)}dt \\ &= -\frac{1}{2\pi i}\int_C \frac{f(t)}{n}\cdot\frac{d}{dt}\left[\frac{1}{w^n(t)}\right]dt \\ &= \frac{1}{2n\pi i}\int_C \frac{f'(t)}{w^n(t)}dt \end{aligned}$$

Applying the residue theorem again,

$$k_n = \frac{1}{n!}\left\{\frac{d^{n-1}}{dz^{n-1}}f'(z)\left[\frac{z}{w(z)}\right]^n\right\}_{z=0} \qquad n = 1, 2, 3, \ldots \tag{105.6}$$

From (105.5) or (105.1) directly, it is clear that $k_0 = f(0)$. The series

$$f(z) = f(0) + k_1 w + k_2 w^2 + k_3 w^3 + \ldots \tag{105.7}$$

is called a Burmann-Lagrange series.

If the function w vanishes only at a point $z = a$ and the series

$$w = c_1(z - a) + c_2(z - a)^2 + \ldots$$

is finite and converges on and within a circle C with center at a, and if $f(z)$ is analytic in the neighborhood of $z = a$, expressions (105.7) and (105.6) take the form

$$f(z) = f(a) + k_1 w + k_2 w^2 + k_3 w^3 + \dots \quad , \tag{105.8}$$

$$k_n = \frac{1}{n!}\left\{\frac{d^{n-1}}{dz^{n-1}} f'(z)\left[\frac{z-a}{w(z)}\right]^n\right\}_{z=a} \qquad n = 1, 2, \dots \tag{105.9}$$

As an illustration of the above method let it be desired to expand $e^{bz} = f(z)$ in powers of $w = ze^{-z}$. In the first place $f(0) = 1$. Then

$$k_n = \frac{1}{n!}\left[\frac{d^{n-1}}{dz^{n-1}} be^{bz}\left(\frac{z}{ze^{-z}}\right)^n\right]_{z=0} = \frac{b}{n!}\left[\frac{d^{n-1}}{dz^{n-1}} e^{(b+n)z}\right]_{z=0}$$

$$k_n = \frac{b(b+n)^{n-1}}{n!}$$

Thus

$$e^{bz} = b\sum_{n=0}^{\infty} \frac{(b+n)^{n-1}}{n!} z^n e^{-nz} \tag{105.10}$$

As another example, let $f(z) = \sinh^2 z$ be expanded in powers of $w = ze^{az}$. In this case $f(0) = 0$, and

$$\begin{aligned} k_n &= \frac{1}{n!}\left(\frac{d^{n-1}}{dz^{n-1}} 2 \sinh z \cosh z\, e^{-anz}\right)_{z=0} \\ &= \frac{1}{n!}\left(\frac{d^{n-1}}{dz^{n-1}} \sinh 2z\, e^{-anz}\right)_{z=0} \\ &= \frac{(-1)^{n-1}}{2n!}[(an-2)^{n-1} - (an+2)^{n-1}] \end{aligned}$$

$$\sinh^2 z = \frac{1}{2}\sum_{n=0}^{\infty} \frac{(-1)^{n-1}}{n!}[(an-2)^{n-1} - (an+2)^{n-1}] z^n e^{-naz} \tag{105.11}$$

106. Inversion of a power series. It frequently happens that one has a function $w(z)$ expressed as a power series in z, so that

$$w = c_1 z + c_2 z^2 + c_3 z^3 + \dots$$

when it would be more useful to have z expressed as a power series in w. The process of doing this is called inverting the series. The problem of obtaining the coefficients

$$z = k_1 w + k_2 w^2 + k_3 w^3 + \dots \tag{106.1}$$

is merely a special case of the more general problem treated in the preceding section in which $f(z)$ is z itself. Thus the Burmann-Lagrange series (105.7) reduces to (106.1), and formula (105.6) for the coefficients becomes

$$k_n = \frac{1}{n!}\left(\frac{d^{n-1}}{dz^{n-1}} \cdot \frac{z^n}{w^n}\right)_{z=0} \tag{106.2}$$

since $f'(z) = dz/dz = 1$. In case

$$w(z) = c_1(z-a) + c_2(z-a)^2 + \dots,$$

then z, in the neighborhood of a, may be expressed as a power series in w by making use of the relations (105.8) and (105.9) with $f(z) = z$.

The inversion process may be illustrated by inverting the series

$$w = ze^{-az} = \sum_{p=0}^{\infty} \frac{(-a)^p}{p!} z^{p+1}$$

Applying formula (106.2),

$$k_n = \frac{1}{n!}\left(\frac{d^{n-1}}{dz^{n-1}} e^{anz}\right)_{z=0} = \frac{1}{n!}(an)^{n-1} \quad n = 1, 2, 3, \ldots$$

$$z = \sum_{n=1}^{\infty} \frac{(an)^{n-1}}{n!} w^n = \sum_{n=1}^{\infty} \frac{(an)^{n-1}}{n!} z^n e^{-anz}$$

If $a = 1$ and $z = 1$, we have the interesting expansion

$$1 = \sum_{n=1}^{\infty} \frac{n^{n-1}}{n!} e^{-n}$$

The above process of inversion often proves useful by providing a power series solution of transcendental equations. As an example, suppose we wish to express z in the neighborhood of a as a function of w when they are related by the transcendental Kepler equation

$$w = \frac{z - a}{\sin z} = c_1(z - a) + c_2(z - a)^2 + \ldots \tag{106.3}$$

This equation is encountered in theoretical astronomy.

In this case, formulas (105.8) and (105.9) must be used. These become

$$k_n = \frac{1}{n!}\left(\frac{d^{n-1}}{dz^{n-1}} \sin^n z\right)_{z=a} = \frac{1}{n!} \cdot \frac{d^{n-1}}{da^{n-1}} \sin^n a$$

and

$$z = a + \sin a\, w + \sin 2a\, w^2 + \frac{1}{3!}\left(\frac{d^2}{da^2} \sin^3 a\right) w^3 + \ldots$$

$$+ \frac{1}{n!}\left(\frac{d^{n-1}}{da^{n-1}} \sin^n a\right) w^n + \ldots$$

107. Expansion of functions with an infinite number of poles. A function $f(z)$ such as $1/\sin z$ which is analytic everywhere in the complex plane except at the isolated poles $a_1, a_2, a_3, \ldots$ can be expanded in a series of rational fractions. Let $f(z)$ have poles at the points $z = a_1, a_2, a_3, \ldots$, which may be placed in order so that $|a_1| \leq |a_2| \leq |a_3| \leq \ldots$. It will also be supposed that circles C_1,

C_2, C_3, with centers at the origin having increasing radii r_1, r_2, r_3, ... can be drawn so that no circle passes through one of the poles a_m, and that on these circles $|f(z)| < M$ independent of the circle considered.

The meaning of these conditions can be made clear by considering the example $1/\sin z$. This function has poles at $z = a_m = \pm m\pi$, $m = 0, 1, 2, 3, \ldots$ as shown in Fig. 107.1. The above conditions concerning the circles C_1, C_2, etc., can be satisfied if $r_m = (m + \frac{1}{2})\pi$. Furthermore $\left|\frac{1}{\sin z}\right| < M$ for all circles.

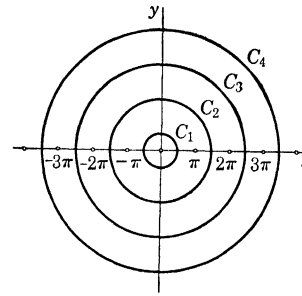


FIG. 107.1

The residue theorem (see Sec. 93) allows us to write

$$\frac{1}{2\pi i}\int_{C_m}\frac{f(z)}{z-t}\,dz = f(t) + \sum_{k=0}^{m} R_{a_k}\left[\frac{f(z)}{z-t}\right] \tag{107.1}$$

where t is any point in the complex plane inside C_m except one of the poles. The right-hand term is the sum of all the residues of $f(z)/(z-t)$ at the poles of $f(z)$ inside C_m. In the neighborhood of $z = a_k$, the function $f(z)$ has the form

$$f(z) = \frac{c_{-1}}{z-a_k} + c_0 + c_1(z-a_k) + c_2(z-a_k)^2 + \ldots$$

and

$$R_{a_k}\left[\frac{f(z)}{z-t}\right] = \lim_{z\to a_k}(z-a_k)\frac{f(z)}{z-t} = \frac{c_{-1}}{a_k - t}$$

However, c_{-1} is just the residue of $f(z)$ at $z = a_k$, so that (107.1) may be written

$$\frac{1}{2\pi i}\int_{C_m}\frac{f(z)}{z-t}\,dz = f(t) + \sum_{k=1}^{m}\frac{R_{a_k}[f(z)]}{a_k - t} \tag{107.2}$$

The integral on the left may be eliminated by writing

$$\frac{1}{2\pi i}\int_{C_m}\frac{f(z)}{z-t}\,dz = \frac{1}{2\pi i}\int_{C_m} f(z)\frac{dz}{z} + \frac{1}{2\pi i}\int_{C_m}\frac{f(z)t}{z(z-t)}\,dz \tag{107.3}$$

Applying formula (107.2) to the first integral on the right gives

$$\frac{1}{2\pi i}\int_{C_m} f(z)\frac{dz}{z} = f(0) + \sum_{k=1}^{m}\frac{R_{a_k}[f(z)]}{a_k}$$

provided $f(z)$ is analytic at the origin. Combining this expression with (107.3) and (107.2),

$$f(t) = f(0) + \sum_{k=1}^{m} R_{a_k}[f(z)]\left(\frac{1}{a_k} + \frac{1}{t - a_k}\right) + \frac{t}{2\pi i}\int_{C_m} \frac{f(z)}{z(z - t)}\,dz \tag{107.4}$$

If the $$\lim_{r_m \to \infty} \int_{C_m} \frac{f(z)}{z(z - t)}\,dz = 0$$

we have the desired expansion. Since

$$\left|\int_{C_m} \frac{f(z)}{z(z - t)}\,dz\right| \leq \int_{C_m} \frac{|f(z)|}{|z(z - t)|}\,|dz| \leq M\int_0^{2\pi} \frac{d\varphi}{|r_m| - |t|}$$

it is clear that the limit is zero.

Consequently if $f(z)$ is an analytic function everywhere in the complex plane including $z = 0$ except at the isolated poles $a_1, a_2, a_3, \ldots$, and $|f(z)| < M$ on all circles C_m, then

$$f(t) = f(0) + \sum_{k=1}^{\infty} R_{a_k}[f(z)]\frac{t}{a_k(t - a_k)} \tag{107.5}$$

The sum extends over all the residues in the complex plane, which may be finite or infinite in number.

This type of expansion will be illustrated by the example $f(z) = \csc z$ mentioned above. Csc z satisfies all the above conditions except it is not analytic at the origin. This can be taken care of by writing $f(z) = \csc z - 1/z$. Here $a_k = \pm k\pi$ and $f(0) = 0$. Furthermore

$$\begin{aligned} R_{a_k}[f(z)] &= \lim_{z \to k\pi}\left(\frac{z - k\pi}{\sin z} - \frac{z - k\pi}{z}\right) \\ &= \lim_{\xi \to 0}\left[\frac{\xi}{\sin(\xi + k\pi)} - \frac{\xi}{\xi + k\pi}\right] \\ &= \lim_{\xi \to 0}\frac{\xi}{\sin(\xi + k\pi)} = (-1)^k \end{aligned}$$

When these expressions are used in (107.5),

$$\csc z = \frac{1}{z} + \sum_{k=-\infty}^{\infty}{}' (-1)^k \frac{z}{k\pi(z - k\pi)} \tag{107.6}$$

The prime over the summation indicates that $k = 0$ is omitted,

108. Infinite product representation of a function. It is possible to represent a certain class of functions by an infinite product. Aside from the practical use of such representations, they often exhibit more clearly the properties of the functions so represented. It will be supposed that $f(z)$ is analytic everywhere in the complex plane and has simple zeros at the points

$$a_1, a_2, a_3, \ldots, \quad \text{where} \quad \lim_{n\to\infty} |a_n| = \infty$$

The function

$$\frac{f'(z)}{f(z)} = \frac{f'(a_k) + f''(a_k)(z - a_k) + \ldots}{f'(a_k)(z - a_k) + [f''(a_k)(z - a_k)^2]/2! + \ldots}$$

will possess poles only at the points $a_1, a_2, a_3, \ldots$. If it satisfies the condition stated in the preceding section, formula (107.5) gives the expansion

$$\frac{f'(z)}{f(z)} = \frac{f'(0)}{f(0)} + \sum_{k=1}^{\infty} R_{a_k}\left[\frac{f'(z)}{f(z)}\right]\left(\frac{1}{z - a_k} + \frac{1}{a_k}\right)$$

Furthermore, the residue

$$R_{a_k}\left[\frac{f'(z)}{f(z)}\right] = \lim_{z\to a_k}(z - a_k)\frac{f'(z)}{f(z)} = 1$$

When this result is used and the above expansion is integrated from 0 to z, term by term,

$$\log f(z) = \log f(0) + \frac{f'(0)}{f(0)}z + \sum_{k=1}^{\infty}\left\{\log(z - a_k) - \log(-a_k) + \frac{z}{a_k}\right\}$$

or

$$f(z) = f(0)e^{f'(0)z/f(0) + \sum_{k=1}^{\infty}\log\left(1 - \frac{z}{a_k}\right) + \frac{z}{a_k}}$$

$$= f(0)e^{f'(0)z/f(0)}\prod_{k=1}^{\infty}\left[\left(1 - \frac{z}{a_k}\right)e^{z/a_k}\right] \tag{108.1}$$

This infinite product representation brings out clearly the zeros of the function.

As an example, we shall give the infinite product representation of the function $(\sin nx)/x$. In the first place $f(0) = n$ and $f'(0) = 0$. The zeros occur at

$$nx = k\pi, \; k = \pm 1, \pm 2, \pm 3, \ldots \quad \text{so that} \quad a_k = \frac{k\pi}{n}$$

Formula (108.1) then gives

$$\frac{\sin nx}{x} = n\prod_{k=-\infty}^{\infty}{}' \left(1 - \frac{nx}{k\pi}\right)e^{nx/k\pi}$$

$$= n\prod_{k=1}^{\infty}\left(1 - \frac{nx}{k\pi}\right)e^{nx/k\pi}\left(1 + \frac{nx}{k\pi}\right)e^{-nx/k\pi}$$

$$= n\prod_{k=1}^{\infty}\left(1 - \frac{n^2x^2}{k^2\pi^2}\right) \tag{108.2}$$

The special formula when $n = 1$ will be required later, i.e.,

$$\frac{\sin z}{z} = \prod_{k=1}^{\infty} \left(1 - \frac{z^2}{k^2\pi^2}\right) \tag{108.3}$$

109. Infinite product representation of the gamma function. Certain properties of the gamma functions follow easily from the infinite product representation. Since the poles of the gamma function have not been given, it is not possible to find the product representation directly from (108.1). However, we may proceed from the relation

$$\Gamma(z+1) = z\Gamma(z) \tag{109.1}$$

derived in Sec. 40. When this relation is applied repeatedly to $\Gamma(z+n+1)$, we have

$$\begin{aligned}\Gamma(z+n+1) &= (z+n)\Gamma(z+n) \\ &= (z+n)(z+n-1)\ldots z\Gamma(z)\end{aligned}$$

By taking the logarithm of both sides,

$$\log \Gamma(z+n+1) = \sum_{k=0}^{n} \log (z+k) + \log \Gamma(z)$$

On taking the second derivative of both sides with respect to z, the functional equation

$$\frac{d^2 \log \Gamma(z)}{dz^2} = \sum_{k=0}^{n} \frac{1}{(z+k)^2} + \frac{d^2}{dz^2} \log \Gamma(z+n+1)$$

satisfied by $d^2 \log \Gamma(z)/dz^2$ is obtained. It is readily seen that the function $f(z) = \sum_{k=0}^{\infty} 1/(z+k)^2$ satisfies exactly the same functional equation, since

$$\begin{aligned}f(z) &= \sum_{k=0}^{n} \frac{1}{(z+k)^2} + \sum_{k=0}^{\infty} \frac{1}{(z+n+1+k)^2} \\ &= \sum_{k=0}^{n} \frac{1}{(z+k)^2} + f(z+n+1)\end{aligned}$$

Consequently,

$$\frac{d^2 \log \Gamma(z)}{dz^2} = f(z) = \sum_{k=0}^{\infty} \frac{1}{(z+k)^2} \tag{109.2}$$

We are now in a position to obtain the product representation of $\Gamma(z)$ by integrating the above relation from 1 to z, which yields

$$\begin{aligned}\frac{\Gamma'(z)}{\Gamma(z)} - \frac{\Gamma'(1)}{\Gamma(1)} &= \sum_{k=0}^{\infty} \frac{1}{1+k} - \frac{1}{z+k} \\ &= -\frac{1}{z} + \sum_{k=1}^{\infty} \left(\frac{1}{k} - \frac{1}{z+k}\right)\end{aligned}$$

or
$$\frac{z\Gamma'(z) + \Gamma(z)}{z\Gamma(z)} = \frac{\Gamma'(1)}{\Gamma(1)} + \sum_{k=1}^{\infty}\left(\frac{1}{k} - \frac{1}{z+k}\right)$$

When the relation (109.1) is used, the left side is readily seen to be $\frac{d}{dz}\log\Gamma(z+1)$, whereupon the equation can be integrated again from 0 to z, giving the result

$$\log\Gamma(z+1) = \frac{\Gamma'(1)}{\Gamma(1)}z + \sum_{k=1}^{\infty}\frac{z}{k} - \log\left(1+\frac{z}{k}\right) \tag{109.3}$$

The value of $\Gamma(1)$ is of course 1, and the value of $\Gamma'(1)$ can be found by placing $z = 1$ in eq. (109.3), so that

$$-\Gamma'(1) = \sum_{k=1}^{\infty}\frac{1}{k} - \log\left(1+\frac{1}{k}\right) = C = 0.57721 \tag{109.3a}$$

The constant C is referred to as Euler's constant. From (109.3) the product representation immediately follows.

$$\Gamma(z) = \frac{e^{-Cz}}{z}\prod_{k=1}^{\infty}\frac{e^{z/k}}{(1+z/k)} \tag{109.4}$$

This form shows clearly that $\Gamma(z)$ is analytic at all points of the complex plane except at the infinite number of poles at $z = 0, -1, -2, -3, \ldots$. Furthermore, it does not vanish anywhere, so that its reciprocal

$$\frac{1}{\Gamma(z)} = ze^{Cz}\prod_{k=1}^{\infty}\frac{1+z}{k}e^{-z/k} \tag{109.5}$$

is an analytic function everywhere in the complex plane, having zeros at $z = 0, -1, -2, -3, \ldots$.

Several very useful relations satisfied by the gamma function can be obtained from the above representation. As an example, (109.4) gives

$$\Gamma(z)\Gamma(-z) = -\frac{1}{z^2}\prod_{k=1}^{\infty}\frac{1}{(1-z^2/k^2)} \tag{109.6}$$

But from (108.3),

$$\prod_{k=1}^{\infty}\left(\frac{1}{1-z^2/k^2}\right) = \frac{\pi z}{\sin \pi z}$$

so that
$$\Gamma(z)\Gamma(-z) = \frac{-\pi}{z\sin \pi z} \tag{109.6a}$$

or using (109.1),

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z} \tag{109.7}$$

Another very useful formula can be derived from the product representation, namely, the multiplication formula of Gauss and Legendre.

$$\Gamma(z)\Gamma(z+1/n)\Gamma(z+2/n)\ldots\Gamma[z+(n-1)/n] = (2\pi)^{(n-1)/2}n^{1/2-nz}\Gamma(nz) \tag{109.8}$$

An important special case of this formula gives a formula of great use in converting gamma functions of half integer values of the argument. This is obtained by letting $n = 2$. Then

$$\Gamma(z)\Gamma\left(z+\frac{1}{2}\right) = \sqrt{2\pi}2^{1/2-2z}\Gamma(2z)$$
$$= \sqrt{\pi}2^{1-2z}\Gamma(2z) \tag{109.9}$$

110. Properties of asymptotic expansions. When the result of calculations is expressed in terms of complicated functions or integrals which cannot be evaluated in terms of ordinary functions, it may suffice to have an asymptotic form of the result valid for large values of the variable or some parameter. In certain cases an asymptotic form may be a series of inverse powers of the large quantity, in which case one speaks of the asymptotic expansion of the original function. A simple example will serve to illustrate the nature of such expansions. Let $f(x)$ be defined by the integral

$$f(x) = \int_x^\infty t^{-1}e^{x-t}\,dt \tag{110.1}$$

where x is real and positive. After performing n integrations by parts, $f(x)$ may be written

$$f(x) = \frac{1}{x} - \frac{1}{x^2} + \frac{2!}{x^3} - \frac{3!}{x^4} + \ldots + \frac{(-1)^{n-1}(n-1)!}{x^n} + (-1)^n n!\int_x^\infty \frac{e^{x-t}\,dt}{t^{n+1}}$$

It is instructive to inquire into the convergence of the series $\sum_{m=0}^{\infty} u_m$ where

$$u_{m-1} = \frac{(-1)^{m-1}(m-1)!}{x^m}$$

If the ratio test (see Sec. 76) for convergence is applied,

$$\lim_{m\to\infty}\frac{|u_m|}{|u_{m-1}|} = \lim_{m\to\infty}\frac{m}{x} \to \infty \quad \text{for all } x$$

This shows that no value of x exists for which the series converges. However, even though this series diverges for all values of x, a certain number of terms can nevertheless be used to compute the value of $f(x)$ very accurately for sufficiently large values of x. For this reason such an asymptotic series is sometimes called a semiconvergent series. In

order to see how this happens, let

$$S_n = \sum_{m=0}^{n} u_m$$

be the sum of $n + 1$ terms of the series, and consider the difference

$$f(x) - S_n(x) = (-1)^{n+1}(n+1)!\int_x^\infty \frac{e^{x-t}}{t^{n+2}}\,dt$$

The right-hand side of this equation provides a measure of the accuracy with which $f(x)$ can be computed by any choice of x and n. An estimate of this can be found by writing

$$\begin{aligned} |f(x) - S_n(x)| &\leq (n+1)!\int_x^\infty \frac{|e^{x-t}|}{|t^{n+2}|}\,dt \\ &\leq (n+1)!\int_x^\infty \frac{dt}{t^{n+2}} = \frac{n!}{x^{n+1}} \end{aligned} \tag{110.2}$$

Thus the error in computing $f(x)$ by means of S_n is less than the term on the right. It is important to notice that for a given value of x there is a value of $n = N$ for which $n!/x^{n+1}$ is a minimum. Consequently $N!/x^{N+1}$ is a measure of the ultimate accuracy with which $f(x)$ can be computed by means of the semiconvergent series.

It is of interest to determine the value of N. This can conveniently be done approximately by noting that $n!/x^{n+1} < n^{n-1}/x^{n+1}$ and finding the value of n which gives the minimum value of n^{n-1}/x^{n+1} in the usual manner. This leads to the relation

$$n\left(1 + \log\frac{n}{x}\right) = 1 \tag{110.3}$$

An approximate solution of this equation is

$$N = xe^{e/x-1} \tag{110.4}$$

For the case where $x = 10$, $N = 5$, so that the greatest accuracy would be obtained by using 5 terms in the series, and the error would be less than 0.00012.

The above example brings out certain of the general characteristics of asymptotic expansions, the formal definition of which is given below. A divergent series

$$a_0 + \frac{a_1}{z} + \frac{a_2}{z^2} + \ldots + \frac{a_n}{z^n} + \ldots$$

in which the sum of the first $n + 1$ terms is $S_n(z)$, is said to be an asymptotic expansion of $f(z)$ for a given range of values of the arg z, if

$$R_n(z) = z^n[f(z) - S_n(z)]$$

is such that $\lim_{|z|\to\infty} R_n(z) = 0$ for a fixed n, even though $\lim_{n\to\infty} |R_n(z)| \to \infty$

for a fixed z. Since $f(z)$ is never exactly given by any number of terms of its asymptotic expansion, one usually writes

$$f(z) \approx \sum_{n=0}^{\infty} \frac{a_n}{z^n}$$

It is readily verified that the series considered in the example above satisfies the formal definition of an asymptotic series, since for positive x,

$$R_n(x) = x^n[f(x) - S_n(x)] < \frac{n!}{x}$$

so that $$\lim_{|x|\to\infty} R_n(x) = 0 \quad \text{for } n \text{ fixed}$$

and $$\lim_{n\to\infty} R_n(x) = \infty \quad \text{for } x \text{ fixed.}$$

111. Calculations with asymptotic expansions. A few useful facts concerning manipulations of these expansions such as multiplication, integration, etc., are given below.

In the first place it should be noted that every function does not possess a unique asymptotic expansion. This means that several distinct functions might happen to have one and the same asymptotic expansion.

Multiplication. Given the asymptotic expansions of two functions $f(z)$ and $g(z)$,

$$f(z) \approx \sum_{n=0}^{\infty} a_n z^{-n}, \qquad g(z) \approx \sum_{n=0}^{\infty} b_n z^{-n}$$

valid for a common domain of values of z, the two series can be multiplied together in the same manner as ordinary series, and the result will be the asymptotic expansion of the function

$$f(z)g(z) \approx \sum_{n=0}^{\infty} c_n z^{-n}$$

where $$c_n = a_0 b_n + a_1 b_{n-1} + a_2 b_{n-2} + \ldots + a_n b_0$$

Integration and Differentiation. The asymptotic expansion of $f(z)$ can be integrated term by term over a range $x < z < \infty$, and the result will be the asymptotic expansion of $\int_x^{\infty} f(z)dz$.

It is not in general allowable to differentiate an asymptotic expansion term by term. Simple criteria as to when this is possible cannot be given. For certain theorems regarding the process of differentiation of such series the reader is referred to a paper by Ritt, *Bull. Am. Math. Soc.*, **24**, 225 (1918).

112. Methods of obtaining asymptotic expansions. There are several ways of obtaining asymptotic forms of a given function; the particular method to be used depends on the nature of the function and how many terms in the expansion are required. The most useful methods will be used on typical practical cases in order to exhibit the various aspects of the problem. One nearly always proceeds from an integral representation of the function in question. This may be either an ordinary definite integral or a contour integral over some path in the complex plane. One way of proceeding from an ordinary integral representation has already been illustrated in Sec. 110. The first part of the problem is to obtain an integral representation of the function.

As a first example, consider the function

$$f(x) = \sum_{n=0}^{\infty} \frac{(-x)^n}{\Gamma(n\alpha+1)} \qquad 0 < \alpha < 1 \tag{112.1}$$

defined by a series convergent for all finite values of x. In order to obtain an integral representation of $f(x)$, an integral representation of the reciprocal of the gamma function is needed. It is convenient to start with the representation

$$\Gamma(t) = \frac{-1}{2i \sin \pi t} \int_{\infty}^{(0+)} (-z)^{t-1} e^{-z}\, dz \tag{112.2}$$

given already in (102.3). This may be converted into a representation of $1/\Gamma(t)$ by making use of formula (109.7), whence it is seen that

$$\frac{1}{\Gamma(t)} = \frac{(\sin \pi t)\Gamma(1-t)}{\pi}$$

Consequently

$$\frac{1}{\Gamma(t)} = \frac{i}{2\pi} \int_{\infty}^{(0+)} (-z)^{-t} e^{-z}\, dz \tag{112.3}$$

When (112.3) is used in (112.1), the function $f(x)$ becomes

$$f(x) = \frac{i}{2\pi} \sum_{n=0}^{\infty} (-x)^n \int_{\infty}^{(0+)} (-z)^{-n\alpha-1} e^{-z}\, dz$$

Inverting the order of integration and summation,

$$f(x) = \frac{i}{2\pi} \int_{\infty}^{(0+)} \frac{e^{-z}}{(-z)} \sum_{n=0}^{\infty} \left[\frac{-x}{(-z)^{\alpha}}\right]^n dz$$

$$= \frac{i}{2\pi} \int_{\infty}^{(0+)} \frac{e^{-z}}{(-z)} \frac{1}{1 + x/(-z)^{\alpha}}\, dz$$

The series has been summed by noting that

$$(1+y)^{-1} = \sum_{n=0}^{\infty} (-y)^n$$

The above expression may be rewritten as

$$f(x) = \frac{i}{2\pi x}\int_{\infty}^{(0+)} \frac{e^{-z}}{(-z)^{1-\alpha}}\frac{dz}{(1+(-z)^{\alpha}/x)} \tag{112.4}$$

and the denominator may be expanded in a series of inverse powers of x, so that

$$\left[1+\frac{(-z)^{\alpha}}{x}\right]^{-1} = 1 - \frac{(-z)^{\alpha}}{x} + \left[\frac{(-z)^{\alpha}}{x}\right]^2 - \ldots + (-1)^n\left[\frac{(-z)^{\alpha}}{x}\right]^n + R_n(z,x)$$

When this is placed in (112.4) and integrated term by term, the resulting integrals are gamma functions according to (112.2).

$$f(x) = \frac{1}{\pi x}\left[\Gamma(\alpha)\sin\pi\alpha - \frac{\Gamma(2\alpha)\sin 2\pi\alpha}{x} + \frac{\Gamma(3\alpha)\sin 3\pi\alpha}{x^2} - \ldots + (-1)^{n-1}\frac{\Gamma(n\alpha)\sin n\pi\alpha}{x^{n-1}}\right] + G_n(x) \qquad 0<\alpha<1 \tag{112.5}$$

The number of terms required to compute $f(x)$ as accurately as possible for a given x can be determined by the method outlined in Sec. 110.

113. Asymptotic form of the gamma function. The asymptotic form for the gamma function will be found by two methods, each of which illustrates quite a different mode of attack. To illustrate the first method we start with formula (109.2).

$$\frac{d^2\log\Gamma(z)}{dz^2} = \sum_{k=0}^{\infty}\frac{1}{(z+k)^2} \tag{113.1}$$

The first task is to find an integral representation of the series on the right. In doing this a method will be followed which can be used quite generally for series of this type. The guiding thought here is to look for a contour integral of a function for which the sum of the residues is essentially the right-hand side of (113.1).

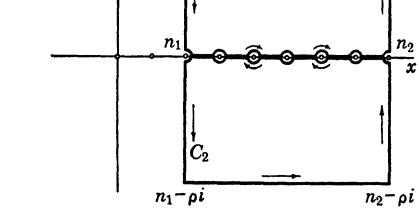


FIG. 113.1

We are led to try a contour integral of the form

$$\int_{C_1}\frac{f(t)dt}{e^{-2\pi it}-1}$$

for reasons which will be immediately apparent. If $f(t)$ is analytic, the integrand has poles at all

integer values of t. The contour C_1 to be selected is shown in Fig. 113.1; it traverses the indented rectangle with corners at n_1, n_2, $n_2 + \rho i$, $n_1 + \rho i$, where n_1 and n_2 are positive integers. The indentations occur at the poles of the integrand. As was shown in Sec. 100, the contribution to the integral over the semicircular indentations will be minus one-half the residues at the points in question. When the indentation is a quarter circle, the contribution will be one-fourth of the residue. With this in mind one may write

$$\int_{C_1} \frac{f(t)}{e^{-2\pi it} - 1}\, dt =$$

$$\int_{n_1}^{n_2} \frac{f(x)}{e^{-2\pi ix} - 1}\, dx + \frac{1}{4}f(n_1) + \frac{1}{2}f(n_1+1) + \ldots + \frac{1}{2}f(n_2-1) + \frac{1}{4}f(n_2)$$

$$+\int_0^{\rho} \frac{f(n_2+iy)i}{e^{+2\pi y} - 1}\, dy + \int_{n_2}^{n_1} \frac{f(x+\rho i)}{e^{-2\pi i(+\rho i)} - 1}\, dx + \int_{\rho}^{0} \frac{f(n_1+iy)i}{e^{+2\pi y} - 1}\, dy = 0$$

The contour integral must be zero, since C_1 contains no poles of the integrand. It will be assumed that $f(t)$ is such a function that

$$\lim_{\rho \to \infty} \int_{n_2}^{n_1} \frac{f(x+\rho i)}{e^{+2\pi\rho - 2\pi ix} - 1}\, dx$$

vanishes. Then

$$\frac{1}{4}f(n_1) + \frac{1}{2}\sum_{k=1}^{n_2-1-n_1} f(n_1+k) + \frac{1}{4}f(n_2) + \int_{n_1}^{n_2} \frac{f(x)}{e^{-2\pi ix} - 1}\, dx$$

$$+\int_0^{\infty} \frac{f(n_2+iy) - f(n_1+iy)}{e^{2\pi y} - 1} i\, dy = 0 \tag{113.2}$$

This result can be simplified by a similar expression found by integrating $f(t)/(e^{2\pi it} - 1)$ around the contour C_2, which is also indicated in Fig. 113.1. This gives

$$-\left[\frac{1}{4}f(n_1) + \frac{1}{2}\sum_{k=1}^{n_2-1-n_1} f(n_1+k) + \frac{1}{4}f(n_2)\right] + \int_{n_2}^{n_1} \frac{f(x)}{e^{2\pi ix} - 1}\, dx$$

$$+\int_0^{\infty} \frac{[f(n_2-iy) - f(n_1-iy)]\, i}{e^{2\pi y} - 1}\, dy = 0 \tag{113.3}$$

By subtracting (113.2) from (113.3),

$$\frac{1}{2}f(n_1) + \sum_{k=1}^{n_2-1-n_1} f(n_1+k) + \frac{1}{2}f(n_2)$$

$$= \int_{n_1}^{n_2} f(x)dx + i\int_0^{\infty} \frac{[f(n_2-iy) - f(n_1-iy) - f(n_2+iy) + f(n_1+iy)]}{e^{2\pi y} - 1} dy \tag{113.4}$$

The left-hand side of this equation can be made equal to the right-hand side of (113.1) by letting

$$f(t) = \frac{1}{(z+t)^2} \quad \text{and} \quad n_1 = 0$$

and passing to the limit $n_2 \to \infty$, and finally adding a term $\frac{1}{2}f(0)$. Consequently,

$$\frac{d^2 \log \Gamma(z)}{dz^2} = \sum_{k=0}^{\infty} \frac{1}{(z+k)^2} = +\frac{1}{2z^2} + \int_0^\infty \frac{dx}{(z+x)^2} + i\int_0^\infty \left[\frac{1}{(z+iy)^2} - \frac{1}{(z-iy)^2}\right] \frac{dy}{e^{2\pi y} - 1}$$

$$= \frac{1}{2z^2} + \frac{1}{z} + 4\int_0^\infty \frac{zy}{(z^2+y^2)^2} \frac{dy}{e^{2\pi y} - 1} \tag{113.5}$$

This expression may be integrated twice with respect to z, so that

$$\log \Gamma(z) = \left(z - \frac{1}{2}\right) \log z + (A-1)z + B + 2\int_0^\infty \frac{\tan^{-1}(y/z)dy}{e^{2\pi y} - 1} \tag{113.6}$$

where A and B are constants of integration, the evaluation of which we shall defer until later. The remaining integral may readily be expanded in a power series in $1/z$ by noting that

$$\tan^{-1}\frac{y}{z} = \frac{y}{z} - \frac{1}{3}\frac{y^3}{z^3} + \frac{1}{5}\frac{y^5}{z^5} + \ldots + \frac{(-1)^{n-1}}{2n-1}\left(\frac{y}{z}\right)^{2n-1} + \ldots$$

Thus

$$2\int_0^\infty \frac{\tan^{-1}(y/z)dy}{e^{2\pi y} - 1} = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}B_n}{2n}\left(\frac{1}{z}\right)^{2n-1}$$

where

$$\frac{B_n}{4n} = \int_0^\infty \frac{y^{2n-1}\,dy}{e^{2\pi y} - 1}$$

is an integral expression for the nth Bernouli number (see eq. (104.5) for another definition of B_n). Taking the exponential of eq. (113.6) and inserting the values $A = 0$; $B = \frac{1}{2}\log 2\pi$ as determined in the next few paragraphs, one has the asymptotic form for $\Gamma(z)$ for large positive values of z.

$$\Gamma(z) = \sqrt{2\pi}e^{-z}z^{z-1/2}\left(1 + \frac{1}{12z} + \frac{1}{288z^2} - \frac{139}{51{,}840z^3} + \ldots\right) \tag{113.7}$$

Another rather different method of approach starts from the definite integral representation of the gamma function (see Sec. 40),

$$\Gamma(x+1) = \int_0^\infty t^x e^{-t}\,dt$$

It is convenient to change the variable of integration to $t = xs$, so that

$$\Gamma(x+1) = x^{x+1}\int_0^\infty s^x e^{-xs}\,ds$$

$$= x^{x+1}e^{-x}\int_0^\infty e^{-x(s-1-\log s)}\,ds$$

$$= x^{x+1}e^{-x}\int_0^\infty e^{-xf(s)}\,ds \tag{113.8}$$

The purpose of the change in variable was to make the large quantity x multiply everything in the exponential and thus make the integrand a rapidly varying function of s. The function

$$e^{-xf(s)} = e^{-x(s-1-\log s)}$$

is plotted in Fig. 113.2 for several values of x. It is readily seen that this function has a maximum value of unity at $s = 1$ for all positive values of x. The larger the value of x, however, the more rapidly the value of the function falls to zero on either side of $s = 1$, and the greater

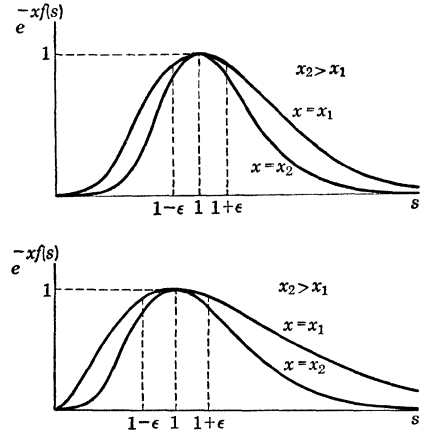


FIG. 113.2

the contribution to the value of the integral that occurs for values of s in a given interval 2ϵ about the point $s = 1$. Consequently, we write

$$\int_0^\infty e^{-xf(s)}\,ds = \int_0^{1-\epsilon} + \int_{1-\epsilon}^{1+\epsilon} + \int_{1+\epsilon}^\infty e^{-xf(s)}\,ds \qquad (113.9)$$

The aim here is to select a value of ϵ such that for the range of x to be used, the contributions of the first and last integrals can be neglected. It is consequently necessary to establish an upper limit for the contributions of these integrals. This may readily be done by assuming that $0 < \epsilon < \frac{1}{2}$. Then since

$$s - 1 - \log s = \int_s^1 \left(\frac{1}{\xi} - 1\right) d\xi \geq \int_s^1 (1 - \xi)\,d\xi = \frac{1}{2}(s-1)^2$$

$$\tfrac{1}{2} \leq s \leq 1 \qquad (113.10)$$

$$s - 1 - \log s = \int_s^1 \left(\frac{1}{\xi} - 1\right) d\xi \geq \int_1^s \frac{(\xi - 1)}{4}\, ds = \frac{1}{8}(s-1)^2$$

$$1 \leq s \leq 4 \qquad (113.11)$$

$$s - 1 - \log s \geq \frac{3s}{4} - \log s > \frac{1}{2}s \qquad 4 < s \qquad (113.12)$$

it is clear that

$$\begin{aligned}
&\int_0^{1-\epsilon} e^{-xf(s)}\, ds + \int_{1+\epsilon}^{\infty} e^{-xf(s)}\, ds \\
&\quad \leq \int_0^{1-\epsilon} e^{-(x/2)(s-1)^2}\, ds + \int_{1+\epsilon}^{4} e^{-(x/8)(s-1)^2}\, ds + \int_4^{\infty} e^{-xs/2}\, ds \\
&\quad \leq \int_0^{1-\epsilon} e^{-x\epsilon^2/2}\, ds + \int_{1+\epsilon}^{4} e^{-x\epsilon^2/8}\, ds + \frac{2}{x} e^{-2x} \\
&\quad \leq e^{-x\epsilon^2/2}(1-\epsilon) + e^{-x\epsilon^2/8}(3-\epsilon) + \frac{2}{x} e^{-2x} \\
&\quad < 5e^{-x\epsilon^2/8} \qquad x > 2
\end{aligned}$$

Consequently

$$\Gamma(x+1)e^x x^{-x-1} = \int_{1-\epsilon}^{1+\epsilon} e^{-xf(s)}\, ds + \delta \qquad (113.13)$$

where $\delta < 5e^{-x\epsilon^2/8}$. The evaluation of the above integral is accomplished by expanding $f(s)$ about the point $s = 1$, so that

$$f(s) = f(1) + f'(1)(s-1) + \frac{f''(1)}{2!}(s-1)^2 + g(s)(s-1)^3$$

and since $f(1)$ and $f'(1)$ are zero and $f''(1) = 1$,

$$f(s) = \frac{(s-1)^2}{2} + g(s)(s-1)^3$$

Since $|\, g(s)\,| \leq M \approx \frac{1}{2}$ in the interval $(1-\epsilon) \leq s \leq (1+\epsilon)$, we may write

$$\begin{aligned}
&\int_{1-\epsilon}^{1+\epsilon} e^{-xf(s)}\, ds \\
&\quad = \int_{1-\epsilon}^{1+\epsilon} e^{-(x/2)(s-1)^2} \left[1 - xg(s)(s-1)^3 + \frac{x^2 g^2(s)(s-1)^6}{2!} - \ldots\right] ds \\
&\quad = \int_{1-\epsilon}^{1+\epsilon} e^{-(x/2)(s-1)^2} ds \left[\begin{array}{c} 1 + \text{terms not greater than} \\ Mx\epsilon^3 + \dfrac{M^2x^2\epsilon^6}{2} + \ldots \end{array}\right]
\end{aligned}$$

Also

$$\begin{aligned}
&\int_{1-\epsilon}^{1+\epsilon} e^{-(x/2)(s-1)^2} ds \\
&\quad = 2\sqrt{\frac{2}{x}} \int_0^{\sqrt{x/2}\,\epsilon} e^{-v^2}\, dv = 2\sqrt{\frac{2}{x}} \left(\int_0^{\infty} e^{-v^2}\, dv - \int_{\sqrt{x/2}\,\epsilon}^{\infty} e^{-v^2}\, dv\right) \\
&\quad = \sqrt{\frac{2\pi}{x}} \left[\begin{array}{c} 1 + \text{terms not greater than} \\ \dfrac{2}{\sqrt{\pi}}\, \epsilon^{-x\epsilon^2/2}(1 + e^{-\sqrt{2x}\epsilon+1} + \ldots) \end{array}\right]
\end{aligned}$$

The additional terms on the right-hand side of this equation are less than or of the same order of magnitude as δ occurring in eq. (113.13). A value of ϵ should be chosen such that $\lim_{x\to\infty} \delta = \lim_{x\to\infty} 5^{-x\epsilon^2/8} = 0$ and at the same time $\lim_{x\to\infty} \epsilon^3 x = 0$, thereby making it possible to neglect such terms for sufficiently large values of x. A suitable choice of ϵ would be $\epsilon = x^{-2/5}$.

$$\lim_{x\to\infty} 5c^{-x^{1/5}/8} = 0 \quad \text{and} \quad \lim_{x\to\infty} \epsilon^3 x = \lim_{x\to\infty} x^{-1/5} = 0$$

Using the equations above in eq. (113.13) and noting that the dominant term in the terms to be neglected is $Mx^{-1/5}$, we obtain the asymptotic expression

$$\Gamma(x+1) \approx \sqrt{2\pi} e^{-x} x^{x+1/2} \,(1 + \text{a term of order } x^{-1/5}) \qquad (113.14)$$

Thus for very large values of x,

$$\Gamma(x+1) \approx \sqrt{2\pi} e^{-x} x^{x+1/2} \qquad (113.15)$$

For large enough values of x this formula agrees with the more accurate formula (113.7) already obtained.

The values of the constants A and B in (113.6) can be found by taking the logarithm of (113.15) and comparing this with (113.6) in the limit $x \to \infty$, where expression (113.15) becomes exactly the value of $\Gamma(x+1)$. It is readily seen that $A = 0$ and $B = \log \sqrt{2\pi}$.

114. Method of steepest descents or saddle point method. Whenever the function for which an asymptotic expansion is desired is expressed in the form of a contour integral, the method of steepest descents may furnish a very convenient method for obtaining the asymptotic form. The general features of the method may be illustrated by considering the contour integral

$$\int_C e^{xf(t)}\, dt \qquad (114.1)$$

The success of the method requires that the real part of $f(t)$, denoted by $Rf(t)$, become negatively infinite at both ends of the path of integration C. Assuming this to be the case, then when Rz is positive and very large there is only a very small contribution to the value of the integral near the ends of the path C. This contribution becomes progressively less as Rz increases in value.

The immediate object is to deform the path C so that there is very little contribution to the value of the integral over the entire path except in the immediate neighborhood of one or more points on the path where practically the entire contribution takes place. For purposes of illustration, suppose that $Rf(t)$ when plotted above the complex t plane is

the surface shown in Fig. 114.1a. In accomplishing the aim mentioned above, the first thing is obviously to deform the path C so that it passes through the point t_0 of Fig. 114.1b, for which $Rf(t)$ takes on its maximum value for an appropriate integration path. This occurs at P in Fig. 114.1a. The second thing is to orient the path C so that most of the contribution to the integral occurs in as confined a region in the neighborhood of t_0 as possible. This will be realized if the path C on each side of t_0 is chosen so that as we proceed on it away from t_0, the value

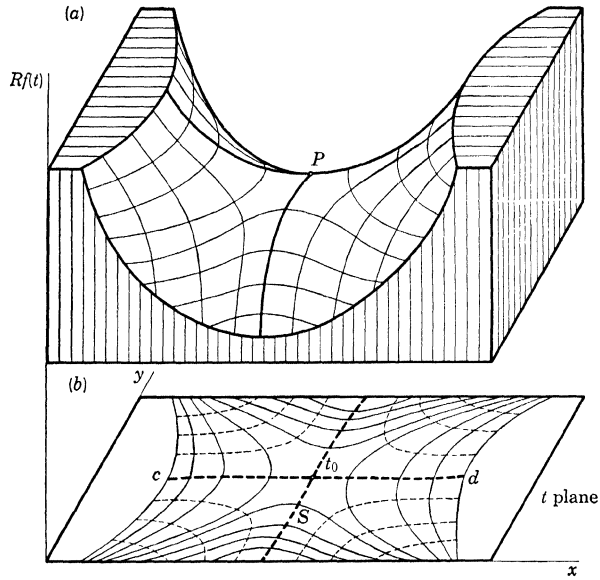


FIG. 114.1

of $Rf(t)$ decreases as fast as possible. The theory of functions of a complex variable provides a very convenient method of determining this path, which we shall call S. In the first place, if planes parallel to the xy plane are allowed to cut the surface $Rf(t)$, the intersecting curves shown in Fig. 114.1a when projected into the xy plane will give rise to the constant level lines shown in Fig. 114.1b. When t is allowed to traverse one of these curves, $Rf(t)$ = constant, so that conversely the equation $Rf(t)$ = *constant* determines the constant level lines in the xy plane. The path S of steepest descent will be one which passes through t_0 and intersects the constant level lines at right angles. The

equation of such paths is given by setting the imaginary part of $f(t)$ equal to a constant, thus $\mathcal{I}f(t) = \text{constant}$. This follows from the property of analytic functions, and can easily be shown by writing

$$f(t) = u(x,y) + iv(x,y)$$

where $u(x,y) = \text{constant}$ gives equation of constant level lines, and $v(x,y) = \text{constant}$ paths of steepest descent. The tangent to the constant levels lines is

$$\left(\frac{dy}{dx}\right)_u = -\frac{\partial u/\partial x}{\partial u/\partial y}$$

The tangent to the paths $v(x,y) = \text{constant}$ is

$$\left(\frac{dy}{dx}\right)_v = -\frac{\partial v/\partial x}{\partial v/\partial y} = \frac{\partial u/\partial y}{\partial u/\partial x} = -\frac{1}{(dy/dx)_u}$$

This is just the condition that the tangents at the points of intersection of the two sets of curves are at right angles. The intermediate step follows from the Cauchy-Riemann equations (82.3). Therefore in order to determine the path of steepest descent it is necessary only to set the imaginary part of $f(t)$ equal to that constant which makes the path go through the point t_0 for which $Rf(t)$ is a maximum.

It should be pointed out that the equation $\mathcal{I}f(t) = \text{constant}$ may have more than one solution, as it would have when the surface $Rf(t)$ has a saddle point as shown in Fig. 114.1a. Then it is important to choose that path for which $Rf(t)$ decreases as one proceeds away from the saddle point t_0. The point t_0 will be one for which

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} = 0 = \frac{df}{dt}$$

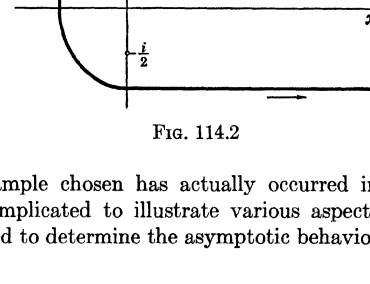


FIG. 114.2

Further features of this method may be brought out best by an example. The example chosen has actually occurred in practice* and is sufficiently complicated to illustrate various aspects of the method. Let it be required to determine the asymptotic behavior of the function

$$F(s) = \int_C e^{-s^{2i}} \left(\frac{t + i/2}{t - i/2}\right)^{k^{2i}/4} dt \qquad (114.2)$$

for large positive values of the parameter k. The path of integration C is shown in Fig. 114.2. The integral (114.2) can be written in the form

* Fritz Sauter, *Z. Physik*, **69**, 748 (1931).

$$F(s) = \int_C e^{h(t)}\, dt$$

which then resembles (114.1). Here

$$h(t) = -s^2 t + \frac{k^2 i}{4} \log \frac{t + i/2}{t - i/2} \tag{114.3}$$

In line with the above discussion, the saddle point t_0 is found by solving the equation

$$\frac{dh(t)}{dt} = -s^2 + \frac{k^2 i}{4}\left(\frac{-i}{t^2 + \frac{1}{4}}\right) = 0$$

There are two solutions of this equation, namely,

$$t_1 = \frac{1}{2}\sqrt{\frac{k^2}{s^2} - 1} \qquad t_2 = -\frac{1}{2}\sqrt{\frac{k^2}{s^2} - 1} \tag{114.4}$$

and two situations to consider. In the first place if $\left|\frac{k}{s}\right| > 1$, then t_1 and t_2 lie on the real axis, but if $\left|\frac{k}{s}\right| < 1$, they lie on the imaginary axis. We shall investigate the two cases separately.

Case 1. $k > s$. When $h(t)$ is separated into its real and imaginary parts,

$$Rh(t) = -s^2 x - \frac{k^2}{4}\left(\tan^{-1}\frac{y + \frac{1}{2}}{x} - \tan^{-1}\frac{y - \frac{1}{2}}{x}\right) \tag{114.5}$$

$$\mathcal{I}h(t) = -s^2 y + \frac{k^2}{8} \log \frac{x^2 + (y + \frac{1}{2})^2}{x^2 + (y - \frac{1}{2})^2} \tag{114.6}$$

It should be noted that at the ends of the path of integration $Rh(t) \to -\infty$ as required. In this case

$$t_1 = x_1 = \frac{1}{2}\sqrt{\frac{k^2}{s^2} - 1} \quad \text{and} \quad t_2 = -x_1$$

To see which or both of these points we require, we must examine $Rh(x_1)$ and $Rh(-x_1)$.

$$Rh(x_1) = -s^2 x_1 - \frac{k^2}{2}\tan^{-1}\frac{1}{2x_1}$$

$$Rh(-x_1) = s^2 x_1 + \frac{k^2}{2}\tan^{-1}\frac{1}{2x_1}$$

and the maximum value of $Rh(t)$ occurs for the point $t_2 = -x_1$, so that we should now find the path of steepest descent passing through t_2. As mentioned above, this path will be found from the equation

$$\mathcal{I}h(t) = \mathcal{I}h(t_2)$$

or using (114.6) and the value $\mathcal{I}h(t_2) = 0$,

$$-s^2 y + \frac{k^2}{8} \log \frac{x^2 + (y + \frac{1}{2})^2}{x^2 + (y - \frac{1}{2})^2} = 0 \tag{114.7}$$

This equation can be rearranged in the following convenient form.

$$e^{2wy} = \frac{x^2 + (y + \frac{1}{2})^2}{x^2 + (y - \frac{1}{2})^2} = 1 + \frac{2y}{x^2 + (y - \frac{1}{2})^2}$$

$$x^2 = \frac{2y}{e^{2wy} - 1} - \left(y - \frac{1}{2}\right)^2 = -\left(y^2 + \frac{1}{4}\right) + y \coth wy$$

Then
$$x = \pm\sqrt{y \coth wy - (y^2 + \tfrac{1}{4})} \tag{114.8}$$

where $w = 4s^2/k^2$. The path represented by eq. (114.8) is the closed path shown in Fig. 114.3. Another solution of eq. (114.7) is obviously the path $y = 0$, also indicated in Fig. 114.3. The arrows show the direction in which $Rf(t)$ is decreasing.

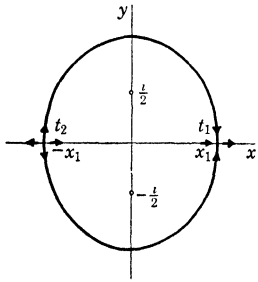


FIG. 114.3

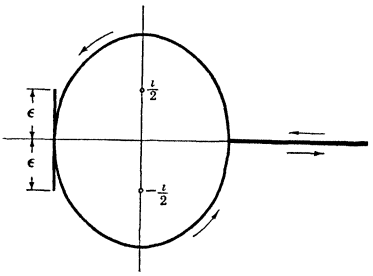


FIG. 114.4

The next step is to see whether or not the original path of integration C can be deformed so as to coincide with one of the paths shown in Fig. 114.4. The path C cannot be deformed so as to coincide with the path $y = 0$, but it can be made to coincide with the path of eq. 114.8, and part of the path $y = 0$. The deformed path is indicated in Fig. 114.4. Here the arrows indicate the direction the path is to be traversed.

In evaluating the integral it is customary to integrate over a segment 2ϵ of a straight-line path tangent to the path of steepest descent at the maximum point, in this case $-x_1$. The actual evaluation of the integral proceeds in much the same manner as outlined in Sec. 113. The function $h(t)$ is expanded about the point $t = t_2 = -x_1$, so that

$$h(t) = h(t_2) + h'(t_2)(t - t_2) + \frac{h''(t_2)}{2!}(t - t_2)^2 + \dots$$

$$= h(t_2) + \frac{2s^4}{k^2}\sqrt{\frac{k^2}{s^2} - 1}\,(t - t_2)^2 + \dots$$

where

$$h(t_2) = +\frac{1}{2}s^2\sqrt{\frac{k^2}{s^2} - 1} + \frac{k^2}{2}\tan^{-1}\frac{1}{\sqrt{k^2/s^2 - 1}}$$

$$= +\frac{1}{2}s^2\sqrt{\frac{k^2}{s^2} - 1} + \frac{k^2}{2}\sin^{-1}\frac{|s|}{k}$$

and

$$h'(t_2) = 0 \quad \text{and} \quad h''(t_2) = \frac{4s^4}{k^2}\sqrt{\frac{k^2}{s^2} - 1}$$

Therefore

$$\int_C e^{h(t)}\,dt \cong e^{+h(t_2)}\int_{-x_1+i\epsilon}^{-x_1-i\epsilon} e^{(2s^4/k^2)\sqrt{k^2/s^2-1}(t+x_1)^2}\,dt$$

$$\cong e^{h(t_2)}\,e^{3\pi i/2}\int_{-\epsilon}^{\epsilon} e^{-(2s^4/k^2)\sqrt{k^2/s^2-1}\xi^2}\,d\xi$$

$$\cong e^{h(t_2)-i\pi/2}\int_{-\infty}^{+\infty} e^{(-2s^4/k^2)\sqrt{k^2/s^2-1}\xi^2}\,d\xi$$

$$\cong e^{h(t_2)-\pi i/2}\,\frac{k\sqrt{\pi/2}}{s^2(k^2/s^2 - 1)^{1/4}}$$

An estimate of the error introduced by neglecting higher order terms in the expansion of $h(t)$ and by increasing the range of integration from ϵ to ∞ and $-\epsilon$ to $-\infty$ can be made as in Sec. 113. The error introduced by extending the path along the tangent will be greater than that introduced if the actual path of steepest descent had been followed.

We have arrived at the asymptotic form

$$F(s) = \sqrt{\pi/2}\,\frac{k e^{-i\pi/2+(s^2/2)\sqrt{k^2/s^2-1}+(k^2/2)\sin^{-1}|s|/k}}{s^2(k^2/s^2 - 1)^{1/4}} \quad \text{for} \quad \frac{k}{s} > 1 \tag{114.9}$$

Case 2. $\left|\frac{k}{s}\right| < 1$. In this case several interesting points arise, which make it instructive to carry out the details of finding the asymptotic form. Now the saddle points are located on the imaginary axis. From (114.4),

$$t_1 = \frac{i}{2}\sqrt{1 - \frac{k^2}{s^2}} = iy_1$$

$$t_2 = -\frac{i}{2}\sqrt{1 - \frac{k^2}{s^2}} = -iy_1$$

Reference to eq. (114.3) shows that $Rh(t_1) = Rh(t_2)$. This means that the two points t_1 and t_2 are of equal importance because the function $e^{-h(t)}$ has identical maxima. As for the preceding case, the paths of steepest descent are given by the solution of the equations

$$\mathcal{J}h(t) = \mathcal{J}h(t_1)$$

$$\mathcal{J}h(t) = \mathcal{J}h(t_2)$$

By using eq. (114.6), the two equations above may be written

$$-s^2y + \frac{k^2}{8}\log\frac{x^2 + (y + \frac{1}{2})^2}{x^2 + (y - \frac{1}{2})^2} = \mp s^2 y_1 + \frac{k^2}{8}\log\frac{(\pm y_1 + \frac{1}{2})^2}{(\pm y_1 - \frac{1}{2})^2}$$

This can be rearranged in the more convenient form,

$$x = \left\{\frac{2y}{[(\pm y_1 + \tfrac{1}{2})^2/(\pm y_1 - \tfrac{1}{2})^2]e^{2w(y \mp y_1)} - 1} - (y - \tfrac{1}{2})^2\right\}^{1/2} \tag{114.10}$$

This equation is satisfied by points on two curves, one being entirely above the x axis and the other being below, as shown in Fig. 114.5. Because $Rh(t)$ is a many-valued function of (x,y) it is necessary to exercise some care in the choice of the path of integration. To make the situation clear, the real part of $h(t)$ divided by s^2 may, from eq. (114.3), be written

$$\frac{Rh(t)}{s^2} = -\left(x + \frac{k^2}{4s^2}\varphi\right) \tag{114.11}$$

where

$$\varphi = \arg\frac{x + i(y + \frac{1}{2})}{x + i(y - \frac{1}{2})} = \theta_1 - \theta_2$$

$$\frac{x + i(y + \frac{1}{2})}{x + i(y - \frac{1}{2})} = \frac{R_1(x,y)e^{i\theta_1(x,y)}}{R_2(x,y)e^{i\theta_2(x,y)}} = Re^{i(\theta_1 - \theta_2)}$$

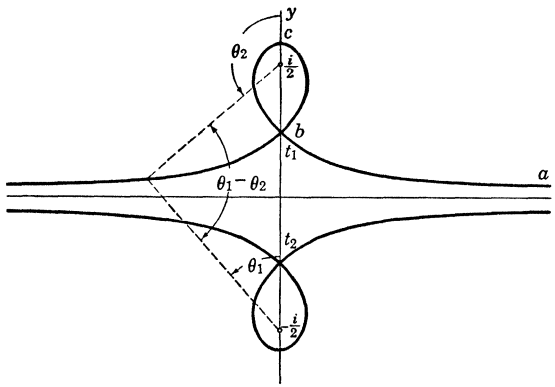


FIG. 114.5

Since only the difference of the angles θ_1 and θ_2 occurs, the direction to which they may both be referred can be arbitrarily chosen. These angles, together with the angle which gives their numerical difference, are also shown in Fig. 114.5. Positive values of θ_1 and θ_2 are measured counterclockwise from the positive imaginary axis. The function

$\log \{[x + i(y + \frac{1}{2})]/[x + i(y - \frac{1}{2})]\}$ will have its principal value along the beginning of the original path of integration C if $\varphi = 0$ at the point $\infty + iy$. Therefore, in deforming the path of integration to coincide with a path of steepest descent, we shall start with $\varphi = 0$. When the two branch points $i/2$ and $-i/2$ have been encircled on the original path C and the point $+\infty - iy$ has been reached, φ is again zero. This can be realized for a path deformed to coincide with the path of steepest descent, as shown by the dotted path in Fig. 114.6. The arrows denote the direction in which the path is traversed. The ends of the path g and g' are connected together at $x = +\infty$ so as to remain on the same Riemann sheet. This will be the case if φ is taken to be -2π at the joining point.

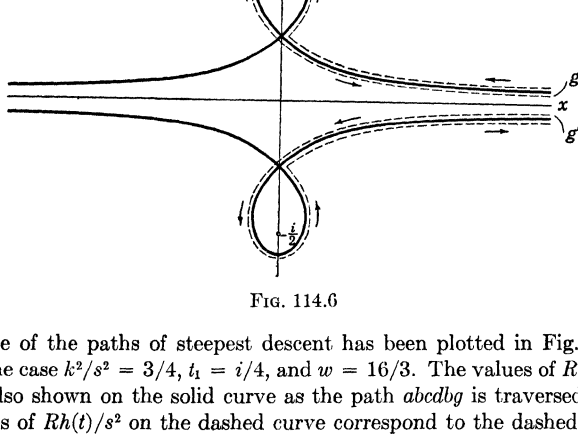


FIG. 114.6

One of the paths of steepest descent has been plotted in Fig. 114.7 for the case $k^2/s^2 = 3/4$, $t_1 = i/4$, and $w = 16/3$. The values of $Rh(t)/s^2$ are also shown on the solid curve as the path *abcdbg* is traversed. The values of $Rh(t)/s^2$ on the dashed curve correspond to the dashed curve *be*. It is evident that the part of the path *be* could not be used because $Rh(t)/s^2$ increases continually, while at the ends a, g, the term $Rh(t)/s^2 \to -\infty$.

The principal contributions to the integral occur over the part *ebe'* of the upper path and a similar section of the lower path. Tangent lines to the path at these points make 45° angles with the x axis, and as in the first case we shall substitute these tangent lines for the paths of steepest descents. As in the previous case $h(t)$ is expanded in the neighborhood of the points t_1 and t_2, so that

$$h(t) = h(\pm iy_1) + \frac{h''(\pm iy_1)}{2!}(t \mp iy_1)^2 + \dots \qquad (114.12)$$

where
$$h''(\pm iy_1) = \mp\frac{k^2}{2}\cdot\frac{iy_1}{(-y_1^2+\frac{1}{4})^2}$$
$$h(\pm iy_1) = \mp is^2y_1 + \frac{k^2i}{4}\log\frac{(\pm y_1+\frac{1}{2})}{(\pm y_1-\frac{1}{2})}$$

On the tangents to the path of steepest descent at iy_1 and $-iy_1$, we may write

$$\left.\begin{aligned} t - iy_1 &= \xi e^{-\pi i/4} \\ t + iy_1 &= \xi e^{-3\pi i/4} \end{aligned}\right\} \quad (114.13)$$

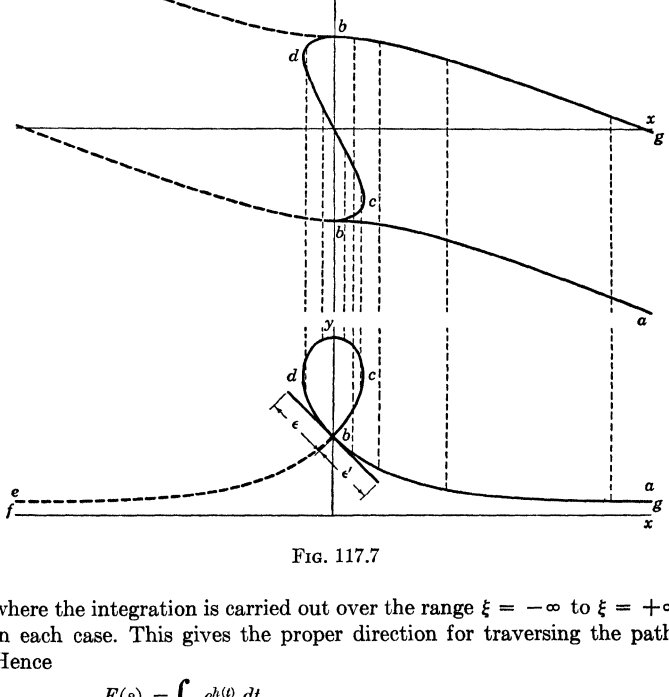


FIG. 117.7

where the integration is carried out over the range $\xi = -\infty$ to $\xi = +\infty$ in each case. This gives the proper direction for traversing the path. Hence

$$\begin{aligned} F(s) &= \int_C e^{h(t)}\,dt \\ &\approx e^{h(iy_1)}\int_{-\infty}^{\infty} e^{[h''(iy_1)/2]e^{-i\pi/2}\xi^2}\,e^{-\pi i/4}\,d\xi \\ &\quad + e^{h(-iy_1)}\int_{-\infty}^{\infty} e^{[h''(-iy_1)/2]e^{-3\pi i/4}\xi^2}\,e^{-3\pi i/4}\,d\xi \end{aligned}$$

When the values of $h''(\pm iy_1)$ are used from (114.12), the above result becomes

$$F(s) \cong (e^{h(iy_1)-\pi i/4} + e^{h(-iy_1)-3\pi i/4}) \int_{-\infty}^{+\infty} e^{[-k^2y_1/4(-y_1^2+1/4)^2]\xi^2}\, d\xi$$

$$\cong (e^{h(iy_1)-\pi i/4} + e^{h(-iy_1)-3\pi i/4}) \frac{2\sqrt{\pi}}{k\sqrt{y_1}} \left(-y_1^2 + \frac{1}{4}\right) \qquad (114.14)$$

When the values of $h(\pm iy_1)$ from (114.12) are used and the value of y_1 is substituted in (114.14), one finally has the desired asymptotic form for $F(s)$, when s and k are large and $\left|\frac{k}{s}\right| < 1$.

The above example has been worked out in some detail in order to indicate clearly some of the situations which may arise in applying this method to a practical problem.

115. Asymptotic solution of a differential equation. It quite often happens that one requires the form of the solution of a differential equation for large values of the independent variable or of some parameter. A method which is sometimes successful consists in finding the solution in terms of a contour integral, after which the methods of finding the asymptotic form outlined above may be applied. The general procedure may well be illustrated by means of the following examples.

Consider the differential equation

$$zL''(z) + (1 - z)L'(z) + nL(z) = 0 \qquad (115.1)$$

which occurs in connection with many problems in quantum theory. The variable z and parameter n will be assumed complex. One starts by letting

$$L(z) = \int_C K(z,t)v(t)dt \qquad (115.2)$$

where $v(t)$ is regarded as the unknown function of the complex variable t, and the function $K(z,t)$ is to be selected so that it will possess certain properties shortly to be indicated. The path of integration C is also to be chosen so that certain conditions will be fulfilled.

If eq. (115.2) is substituted in (115.1) and the order of integration and differentiation is inverted,

$$\int_C \left[z\frac{\partial^2 K}{\partial z^2} + (1 - z)\frac{\partial K}{\partial z} + nK\right]v(t)dt = 0 \qquad (115.3)$$

One then tries to find a function $K(z,t)$ such that derivatives of K with respect to z may be expressed in terms of partial derivatives with respect to t. Simple functions which have this property are

$$K(z,t) = (t + z)^\alpha \qquad (115.4)$$

$$K(z,t) = e^{tz} \qquad (115.5)$$

Furthermore, $K(z,t)$ should be chosen so that the resulting partial derivatives with respect to t can be removed by integrating (115.3) by parts. For the example being considered, the function defined by (115.4) is convenient. Then

$$z\frac{\partial^2 K}{\partial z^2} + (1 - z)\frac{\partial K}{\partial z} + nK$$
$$= (t + z)^{\alpha-2}[\alpha(\alpha - 1)z + \alpha(t + z) - \alpha z(t + z) + n(t + z)^2]$$

It is obvious that a simplification results when $\alpha = n$, in which case

$$z\frac{\partial^2 K}{\partial z^2} + (1 - z)\frac{\partial K}{\partial z} + nK = -t\frac{\partial^2 K}{\partial t^2} + (n + t)\frac{\partial K}{\partial t}$$

and (115.3) becomes

$$\int_C \left[t\frac{\partial^2 K}{\partial t^2} - (n + t)\frac{\partial K}{\partial t} \right] v(t)dt = 0$$

The next step is to remove the derivatives of K with respect to t by an appropriate number of integrations by parts. This can be done because z is not a function of the variable of integration t. After the integrations

$$\int_C \left[t\frac{\partial^2 K}{\partial t^2} - (n + t)\frac{\partial K}{\partial t} \right] v(t)dt$$
$$= \left[tv\frac{\partial K}{\partial t} - K\frac{d}{dt}(v \cdot t) - (n + t)vK \right]_{t_1}^{t_2}$$
$$+ \int_C K\left[\frac{d^2}{dt^2}(t \cdot v) + \frac{d}{dt}(n + t) \cdot v \right] dt \qquad (115.6)$$

where t_1 and t_2 are values of t at the ends of the path of integration. The remaining problem is to determine the function v and the path of integration so that the right-hand side of the above equation vanishes. First, v may be determined by making the integrand of the integral on the right identically zero, so that

$$\frac{d^2}{dt^2}(t \cdot v) + \frac{d}{dt}(n + t)v = 0$$

This differential equation may readily be solved for v. One integration can be effected at once, giving

$$\frac{d}{dt}(t \cdot v) + (n + t)v = 0 \qquad \frac{dv}{v} = -\frac{(n + t + 1)}{t}dt$$

which on further integration gives

$$v = t^{-(n+1)}e^{-t}$$

The path of integration must now be determined so that the terms $tv\dfrac{\partial K}{\partial t}\Big|_{t_1}^{t_2}$ etc. vanish. The natural path to take would be some closed

path, so that t_1 and t_2 would be identical. However, some care must be taken, since the terms under consideration are many-valued functions if n is not an integer, to insure that these functions have the same value after traversing the path once as they had at the starting point. Furthermore, if the closed path is chosen so that no singularities of $K \cdot v$ are enclosed, $L(z)$ will be identically zero and only a trivial solution will result. An inspection of the function $K \cdot v = (t + z)^n t^{-(n+1)} e^{-t}$ reveals two branch points, one at $t = 0$ and the other at $t = -z$, as shown in Fig. 115.1. If we choose the path C so that it encloses both branch points, the right-hand side of (115.6) will vanish. This will be evident by examining the individual terms as such a path is traversed. First consider the term

$$tv\frac{\partial K}{\partial t} = nt^{-n}(t + z)^{n-1}e^{-t}$$

The function e^{-t} is analytic, and its value after traversing the path shown in Fig. 115.1 will return to its value at the start, so that it is necessary only to consider the many-valued function $t^{-n}(t + z)^{n-1}$, which may be written

$$\rho_1^{-n} e^{-in\theta} \rho_2^{n-1} e^{i(n-1)\theta'}$$

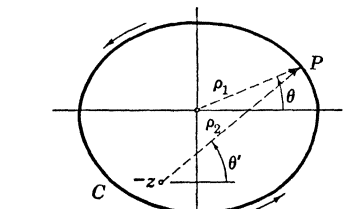


FIG. 115.1

By referring to Fig. 115.1 it is seen that as the point P traverses the path C and returns to its starting point, the value of θ goes from θ_1 to $\theta_1 + 2\pi$, and θ' goes from θ_2 to $\theta_2 + 2\pi$, so that at the start

$$t^n(t + z)^{n-1} = \rho_1^{-n} \rho_2^{n-1} e^{-in\theta_1} e^{i(n-1)\theta_2}$$

After one complete circuit

$$\begin{aligned} t^{-n}(t + z)^{n-1} &= \rho_1^{-n} \rho_2^{n-1} e^{-in(\theta_1 + 2\pi)} e^{i(n-1)(\theta_2 + 2\pi)} \\ &= \rho_1^{-n} \rho_2^{n-1} e^{-in\theta_1} e^{i(n-1)\theta_2} \end{aligned}$$

which is the same as the starting value; consequently $tv\dfrac{\partial K}{\partial t}\Big|_{t_1}^{t_2}$ vanishes over the path chosen. It should be noticed that this would not be the case unless both branch points were enclosed by the path C. A similar examination of the remaining terms in (115.6) shows that they will vanish when integrated over the closed path C. Consequently, we have the desired solution of the differential equation in the form of the contour integral

$$L(z) = \int_C (t + z)^n t^{-n} e^{-t}\, dt \tag{115.7}$$

The asymptotic expansion of (115.7) can be found by applying methods outlined in Sec. 112.

SUGGESTED REFERENCES

E. T. Whittaker and G. N. Watson, *Modern Analysis*, London, Cambridge University Press, 1920.

R. Courant and D. Hilbert, *Methoden der mathematischen Physik*, Vol. 1, Berlin, Julius Springer, 1931.

H. Jeffreys and Bertha S. Jeffreys, *Methods of Mathematical Physics*, London, Cambridge University Press, 1946.

E. T. Copson, *Theory of Functions of a Complex Variable*, New York, Oxford University Press, 1935.

PROBLEMS

105.1. Expand z in terms of $w = z/(1 - z)$, using the method of the text. Check by algebra.

105.2. Expand z in terms of $w = (e^z - 1)e^{-az}$.

105.3. Expand $f(z) = (-1)^n \dfrac{e^{-az/(1-z)}}{(1-z)^{n+1}}$ in terms of $w = z/(1 + z)$.

107.1. Show that

$$\frac{1}{e^z + 1} = \frac{1}{z} - \sum_{\substack{k \neq 0 \\ k = -\infty}}^{\infty} \frac{1}{z - k\pi i} + \frac{1}{k\pi i}.$$

107.2. Expand $e^{az}/(e^z - 1)$ in rational fractions.

109.1. Show that

$$(1 - z)\left(1 + \frac{z}{2}\right)\left(1 - \frac{z}{3}\right) \cdots = \frac{\sqrt{\pi}}{(1 + z/2)\Gamma(\tfrac{1}{2} - z/2)}.$$

109.2. Show that

$$|\Gamma(z)| = \sqrt{\frac{\pi}{y \sinh \pi y}} \qquad \text{where } z = iy.$$

112.1. By integrating by parts, find an asymptotic series for the Fresnel integrals

$$C(x) = \int_0^x \cos \frac{\pi}{2} t^2 \, dt, \qquad S(x) = \int_0^x \sin \frac{\pi}{2} t^2 \, dt.$$

Estimate the error involved in keeping only n terms. *Hint.* Put $t^2 = y$.

112.2. Find an asymptotic series for the exponential integral

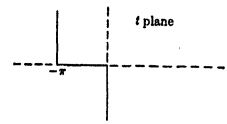
$$Ei(x) = \text{P.V.} \int_{-\infty}^{x} \frac{e^t}{t} \, dt$$

and an estimate of the error when only n terms are retained.

114.1. Find an asymptotic formula for the Hankel function

$$H_\lambda^{(1)}(z) = -\frac{1}{\pi} \int_C e^{-iz \sin t + i\lambda t} \, dt$$

where C is the contour shown below.



PROB. 114.1

114.2. The Hermite polynomials $H_n(z)$ are given by the generating function

$$e^{z^2-(s-z)^2} = \sum_{n=0}^{\infty} \frac{H_n(z)}{n!} s^n$$

Find an asymptotic formula for $H_n(z)$ by the method of steepest descents.

114.3. Determine an asymptotic formula for

$$L(z) = \int_C (t+z)^{\alpha} t^{-\alpha} e^{-t}\, dt$$

where the path of integration is that shown in Fig. 115.1 for two cases: (1) α is real, positive, and noninteger; (2) α is a pure imaginary.

Chapter 9

APPLICATION OF FUNCTIONS OF A COMPLEX VARIABLE TO POTENTIAL AND FLOW PROBLEMS

116. Connection between analytic functions and potential theory. In Sec. 82, it was shown that the real and imaginary parts U and V of any analytic function $w = U + iV = f(z)$ of the complex variable $z = x + iy$ are solutions of Laplace's equation so that

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0 \qquad \text{and} \qquad \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \tag{116.1}$$

Thus V is a potential function and is referred to as the potential function conjugate to U. In Sec. 82 it was shown that the conjugate potential function V could be uniquely determined when U is known except for an additive constant. Therefore since one may start with any potential function U and determine its conjugate V, one constructs an analytic function $f(z) = U + iV$ at the same time, so that the theory of potential functions is equivalent to the theory of analytic functions.

In potential theory, the force or field intensity F is a vector whose components in the x and y directions are given by $-\frac{\partial V}{\partial x}$ and $-\frac{\partial V}{\partial y}$, respectively, when V is taken to be the appropriate potential function. The field intensity F may be expressed directly in terms of the analytic function $w = U + iV = f(z)$ since from Sec. 82 it was shown that

$$i\frac{dw}{dz} = \frac{\partial U}{\partial y} + i\frac{\partial V}{\partial y} = -\frac{\partial V}{\partial x} + i\frac{\partial V}{\partial y} = F_x - iF_y = F^* \tag{116.2}$$

Thus the real part of $i(dw/dz)$ gives the value of the x component of the field intensity, and the imaginary part of $i(dw/dz)$ gives the $-y$ component. The absolute magnitude of F is simply

$$|F| = \left|\frac{dw}{dz}\right| \tag{116.3}$$

In the case of electric fields the charge density at the surface of a conductor (equipotential) surface is $\sigma = |F|/4\pi$, so that

$$\sigma = \frac{1}{4\pi} \left| \frac{dw}{dz} \right|_{\text{at surface}} \tag{116.4}$$

117. Conformal mapping and the solution of the potential problem. The relationship

$$w = U + iV = f(z) \tag{117.1}$$

will now be examined in more detail so as to indicate its usefulness in obtaining solutions for a certain class of potential problems. From a mathematical point of view eq. (117.1) establishes a transformation of a certain region D_z in the z plane to a region D_w in the w plane as long as $f(z)$ is single-valued. When $f(z)$ is not single-valued, the z plane is replaced by the appropriate Riemann surface (see Sec. 92). When $f(z)$ is analytic, the indicated transformation is of a special kind. This can be seen by referring to the corresponding domains D_w and D_z in Fig. 117.1.

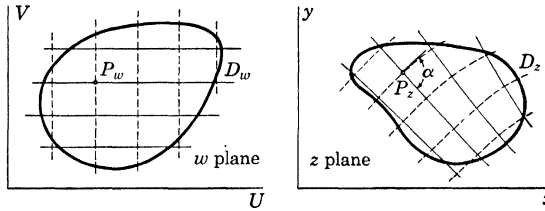


FIG. 117.1

The lines V = constant and U = constant of course intersect at right angles in the w plane and consequently divide D_w into rectangular pieces. Consider the intersection of any two such lines at the point P_w and let the point corresponding to this in the z plane be P_z. The curves U = constant go over into the dotted curves in the z plane and the curves V = constant go over into the solid curves in the z plane. It will be presently evident that these curves in the z plane also intersect at right angles. It has been shown in Sec. 45 that the direction cosines of the tangents to the curves $V(x,y)$ = constant and $U(x,y)$ = constant in the z plane are proportional to 1 and dy/dx, so that

$$\cos\alpha = k\left(1 + \frac{\partial U/\partial x}{\partial U/\partial y}\frac{\partial V/\partial x}{\partial V/\partial y}\right)$$

or the cosine of the angle α between the tangents of the intersecting curves in the z plane is proportional to

$$\frac{\partial U}{\partial y}\frac{\partial V}{\partial y} + \frac{\partial U}{\partial x}\frac{\partial V}{\partial x}$$

Since $f(z)$ is analytic, U and V satisfy the Cauchy-Riemann differential eqs. 82.3, so that the above quantity becomes

$$\frac{\partial U}{\partial y}\frac{\partial V}{\partial y} - \frac{\partial V}{\partial y}\frac{\partial U}{\partial y} = 0$$

and the curves V = constant, U = constant intersect at right angles in the z plane—a property of equipotential curves and lines of force.

The property that curves intersecting at right angles in the z plane also intersect at right angles in the w plane is a special case of a more general property of the transformation $w = f(z)$. In fact it can be shown* that if a transformation from the z plane to the w plane is defined by the analytic function $w = f(z)$, then the magnitude of the angle between two curves intersecting at any point in the z plane, for which $f'(z) \neq 0$, will be identical with that between the intersecting image curves in the w plane. As mentioned in Sec. 52, such transformations are called conformal transformations. In the theory of complex variables one speaks of mapping a domain of the z plane conformally on the corresponding domain in the w plane.

In this and the preceding section, we have seen that the real or imaginary part of any conformal mapping function $w = f(z)$ is a solution of Laplace's equation and is therefore the solution to some potential problem. In any practical problem it is known† that the unique solution is given by that solution of Laplace's equation which yields the prescribed potential distribution over the boundary of the region, which may be made up of surfaces of conductors, etc. Potential problems are mainly of two kinds depending on the nature of the boundary conditions. When the value of the potential $V_0(x,y)$ at all points on the boundary is given, this constitutes a potential boundary value problem of the first kind. On the other hand, when derivatives of the potential are given at the boundary (this is equivalent to prescribing the field intensity there), we have to do with a potential boundary value problem of the second kind.

From the point of view of conformal mapping, a unique solution of a given boundary value potential problem of the first kind (or second kind) will be obtained when a mapping function is found which maps the region under consideration in the z plane on a region of the w plane such that when the boundary of the region in the z plane is mapped on the w plane, the imaginary part of w, i.e., $V(x,y)$ (or its derivative) corresponds to the prescribed boundary values of the potential (or its derivative).

* See for instance Hurwitz and Courant, *Vorlesungen über allgemeine Funktiontheorie*, p. 289.

† See, for instance, L. Lichtenstein, *Neuere Entwicklung der Potentialtheorie; Encyklopedie der mathematischen Wissenschaften*, p. 218.

The validity of this important fundamental assertion rests on the following established theorems* in the theory of functions.

(1) Any simply connected region D_z in the z plane which is different from the entire plane or the plane with a single point removed can be mapped *conformally, uniquely,* and *reversibly* inside the unit circle in the ζ plane in such a way that a definite point and direction in D_z (for instance the zero point and direction of the x axis) correspond in the region inside the circle.

It is hardly necessary to add that this establishes the unique mapping of the z plane on the required w plane, since each can be mapped in the unit circle.

The second important theorem has to do with the mapping of a multiply connected region.

(2) There exists an analytic function† $w = U + iV = f(z)$ which uniquely maps an n-fold multiply connected region in the z plane on n strips in the w plane formed by straight lines parallel to the real axis.

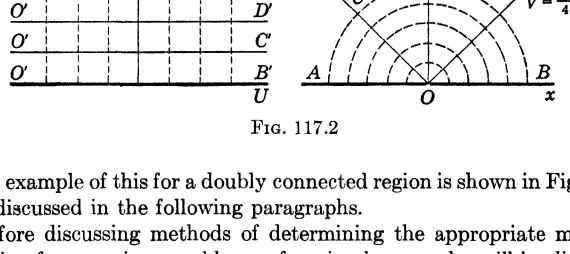


FIG. 117.2

An example of this for a doubly connected region is shown in Fig. 117.3 and discussed in the following paragraphs.

Before discussing methods of determining the appropriate mapping function for any given problem, a few simple examples will be discussed in order to clarify the general statements made in the preceding paragraph. First, let it be required to find the potential at any point in space above two semi-infinite planes, one being at zero potential and the other at potential V_0. The traces of these two planes in the complex z plane coincide with the real axis $y = 0$ as shown in Fig. 117.2. The plane containing the trace $y = 0$ and $x > 0$ will be taken at zero potential, and the plane containing the trace $y = 0$ $x < 0$ will be taken at the potential V_0. The potential at P is required. In accordance with the preceding paragraph, the unique solution will be had if a mapping

* See for instance Hurwitz and Courant, *Vorlesungen über allgemeine Funktionentheorie und elliptische Funktionen,* Berlin, Julius Springer, 1925, p. 365.

† See for instance Hurwitz and Courant, *op. cit.,* p. 438.

function $w = U + iV = f(z)$ can be found which maps the z plane on the w plane in such a way that $V = 0$ when z traverses OB and $V = V_0$ when z traverses AO. A function which does this is $w = (V_0/\pi) \log z$. This may be verified by writing $z = \rho e^{i\theta}$, so that

$$w = U + iV = \frac{V_0}{\pi} \log \rho e^{i\theta}$$

$$= \frac{V_0}{\pi} (\log \rho + i\theta) \qquad 0 \leq \theta \leq \pi \qquad (117.2)$$

and $V = V_0\theta/\pi$. On the trace OB, $\theta = 0$, and therefore $V = 0$ as required. It is readily seen that OB is mapped onto the entire real axis $O'B'$ in the w plane. On the trace AO, $\theta = \pi$ and consequently $V = V_0$ as required. The line OA is mapped onto the line $O'A'$ in the w plane. The mapping function presented above is the solution of our potential problem and

$$V = \frac{V_0\theta}{\pi} = \frac{V_0}{\pi} \tan^{-1} \frac{y}{x} \qquad x > 0$$

$$= \frac{V_0}{\pi} \left(\pi + \tan^{-1} \frac{y}{x}\right) \qquad x < 0$$

The lines of force are given by the equation $U =$ constant, which are therefore circles in the z plane. From eq. (116.3) the absolute value of the field intensity at any point is

$$|F| = \left|\frac{dw}{dz}\right| = \frac{V_0}{\pi} \left|\frac{1}{z}\right| = \frac{V_0}{\pi\rho}$$

It was mentioned in Sec. 116 that the curves $U =$ constant could be taken also as equipotential curves if these satisfied the boundary

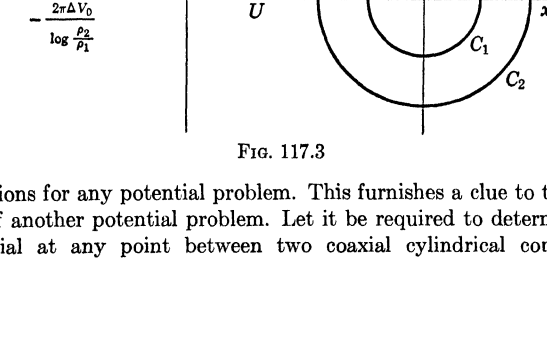


FIG. 117.3

conditions for any potential problem. This furnishes a clue to the solution of another potential problem. Let it be required to determine the potential at any point between two coaxial cylindrical conductors

between which there is a difference of potential ΔV_0. The traces of these cylinders on the complex z plane are shown in Fig. 117.3. Since the real part of $\log z$ = constant maps into a circle in the z plane, this function may be modified so that a circle in the z plane will map into a line in the w plane for which the imaginary part of w is constant. Such a function would be

$$w = U + iV = iA \log z = A(-\theta + i \log \rho)$$

The constant A may be determined, since V (on C_2) $-$ V (on C_1) $= \Delta V_0$, so that

$$V_2 - V_1 = A \log \rho_2 - A \log \rho_1 = \Delta V_0$$

or

$$A = \frac{\Delta V_0}{\log \rho_2/\rho_1} \quad \text{and} \quad w = \frac{i\Delta V_0}{\log \rho_2/\rho_1} \log z$$

The field intensity is

$$|F| = \frac{\Delta V_0}{\log \rho_2/\rho_1} \left|\frac{1}{z}\right| = \frac{\Delta V_0}{\log \rho_2/\rho_1} \frac{1}{\rho}$$

118. Methods for determining the proper mapping functions. In the preceding section it is stated that a suitable mapping function exists which provides a solution to a given potential problem. However in a general case it is difficult to deduce the required mapping function from a knowledge of the geometry of the boundary. Nevertheless, for certain classes of boundaries and boundary conditions the mapping function can be determined. The methods available for this, together with examples, will be given in the following sections.

119. Constant potential over a parametrically defined curve. When the trace of a body at constant potential V_0 in the z plane is defined by the parametric equations

$$x = f(s) \quad \text{and} \quad y = g(s) \qquad s \text{ real} \tag{119.1}$$

the mapping function from the z plane to the w plane can be found directly by writing

$$z = x + iy = f(U + iV) + ig(U + iV) \tag{119.2a}$$

$$= f(w) + ig(w) = F(w) \tag{119.2b}$$

If V is set equal to zero in (119.2a),

$$x + iy = f(U) + ig(U)$$

or since U is real,

$$x = f(U) \quad \text{and} \quad y = g(U)$$

This means that the mapping function (119.2a) or (119.2b) is a solution of the potential problem when the boundary specified by (119.1) is

at zero potential. Since $V + V_0$ is also a solution of Laplace's equation we may add the constant V_0 to the value of V found from (119.2) and thereby satisfy the boundary condition $V = V_0$ on the boundary.

As an example of this method, we shall determine the potential due to an elliptical cyclinder at the constant potential V_0. The trace of such a cylinder on the z plane is of course an ellipse, the parametric equations of which may be taken to be

$$x = a \cos t \qquad y = b \sin t \tag{119.3}$$

According to eq. (119.2) the required mapping function from the z to the w plane is

$$z = x + iy = a \cos w + ib \sin w \tag{119.4}$$

By writing

$$a = c \cos i\alpha = c \cosh \alpha$$
$$ib = c \sin i\alpha = ic \sinh \alpha$$

the mapping function may be written in the convenient form

$$z = c \cos (w - i\alpha) = c \cos [U + i(V - \alpha)] \tag{119.5}$$

Separating the real and imaginary parts of this equation one obtains

$$x = c \cos U \cosh (V - \alpha), \qquad y = -c \sin U \sinh (V - \alpha)$$

In order to solve for V it is only necessary to eliminate U by means of the relation $\cos^2 U + \sin^2 U = 1$. This results in the equation

$$\frac{x^2}{c^2 \cosh^2 (V - \alpha)} + \frac{y^2}{c^2 \sinh^2 (V - \alpha)} = 1 \tag{119.6}$$

Thus the equipotentials are confocal ellipses. In order to make $V = V_0$ on the boundary expressed by (119.3) it is necessary only to write

$$\frac{x^2}{c^2 \cosh^2 (V - V_0 - \alpha)} + \frac{y^2}{c^2 \sinh^2 (V - V_0 - \alpha)} = 1 \tag{119.7}$$

The lines of force are given by values of x and y satisfying the equation

$$\frac{x^2}{c^2 \cos^2 U} - \frac{y^2}{c^2 \sin^2 U} = 1$$

when U = constant, and consequently form a system of confocal hyperbolas.

In passing, it is worth noting that the limiting case $b \to 0$, $c \to a$, $\alpha \to 0$ provides the potential corresponding to a thin conducting sheet of width $2a$ at the potential V_0. The equipotential lines are found from (119.7) by placing $\alpha = 0$ and $c = a$, thereby still forming a system of confocal ellipses with the exception of the boundary, which is a degenerate ellipse.

The mapping of the w plane on the z plane accomplished by the func-

tion (119.5) is indicated in Fig. 119.1. The region external to the given ellipse maps into the region $V_0 \leq V < \infty$, $0 \leq U \leq 2\pi$ in the w plane.

The method outlined above has a rather limited usefulness because it cannot be extended so as to give the potential between two arbitrary confocal ellipses with an arbitrary difference of potential between them.

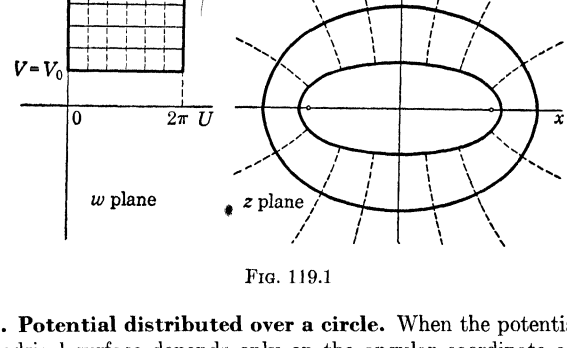


FIG. 119.1

120. Potential distributed over a circle. When the potential over a cylindrical surface depends only on the angular coordinate and the trace of the surface on the z plane is a circle of radius R, the mapping function which yields the potential in the interior may be found in the following manner. Let the potential on the circular boundary be given by the function $V(R,\varphi)$. Let us assume that the desired mapping function is the analytic function $w = U + iV = f(z)$. By means of Cauchy's integral formula [see eq. (86.2)], one may represent $f(z)$ at points interior to the circle in terms of the values of $f(z)$ on the boundary, so that

$$w = \frac{1}{2\pi i}\int_0^{2\pi} \frac{f(t)}{t - z}\, dt$$

where t denotes points on the circular trace. We may take the center of the circle to be at the origin $z = 0$ and write $t = Re^{i\varphi}$. The above integral then becomes

$$w = \frac{1}{2\pi}\int_0^{2\pi} f(R,\varphi)\frac{Re^{i\varphi}}{Re^{i\varphi} - re^{i\theta}}\, d\varphi \tag{120.1}$$

If the point $z = z_1$ lies outside the circle, then from Sec. 85,

$$\frac{1}{2\pi i}\int_0^{2\pi} \frac{f(t)}{t - z_1}\, dt = 0 \tag{120.2}$$

For z_1 we may take $z_1 = R^2/z = R^2e^{-i\theta}/r$ and subtract the zero result from (120.1), so that

$$w = \frac{1}{2\pi}\int_0^{2\pi} f(R,\varphi)\left(\frac{Re^{i\varphi}}{Re^{i\varphi} - re^{i\theta}} - \frac{Re^{i\varphi}}{Re^{i\varphi} - R^2e^{i\theta}/r}\right)d\varphi$$

$$= \frac{1}{2\pi}\int_0^{2\pi} f(R,\varphi)\left(\frac{Re^{i\varphi}}{Re^{i\varphi} - re^{i\theta}} - \frac{re^{i\varphi}}{re^{i\varphi} - Re^{i\theta}}\right)d\varphi$$

or

$$w = U + iV$$

$$= \frac{1}{2\pi}\int_0^{2\pi} [U(R,\varphi) + iV(R,\varphi)]\frac{R^2 - r^2}{R^2 - 2Rr\cos(\theta - \varphi) + r^2}d\varphi \quad (120.3)$$

On equating the imaginary parts of this equation one obtains the potential

$$V(r,\theta) = \frac{1}{2\pi}\int_0^{2\pi} V(R,\varphi)\frac{R^2 - r^2}{R^2 - 2Rr\cos(\theta - \varphi) + r^2}d\varphi \quad (120.4)$$

This formula is called the Poisson integral formula, and represents the unique solution of the potential problem stated above when $V(R,\varphi)$ is continuous on the boundary except for a finite number of finite discontinuities. At a discontinuity of $V(R,\varphi)$, the term $V(r,\theta)$ assumes a value lying between the limits

$$\lim_{\epsilon\to 0} V(R,\varphi+\epsilon) \quad \text{and} \quad \lim_{\epsilon\to 0} V(R,\varphi-\epsilon)$$

In order to determine the complete mapping function w it is necessary to determine U. This can be done by the methods given in Sec. 82, or $w = f(z)$ may be conveniently found by adding (120.1) and (120.2), which gives

$$w = f(z)$$

$$= \frac{1}{2\pi}\int_0^{2\pi} f(R,\varphi)\left(\frac{Re^{i\varphi}}{Re^{i\varphi} - re^{i\theta}} + \frac{re^{i\varphi}}{re^{i\varphi} - Re^{i\theta}}\right)d\varphi$$

$$= U + iV$$

$$= \frac{1}{2\pi}\int_0^{2\pi} [U(R,\varphi) + iV(R,\varphi)]\left[1 + i\frac{2Rr\sin(\theta - \varphi)}{R^2 - 2Rr\cos(\theta - \varphi) + r^2}\right]d\varphi$$

Equating the real parts of this equation, one obtains

$$U = \frac{1}{2\pi}\int_0^{2\pi} U(R,\varphi)d\varphi - \frac{1}{\pi}\int_0^{2\pi} V(R,\varphi)\frac{Rr\sin(\theta - \varphi)}{R^2 - 2Rr\cos(\theta - \varphi) + r^2}d\varphi$$

The first integral is a constant, and in fact is the value of U at $r = 0$ or at the center of the circle, so that

$$U(r,\theta) = U(0) - \frac{1}{\pi}\int_0^{2\pi} V(R,\varphi)\frac{Rr\sin(\theta - \varphi)}{R^2 - 2Rr\cos(\theta - \varphi) + r^2}d\varphi \quad (120.5)$$

The function U is thus determined in terms of the known potential $V(R,\varphi)$ on the boundary except for the additive constant $U(0)$. By combining (120.5) and (120.4), the mapping function $f(z)$ is found to be

$$
\begin{aligned}
w = f(z) &= U + iV \\
&= U(0) + \frac{1}{2\pi}\int_0^{2\pi} V(R,\varphi)\left[\frac{i(R^2 - r^2) - 2Rr\sin(\theta - \varphi)}{R^2 - 2Rr\cos(\theta - \varphi) + r^2}\right]d\varphi \\
&= U(0) + \frac{i}{2\pi}\int_0^{2\pi} V(R,\varphi)\frac{(Re^{i\varphi} + re^{i\theta})(Re^{-i\varphi} - re^{-i\theta})}{(Re^{i\varphi} - re^{i\theta})(Re^{-i\varphi} - re^{-i\theta})}\,d\varphi \\
&= U(0) + \frac{1}{2\pi i}\int_C V(t)\frac{t + z}{t - z}\frac{dt}{t} \qquad (120.6)
\end{aligned}
$$

where the integration is carried over the bounding circle in the last integral.

As an illustration of the use of the mapping function (120.6) let it be required to calculate the potential inside a circle of radius R when the potential on the boundary is given by $V_0 = \sin\varphi$. In terms of the complex variable t which traverse the circle $\sin\varphi = (1/2i)(t/R - R/t)$, and (120.6) becomes

$$
\begin{aligned}
w &= U(0) + \frac{1}{4\pi i}\int_C\left(\frac{t}{R} - \frac{R}{t}\right)\frac{t + z}{t - z}\frac{dt}{t} \\
&= U(0) + \frac{1}{4\pi i}\frac{1}{R}\int_C \frac{t + z}{t - z}\,dt - \frac{R}{4\pi i}\int_C \frac{t + z}{t - z}\frac{dt}{t^2}
\end{aligned}
$$

The above integrals can very easily be evaluated by the method of residues (see Sec. 94). The integrand of the first integral has a simple pole at $t = z$, while that of the second has poles at $t = z$ and $t = 0$. Calculating the residues in the usual way,

$$w = U(0) + \frac{z}{R}$$

or

$$w = U + iV = U(0) + \frac{r}{R}\cos\theta + \frac{ir}{R}\sin\theta$$

Since the imaginary part of $w(z)$ gives the potential, one obtains the immediate result

$$V = \frac{r}{R}\sin\theta$$

which satisfies the prescribed boundary condition.

In case the potential on the boundary has the constant value V_0, the theory of residues furnishes the immediate result

$$w = U(0) + iV_0$$

and the potential inside the boundary is everywhere equal to V_0.

The result (120.6), together with the theorems in Sec. 117, yield the solution of any potential problem as soon as the function which maps the region in question on the inside of a circle is found so that the boundary maps onto the circle.

121. Polygonal boundaries. Schwarz-Christoffel formula. The method of conformal mapping in potential theory can be used to great advantage when the boundary on which the potential is known is a polygon. This is because it is always possible to determine the mapping function which maps the interior and boundary of a polygon on a circle and its interior, after which the final mapping function in the w plane, and consequently the potential, can be determined by means of eq. (120.6). By a polygonal region is meant any region bounded by a finite number of straight lines. Thus a half plane, the strip between two parallel lines (Fig. 121.1a), an angle region (Fig. 121.1b), an "angle iron" section (Fig. 121.1c), the interior or exterior of a rectangle, the exterior of a semi-infinite line (Fig. 121.1d), and the exterior of a line e of finite length are all polygonal regions.

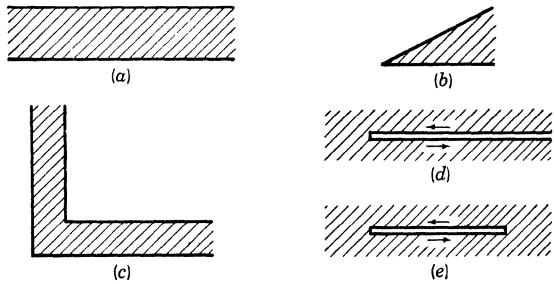


FIG. 121.1

When the potential over one part of a polygonal boundary is constant and the potential over the remainder is another constant, it is customary to map the polygonal region in the z plane on the w plane without first mapping the region in the z plane on to the circle in the ζ plane and then using eq. (120.6). Mapping the polygonal region on the w plane is most conveniently carried out by mapping the upper half of an auxiliary t plane ($\mathcal{I}mt \geq 0$) on the interior of the polygon in the z plane. Then the upper half of the t plane is mapped on to the w plane, and the variable t is eliminated from both mapping functions, thereby obtaining the mapping function from the z to the w plane.

The key to the above procedure is the determination of the function which maps the upper half of the t plane within the prescribed polygon in the z plane. This is accomplished by means of the Schwarz-Christoffel formula

$$z(t) = A\int (t - t_1)^{-p_1}(t - t_2)^{-p_2} \dots (t - t_n)^{-p_n}\, dt + B \qquad (121.1)$$

which maps the upper half of the t plane within a polygon having n sides

in such a way that the real axis in the t plane maps into the boundary of the polygon. Here $t_1, t_2, \ldots, t_n$ represent n points on the real axis in the t plane, which we shall regard as ordered so that

$$t_1 < t_2 < t_3 < \ldots < t_n; \quad p_1, p_2, p_3 \ldots p_n$$

are n real numbers which serve to measure the exterior angles of the polygon; A and B are arbitrary complex constants (except that $A \neq 0$) which serve to locate and orient the polygon in the z plane but do not affect its shape.

In order to see how eq. (121.1) accomplishes the desired mapping, consider the segment s_k of the real axis between the points t_{k-1} and t_k. As t traverses s_k, the quantity

$$\begin{aligned} \frac{dz}{dt} &= \frac{dx}{dt} + i\frac{dy}{dt} \\ &= A(t - t_1)^{-p_1}(t - t_2)^{-p_2} \ldots (t - t_n)^{-p_n} \end{aligned} \tag{121.2}$$

is a vector tangent to the curve into which the segment s_k transforms in the z plane. When dz/dt is defined by eq. (121.2), the image of s_k is a straight line having the direction $\arg (dz/dt)$. This may readily be seen by writing

$$\begin{aligned} \frac{dz}{dt} = {} & |A|\,|(t - t_1)|^{-p_1} \ldots \\ & \cdot |t - t_{k-1}|^{-p_{k-1}} |t - t_k|^{-p_k} \ldots |t - t_n|^{-p_n} e^{\arg (dz/dt)} \end{aligned}$$

where

$$\begin{aligned} \arg \frac{dz}{dt} = {} & i \arg A - ip_1 \arg (t - t_1) - \\ & \ldots - ip_{k-1} \arg (t - t_{k-1}) - ip_k \arg (t - t_k) - \\ & \ldots - ip_n \arg (t - t_n) \end{aligned} \tag{121.3}$$

On s_k, $t > t_{k-1} > t_{k-2} > \ldots > t_1$ and the arguments of

$$(t - t_1), (t - t_2), \ldots, (t - t_{k-1})$$

are zero for all $t > t_{k-1}$, while the arguments of

$$(t - t_k), (t - t_{k+1}), \ldots, (t - t_n)$$

are equal to π for all t on s_k, since

$$t < t_k < t_{k+1} < \ldots < t_n$$

Consequently, s_k maps into a straight line in the z plane, making an angle

$$\arg A - p_k\pi - p_{k+1}\pi - \ldots - p_n\pi$$

with the real axis in the z plane.

As t passes over the point t_k, the $\arg (dz/dt)$ jumps an amount $p_k\pi$, so that the image z_k of the point t_k in the z plane is a corner of the poly-

gon, and the exterior angle of the polygon at z_k is $p_k\pi$. The association of the points t_k and z_k and the appropriate angles are shown in Fig. 121.2.

There is a certain amount of arbitrariness in the choice of the values $t_1, t_2, \ldots, t_n$, which correspond to the corners of the polygon. This arises from the fact that a polygon with n corners is completely specified by the $2n$ coordinates of the n corners, while the Schwarz-Christoffel formula contains $2n + 3$ constants composed of n values of t_k, $n - 1$ values of p_k (because of the relation $\sum_{k=1}^{n} p_k\pi = 2\pi$), and the 4 real constants contained in the complex constants A and B. Consequently, three of the values t_k can be assigned at will, and the remainder will be uniquely determined.

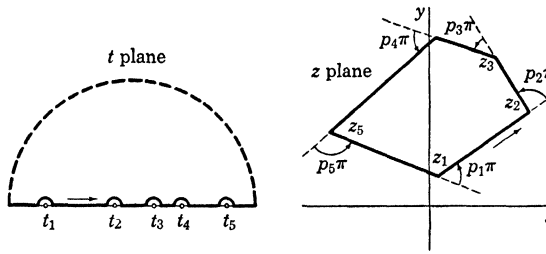


FIG. 121.2

It is important to notice the difference between the cases $p_k < 1$ and $p_k = 1$. When $p_k < 1$, we are dealing with a corner z_k of a polygon whose exterior angle is less than π. Furthermore, an examination of the mapping function (121.1) shows it to be continuous at $t = t_k$, and the corner z_k lies in a finite part of the z plane, i.e., $|z_k| < \infty$. The case $p_k = 1$ means an exterior angle of π, which at first sight would appear to mean that two sides of the polygon must coincide in part. This need not be the case, however, because an examination of the mapping function (121.1) shows that in the neighborhood of $t = t_k$ the mapping function has the form

$$z = C_1 \log (t - t_k) + C_2 \tag{121.4}$$

so that the corner z_k lies at infinity. The situation at $t = t_k$ can be made clear by assuming that C_1 and C_2 are real constants. Then for $t > t_k$, the $\arg t - t_k = 0$, so that

$$z = C_1 \log |t - t_k| + C_2 \qquad t > t_k$$

and for the range $t_k \leq t \leq \infty$, z traverses the real axis in the z plane.

However, when $t < t_k$, the arg $(t - t_k) = \pi$, and

$$z = C_1 \log | t - t_k | + C_2 + iC_1\pi \qquad t < t_k$$

which means that for $-\infty \leq t \leq t_k$, the variable z traverses a line parallel to the real z axis but a distance $C_1\pi$ from it. Consequently (121.4) provides the means of mapping the upper half of the t plane into a uniform strip of width $C_1\pi$ in the z plane. This function is of great use in mapping the t plane on the w plane referred to at the beginning of this section.

122. Mapping function for rectangular poles. As an example illustrating the use of the method outlined above, let it be required to determine the fringing field due to two poles of rectangular cross section large compared to the gap between them. The poles are shown in Fig. 122.1. Since the field in the gap at the center of the poles will not be

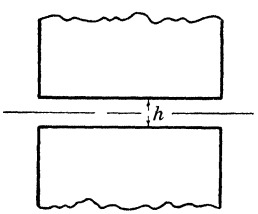


FIG. 122.1

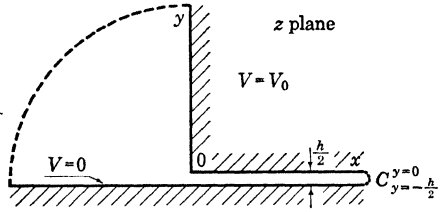


FIG. 122.2

appreciably different from the uniform field characteristic of the difference of potential between the poles, we may extend the gap to infinity on the right. Also the dashed line is an equipotential differing in potential from either of the poles by half the total difference of potential $2V_0$ between the poles. Consequently it will suffice to solve the potential problem for the configuration shown in Fig. 122.2.

The first step is to map the polygon of Fig. 122.2 on the upper half of the auxiliary t plane. Since this polygon has only three corners, all three values of the points t_k in the t plane corresponding to them can be chosen at will. A convenient choice is shown in the following table.

z_k	t_k	p_k
C, ∞	0	1
0	1	$-\frac{1}{2}$
. . . . .		
$-\infty$	∞	$\frac{3}{2}$
. . . . .		

The point enclosed in the dotted rectangle does not enter into the mapping function. The mapping function (121.1) may now be written.

$$z(t) = A\int t^{-1}(t-1)^{1/2}\,dt + B \tag{122.1}$$

The choice of making the point C in the z plane transform into $t = 0$ is convenient because this makes the parts of the polygon at different potentials go over into the negative and positive real axis in the t plane.

The constant A may be found by noting that when t traverses the point $t = 0$ on the real axis, the value of z must jump from $z = \infty - ih/2$ to $z = \infty$. Instead of actually going through the point $t = 0$ we may go around it on a small semicircle in the positive half plane by writing $t = \delta e^{i\theta}$. The jump in z will be given by integrating (122.1) over this semicircle and then taking the limit $\delta = 0$. On carrying out this procedure, one finds for the change in z,

$$\begin{aligned} &A\int_{\pi}^{0} \frac{e^{-i\theta}}{\delta}\,(\delta e^{i\theta} - 1)^{1/2}\delta i e^{i\theta}\,d\theta \\ &= A\int_{\pi}^{0} i\,d\theta\, i(1 - \tfrac{1}{2}\delta e^{i\theta} + \ldots) \\ &= A\pi + \text{terms in } \delta \end{aligned}$$

hence the limit as $\delta \to 0$ is just $A\pi$. This jump must be equal to $ih/2$; hence

$$A\pi = \frac{ih}{2} \quad \text{or} \quad A = \frac{ih}{2\pi}$$

Consequently the mapping function becomes

$$z(t) = \frac{ih}{2\pi}\int \frac{\sqrt{t-1}}{t}\,dt + B$$

This can be integrated by elementary means (see Sec. 28), so that

$$z = \frac{ih}{\pi}\,(\sqrt{t-1} - \tan^{-1}\sqrt{t-1}) + B$$

The constant B may easily be determined from the fact that the point $t = 1$ is required to map into the point $z = 0$, which requires that $B = 0$. Hence the complete mapping function is

$$z = \frac{ih}{\pi}(\sqrt{t-1} - \tan^{-1}\sqrt{t-1}) \tag{122.2}$$

It is instructive actually to check the correctness of this function. In the first place the negative real t axis should map into the line $y = -h/2$ in the z plane. To check this we write $t - 1 = |t-1|e^{i\pi}$ for $t < 0$, so that

$$\sqrt{t-1} = |t-1|^{1/2}e^{i\pi/2} = se^{i\pi/2} \quad \text{where} \quad s > 1 \tag{122.3}$$

If $\theta = \tan^{-1}\sqrt{t-1}$, then

$$\tan\theta = \frac{e^{i\theta} - e^{-i\theta}}{i(e^{i\theta} + e^{-i\theta})} = \frac{e^{2i\theta} - 1}{(e^{2i\theta} - 1)e^{i\pi/2}} = se^{i\pi/2}$$

Therefore

$$e^{2i\theta} = \frac{1 + se^{i\pi}}{1 - se^{i\pi}} = \frac{1 - s}{1 + s}$$

and

$$\theta = \frac{1}{2i}\log\frac{1 - s}{s + 1}$$

$$= \frac{1}{2i}\log\frac{e^{i\pi}(s-1)}{s+1} = \frac{1}{2i}\log\frac{s-1}{s+1} + \frac{\pi}{2} \qquad (122.4)$$

where $(s-1)/(s+1)$ is positive and real for $t < 0$. Consequently

$$z = \frac{ih}{\pi}\left(se^{i\pi/2} - \frac{\pi}{2} - \frac{1}{2i}\log\frac{s-1}{s+1}\right)$$

$$= \frac{-h}{\pi}\left(s + \frac{1}{2}\log\frac{s-1}{s+1}\right) - \frac{ih}{2}$$

As s traverses the range $\infty \geq s \geq 1$, $(-\infty \leq t \leq 0)$ the real part of z traverses the range $-\infty \leq Rz \leq +\infty$, so that z traverses a line parallel to the real z axis from $-\infty$ to $+\infty$ and at a distance $-h/2$ from it.

The next range to be tested is $0 \leq t \leq 1$. In this range

$$\sqrt{t-1} = |t-1|^{1/2}e^{i\pi/2} = s_1e^{i\pi/2} \quad \text{where} \quad 0 < s_1 < 1$$

A reference to eq. (122.4) shows that

$$\theta = \tan^{-1}\sqrt{t-1} = \frac{1}{2i}\log\frac{1 - s_1}{1 + s_1}$$

For this range,

$$z = \frac{ih}{\pi}\left(s_1e^{i\pi/2} - \frac{1}{2i}\log\frac{1 - s_1}{1 + s_1}\right)$$

$$= -\frac{h}{\pi}\left(s_1 + \frac{1}{2}\log\frac{1 - s_1}{1 + s_1}\right) = \frac{h}{\pi}\left(\log\frac{1 + s_1}{1 - s_1} - s_1\right)$$

The right-hand side of this equation is real and positive, and thus corresponds to the positive real axis of the z plane. It is easily verified that as $t \to 0$, $s_1 \to 1$, and $z \to \infty$, and as $t \to 1$, $s_1 \to 0$, and $z \approx \frac{h}{\pi}s_1 \to 0$, which shows that the real axis is traversed in the right direction.

On the final side of the polygon, $1 \leq t \leq \infty$, so that $\sqrt{t-1} = s_2$ is purely real and

$$z = \frac{ih}{\pi}(s_2 - \tan^{-1}s_2)$$

is purely imaginary, corresponding to the imaginary axis in the z plane. This completely verifies the correctness of the mapping function (122.2).

The next step is to map the w plane on the auxiliary t plane and then eliminate the variable t. Figure 122.3 shows the region of the w plane which is to be mapped on the upper half of the t plane, namely, that enclosed between the two equipotentials $V = 0$ and $V = V_0$. This can also be done with the help of the Schwarz-Christoffel formula (121.1). Since the point C, Fig. 122.2, where the potential jump occurred, mapped into the point $t = 0$ in the t plane, we shall make the point C' in the w plane go over into the point $t = 0$ in the t plane. The mapping function will be characterized by the points

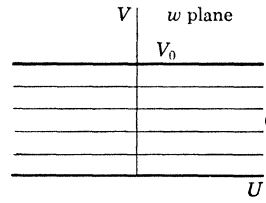


FIG. 122.3

w_k	t_k	p_k
$+\infty$	0	1
iV_0	1	0

Substituting these values in (121.1),

$$w(t) = A\int (t-0)^{-1}(t-1)^0\,dt + B = A\log t + B$$

When $t = 1$, $w = iV_0$, so that $iV_0 = B$. The constant A must be chosen so that $w(t)$ traverses the real axis when $-\infty < t < 0$, and traverses the line $V = V_0$ for $0 < t < \infty$. This requirement is fulfilled when $A = -V_0/\pi$ and

$$w(t) = \frac{-V_0}{\pi}\log t + iV_0 \tag{122.5}$$

By means of this mapping formula and that given by (122.2), we have mapped the z plane and the w plane on the same auxiliary t plane, and the problem will be completed when the z plane is mapped directly on the w plane by eliminating the variable t between expressions (122.2) and (122.5). From (122.5),

$$t = e^{-(w-iV_0)\pi/V_0} \tag{122.6}$$

and formula (122.2) then gives

$$z = \frac{ih}{\pi}\left\{\left[e^{\pi(iV_0-w)/V_0} - 1\right]^{1/2} - \tan^{-1}\left[e^{\pi(iV_0-w)/V_0} - 1\right]^{1/2}\right\} \tag{122.7}$$

The equipotential lines can be mapped out by holding V constant in $w = U + iV$, and letting U traverse the range $-\infty < U < \infty$, or the lines of force can be mapped by holding U constant and allowing V to traverse the range $0 \leq V \leq V_0$.

The field intensity at any point is found from formula (116.3), so that

$$F = -i\left(\frac{dw}{dz}\right)^* = -\frac{i}{(dz/dw)^*} = -i\left(\frac{dw}{dt}\frac{dt}{dz}\right)^* = -i\left(\frac{dw}{dt}\Big/\frac{dz}{dt}\right)^*$$

From (122.2),

$$\frac{dz}{dt} = \frac{ih}{2\pi}\frac{(t-1)^{1/2}}{t}$$

and from (122.5),

$$\frac{dw}{dt} = -\frac{V_0}{\pi t}$$

Consequently,

$$F = \frac{-2V_0}{h(t^*-1)^{1/2}}$$

From (122.6),

$$|t-1|^2 = \{e^{\pi[U+i(V-V_0)]/V_0} - 1\}\ \{e^{\pi[U-i(V-V_0)]/V_0} - 1\}$$
$$= 2e^{\pi U/V_0}\left[\cosh\frac{\pi U}{V_0} - \cos\frac{\pi(V-V_0)}{V_0}\right]$$

which allows us to express the absolute value of the field strength in the form

$$|F| = \frac{2^{3/4}V_0}{h}e^{-\pi U/4V_0}\left[\cosh\frac{\pi U}{V_0} - \cos\frac{\pi(V-V_0)}{V_0}\right]^{-1/4}$$

This essentially completes the determination of the potential distribution for the rectangular pole electrodes shown in Fig. 122.1.

123. Electric field in the cyclotron dees. As a further example of the Schwarz-Christoffel mapping function, which is of considerable practical importance, we shall calculate the electric field inside the accelerating electrodes used in a cyclotron. These electrodes are shown in Fig. 123.1. We shall suppose that the electrode A is at a potential V_0, and the potential of electrode B is $-V_0$. In a practical case the dimensions $2h$ and $2d$ are small compared with l_1 and l_2, so that we may assume that $l_1 = l_2 = \infty$. Consequently we have a two-dimensional problem, the geometry of which can be further simplified as shown in Fig. 123.2. The association of corners of the polygon in the z plane with the points on the t plane, and the values of the constants p_k to be used in the mapping formula (121.1) are also shown in Fig. 123.2. Considerations of sym

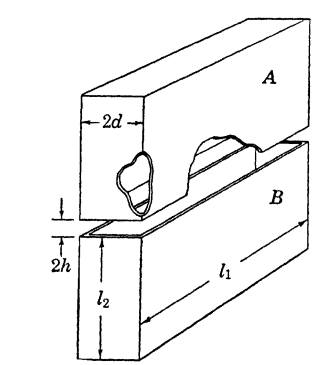


FIG. 123.1

metry lead to the association of the points on the t axis with the corners of the polygon. Because only three points of the polygon can be arbitrarily associated with three specified points on the t axis, the remainder being uniquely determined, the point $t = a$ cannot be prescribed, but will be determined later.

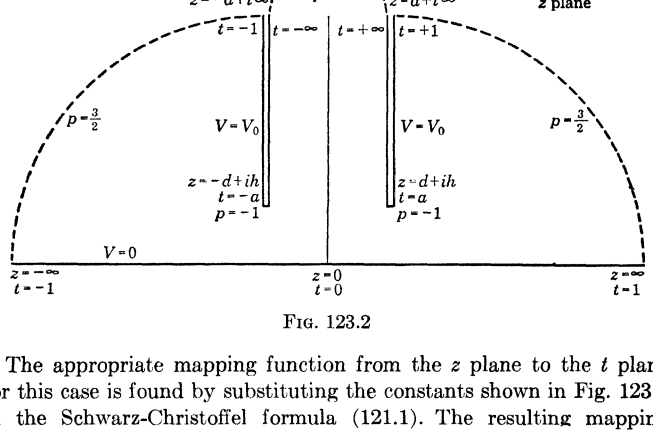


FIG. 123.2

The appropriate mapping function from the z plane to the t plane for this case is found by substituting the constants shown in Fig. 123.2 in the Schwarz-Christoffel formula (121.1). The resulting mapping function is

$$z = A\int (t + a)^1(t + 1)^{-3/2}(t - a)^1(t - 1)^{-3/2}\,dt + B$$
$$= A\int \frac{(t^2 - a^2)}{(t^2 - 1)^{3/2}}dt + B$$

The integration may readily be performed (cf. Sec. 28) by elementary methods, after which we find

$$z = A\left[\frac{(a^2 - 1)t}{\sqrt{t^2 - 1}} + \log\,(t + \sqrt{t^2 - 1})\right] + B \qquad (123.1)$$

The complex constants A and B and the real constant a must still be determined. The constants A and B may be determined in the following way. First it is to be noted that the quantity in brackets in eq. (123.1) is purely imaginary when $|\,t\,| < 1$, because

$$\frac{t}{\sqrt{t^2 - 1}} = -\frac{it}{\sqrt{1 - t^2}} \qquad t < 1$$

and

$$\log\,(t + \sqrt{t^2 - 1}) = i\,\frac{t}{|\,t\,|}\tan^{-1}\frac{\sqrt{1 - t^2}}{|\,t\,|} \qquad t < 1$$

Since the values of z for $|t| < 1$ must lie on the real axis in the z plane, A must be purely imaginary and B must be real. On the other hand, when $t > 1$ the quantity in brackets in expression (123.1) is real, which with A purely imaginary, requires that $B = d$, since the real part of z must be constant and equal to d for $t > 1$. Thus far we may write

$$z = iC\left[\frac{(a^2-1)t}{\sqrt{t^2-1}} + \log\left(t+\sqrt{t^2-1}\right)\right] + d \tag{123.2}$$

where C is real. For $t < -1$, this expression becomes

$$z = iC\left[\frac{-(a^2-1)|t|}{\sqrt{t^2-1}} + \log\left(|t| - \sqrt{t^2-1}\right)\right] - C\pi + d$$

In this expression the quantity in brackets is again real, and since for $t < -1$ the real part of z must be equal to $-d$, this requires that $C\pi = 2d$. The mapping function is then

$$z = \frac{2di}{\pi}\left[\frac{(a^2-1)t}{\sqrt{t^2-1}} + \log\left(t+\sqrt{t^2-1}\right)\right] + d \tag{123.3}$$

The real constant a can now be determined by noting that $z = d + ih$ when $t = a$, thereby requiring that

$$\frac{2d}{\pi}\left[a\sqrt{a^2-1} + \log\left(a+\sqrt{a^2-1}\right)\right] = h \tag{123.4}$$

With the value of a thus determined, the polygon in Fig. 123.2 is mapped on the half t plane by the function (123.3).

The completion of the potential problem requires the mapping of the w plane on the t plane. For this purpose the strip in the w plane bounded by the line $V = 0$ and $V = V_0$, shown in Fig. 123.3, can be regarded as a polygon. The association of values of t with the corners of the

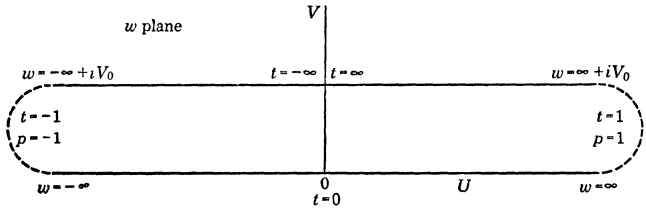


FIG. 123.3

polygon in the w plane is shown in Fig. 123.3. Using the values of t and p in the mapping formula (121.1) gives

$$w = A\int (t+1)^{-1}(t-1)^{-1}\,dt + B = A\int \frac{dt}{t^2-1} + B$$

$$= \frac{A}{2}\log\frac{t-1}{t+1} + B \tag{123.5}$$

When $|t| < 1$, the function (123.5) assumes the two forms

$$w = \frac{A}{2}\log\frac{1-t}{1+t} + \frac{A\pi i}{2} + B \qquad 0 < t < 1$$

$$w = \frac{A}{2}\log\frac{1+|t|}{1-|t|} + \frac{A\pi i}{2} + B \qquad -1 < t < 0$$

Since the logarithmic terms are real and w must have values on the real axis for $|t| < 1$, it is evident that A must be real and $B = -A\pi i/2$. On the other hand, when $|t| > 1$, the mapping function w assumes the two forms

$$w = \frac{A}{2}\log\frac{t-1}{t+1} + B \qquad t > 1$$

$$w = \frac{A}{2}\log\frac{|t|+1}{|t|-1} + B \qquad t < -1$$

Because A has been found to be real, and since the logarithmic terms in the above two expressions are real, the value of B must be

$$B = iV_0$$

to give the correct imaginary part of w for $t > 1$. This relation, together with the relation $B = -A\pi i/2$ established above, determines A to be

$$A = -\frac{2V_0}{\pi}$$

The required mapping function is then

$$w = -\frac{V_0}{\pi}\log\frac{t-1}{t+1} + iV_0 \tag{123.6}$$

The next step in completing the potential problem is the elimination of the complex variable t from eqs. (123.3) and (123.6). From eq. (123.6),

$$t = \frac{-[1 + e^{(\pi/V_0)(w-iV_0)}]}{1 - e^{(\pi/V_0)(w-iV_0)}}$$

$$\sqrt{t^2-1} = \frac{2e^{(\pi/2V_0)(w-iV_0)}}{1 - e^{(\pi/V_0)(w-iV_0)}}$$

Substituting these expressions in eq. (123.3),

$$z = \frac{2di}{\pi}\left\{(1-a^2)\frac{1 + e^{(\pi/V_0)(w-iV_0)}}{2e^{(\pi/2V_0)(w-iV_0)}} + \log\frac{[2e^{(\pi/2V_0)(w-iV_0)} - e^{(\pi/V_0)(w-iV_0)} - 1]}{1 - e^{(\pi/V_0)(w-iV_0)}}\right\} + d$$

$$= \frac{2di}{\pi}\left[(1-a^2)\frac{e^{-\pi w/2V_0} - e^{\pi w/2V_0}}{-2i} + \log\frac{2 + i(e^{\pi w/2V_0} - e^{-\pi w/2V_0})}{i(e^{\pi w/2V_0} + e^{-\pi w/2V_0})}\right] + d$$

$$= \frac{2d}{\pi}\left[(1-a^2)\sinh\frac{\pi w}{2V_0} + i\log\frac{\sinh(\pi w/2V_0) - i}{\cosh(\pi w/2V_0)}\right] + d \tag{123.7}$$

This formula may be used to map the equipotential lines by making V in w a constant between zero and V_0 and allowing U to take on all values in the range $-\infty < U < \infty$.

The electric force may be found with the aid of eqs. (116.3), (123.3) and (123.6).

$$F = -i\left(\frac{dw}{dz}\right)^* = -i\left(\frac{dw/dt}{dz/dt}\right)^*$$
$$= -i\left[\frac{-(2V_0/\pi)[1/(t^2-1)]}{(2di/\pi)(t^2-a^2)/(t^2-1)^{3/2}}\right]^* \tag{123.8}$$

This completes the main features involved in the solution of the electric field produced by the accelerating electrodes in a cyclotron.

124. Mapping of a polygon on a circle. The method of determining the potential on the inside of a long cylinder when the potential on the cylinder is a function of the angular position was treated in Sec. 120 by the method of conformal mapping. It is evident that the results found there can be used to determine the potential on the inside of a long polygonal cylinder when the potential on the polygonal boundary is a function of angular position if it is possible to find the mapping function which maps the interior of a polygon on the interior of a circle. This can readily be accomplished by first mapping the positive imaginary half of the t plane into the interior of a unit circle. The function which does this is

$$s = Re^{i\varphi} = \frac{t-i}{t+i} \tag{124.1}$$

The general features of this transformation can be seen by letting $t = \rho e^{i\theta}$, so that

$$s = Re^{i\varphi} = \frac{\rho e^{i\theta} - i}{\rho e^{i\theta} + i}$$

The expressions for R and φ, found in the usual way, are

$$R = \sqrt{\frac{1+\rho^2-2\rho\sin\theta}{1+\rho^2+2\rho\sin\theta}} \tag{124.2}$$

$$\varphi = \tan^{-1}\frac{(-2\rho\cos\theta)}{\rho^2-1} \tag{124.3}$$

The first of these expressions shows that for the imaginary part of t greater than or equal to zero, i.e., $0 \leq \theta \leq \pi$, $R \leq 1$, thereby indicating that the upper half of the t plane maps into the interior of the unit circle. From eq. (124.3), it is readily ascertained that as t traverses the real axis from $t = -\infty$ to $t = +\infty$, φ traverses the region $0 \leq \varphi \leq 2\pi$, thereby showing that the real t axis is transformed into a unit circle in the s plane.

Since eq. (121.1) maps a polygon on the t plane, and eq. (124.1) maps a unit circle on the t plane, the polygon may be mapped directly on the unit circle by eliminating t from these two expressions. Solving for t from eq. (124.1),

$$t = \frac{i(1+s)}{1-s} \tag{124.4}$$

Terms of the form $t - t_k$ are required in eq. (120.1). These can be written in the following convenient form

$$t - t_k = i\left(\frac{1+s}{1-s} - \frac{1+s_k}{1-s_k}\right) = \frac{i}{1-s}\left[1 + s - \frac{(1+s_k)(1-s)}{1-s_k}\right]$$
$$= \frac{2i(s-s_k)}{(1-s)(1-s_k)} = \frac{-2s_k i}{(1-s_k)(1-s)}\left(1 - \frac{s}{s_k}\right)$$

Substituting in eq. (121.1),

$$z = C\int (1-s)^{p_1+p_2+\quad+p_n}\left(1-\frac{s}{s_1}\right)^{-p_1}\left(1-\frac{s}{s_2}\right)^{-p_2}$$
$$\cdots\left(1-\frac{s}{s_n}\right)^{-p_n}\frac{ds}{(1-s)^2} + D$$

It has been assumed that $t = \infty$ does not correspond to a corner of the polygon and that the point $z = \infty$ is not contained in the polygon. Since the algebraic sum of the exterior angles of a polygon is 2π, it follows that $p_1 + p_2 + \ldots + p_n = 2$, and the above mapping function may be written

$$z = C\int \left(1-\frac{s}{s_1}\right)^{-p_1}\left(1-\frac{s}{s_2}\right)^{-p_2}\cdots\left(1-\frac{s}{s_n}\right)^{-p_n} ds + D \tag{124.5}$$

The above formula represents the most general mapping of the unit circle on a simply connected polygonal region with n sides not containing the point $z = \infty$.

By means of the function (124.5), the known potential $V_0(z)$ on the boundary of the polygon can be expressed as a known potential $V_1(s)$ on a unit circle in the s plane, which can then be used in formula (120.6) to determine the potential off the boundary.

125. Two-dimensional flow of an ideal fluid. The motion of a fluid in a plane having the same density at all points can be completely specified by the velocity components v_x and v_y. In the case of steady flow, v_x and v_y will be functions of x and y and independent of the time. If the region in the z plane (x,y plane) over which the flow is to be described is free of sources or sinks, so that any change in the state of the fluid in this region is a result of flowing across the boundary, certain conditions must be satisfied by the components of velocity v_x and v_y.

In order to see what these are, consider a closed curve C in the region of flow which contains no sources or sinks. The net flow of fluid must be zero into the region enclosed by C as measured by the quantity

$$\int_C v_n \, ds \tag{125.1}$$

where ds is an element of length along the curve C, and v_n is the component of velocity normal to the curve C and directed into the region enclosed by C. When v_n is expressed in terms of the components v_x and v_y and the direction cosines of the normal to C, it becomes

$$v_n = v_y \cos \alpha + v_y \cos \beta = -v_x \frac{dy}{ds} + v_y \frac{dx}{ds}$$

The condition that the net flow into the region enclosed by C is zero then becomes

$$\int_C v_n \, ds = \int_C v_y \, dx - v_x \, dy = 0 \tag{125.2}$$

In Sec. 55 we have seen that when a line integral of the above form vanishes for all choices of C, the functions v_x and v_y must satisfy the conditions

$$\frac{\partial v_x}{\partial x} = -\frac{\partial v_y}{\partial y} \tag{125.3}$$

which gives one of the conditions that the velocity components must satisfy.

The circulation of the fluid is measured by the quantity

$$\int_C v_s \, ds \tag{125.4}$$

where the curve C is traversed in the positive sense, and v_s is the component of the velocity tangential to the curve C. If the circulation for every closed curve C in the flow region vanishes, the flow is said to be vortex-free. Since

$$v_s = v_x \frac{dx}{ds} + v_y \frac{dy}{ds}$$

the vanishing of the circulation requires that

$$\int_C v_x \, dx + v_y \, dy = 0 \tag{125.5}$$

If this integral vanishes for all choices of the curve C, then

$$\frac{\partial v_x}{\partial y} = \frac{\partial v_y}{\partial x} \tag{125.6}$$

is a second condition which the components of velocity must satisfy. Furthermore a necessary and sufficient condition (cf. Sec. 55) for sat-

isfying conditions (125.5) is that the integrand be the differential of some function $\Phi(x,y)$. This means that

$$d\Phi = \frac{\partial \Phi}{\partial x}\,dx + \frac{\partial \Phi}{\partial y}\,dy = v_x\,dx + v_y\,dy$$

which in turn requires that

$$v_x = \frac{\partial \Phi}{\partial x} \tag{125.7}$$

$$v_y = \frac{\partial \Phi}{\partial y} \tag{125.8}$$

Consequently, when the flow of an incompressible fluid is vortex-free, the components of velocity are derivable from a velocity potential Φ. The conditions imposed on Φ by eqs. (125.7), (125.8), and (125.3) require Φ to be a solution of Laplace's equation. In Sec. 82, it was shown that the real and imaginary parts of any analytic function w of the complex variable z satisfied Laplace's equation, so that we are led to regard the velocity potential Φ as the real part of an analytic function w which we may write

$$w = \Phi + i\Psi \tag{125.9}$$

Since we know that the curves Ψ = constant intersect at right angles with the curves Φ = constant, we are led to associate the former curves with the stream lines. That this is actually possible can be seen by noting that the equation of the stream lines is

$$\frac{dx}{v_x} = \frac{dy}{v_y}$$

or from eqs. (125.7) and (125.8) this becomes

$$\frac{\partial \Phi}{\partial y}\,dx - \frac{\partial \Phi}{\partial x}\,dy = 0$$

This equation may be expressed in terms of ψ by means of the Cauchy-Riemann equations (cf. Sec. 82), in which case it becomes

$$\frac{\partial \psi}{\partial x}\,dx + \frac{\partial \psi}{\partial y}\,dy = 0$$

The general solution of this equation is $\psi(x,y)$ = constant, thereby showing that the stream lines are mapped on the z plane when the imaginary part of w is placed equal to a constant.

The vector velocity can be expressed in terms of the derivative of the function w, since

$$\frac{dw}{dz} = \frac{\partial w}{\partial x} = \frac{\partial \Phi}{\partial x} + i\frac{\partial \psi}{\partial x} = \frac{\partial \Phi}{\partial x} - i\frac{\partial \Phi}{\partial y} = v_x - iv_y$$

Consequently

$$v = v_x + iv_y = \left(\frac{dw}{dz}\right)^* \tag{125.10}$$

From the above discussion it is evident that the methods of conformal mapping already outlined for the solution of potential problems can be applied to the flow of an ideal fluid. The boundary conditions are mainly of two distinct types or a mixture of these. One type is that for which Ψ = constant on the boundary. This corresponds to a case where the boundaries of the region of flow are made up of fixed surfaces. Since there can be no shear in an ideal fluid, the traces of these surfaces in the z plane must coincide with a stream line, that is, a line over which Ψ = constant. In the other main type of boundary condition, the components of the velocity are known, which means the function Φ must be such that the first-order partial derivatives $\partial\Phi/\partial x$ and $\partial\Phi/\partial y$ give the prescribed values of v_x and v_y over the boundary. Flow problems involving a mixture of these boundary conditions often occur, in particular when a fluid flows around material obstacles.

It should be mentioned that cases often arise where the mapping function w may fail to be analytic at isolated points of the z plane. Such singular points then introduce sources, sinks, or vortex points into the flow region. This situation may be made clear by an examination of the mapping function

$$w = \frac{q}{2\pi} \log z = \frac{q}{2\pi} (\log \rho + i\varphi) \tag{125.11}$$

In accordance with the above discussion the stream lines are given by the equation $q\varphi/2\pi$ = constant, which are radial lines with origin at $z = 0$. From eq. (125.10) the vector velocity is

$$v = \left(\frac{dw}{dz}\right)^* = \frac{q}{2\pi z^*} = \frac{q}{2\pi\rho} e^{i\varphi}$$

If a circle of radius ρ is described with center at $z = 0$, the quantity of fluid flowing across this boundary from inside to outside is proportional to

$$\int_C v_\rho \, ds = \int_0^{2\pi} \frac{q}{2\pi\rho} \rho \, d\varphi = q \tag{125.12}$$

where q may be called the discharge per unit length of line source. The type of flow here considered is that due to a line source at $z = 0$ and a sink at $z = \infty$.

If in eq. (125.11) q is replaced by $i\Gamma$, we have the mapping function

$$w = \frac{i\Gamma}{2\pi} \log z = \frac{i\Gamma}{2\pi} (\log \rho + i\varphi) \tag{125.13}$$

from which the equation for the stream lines is $(\Gamma/2\pi) \log \rho = \text{constant}$, which in this case are circles. The vector velocity is now

$$v = \left(\frac{dw}{dz}\right)^* = \left(\frac{i\Gamma}{2\pi z}\right)^* = \frac{\Gamma}{2\pi\rho} e^{i\varphi - i\pi/2}$$

Thus the velocity has the magnitude $\Gamma/2\pi\rho$ in a direction tangent to the circle with radius ρ, and directed in the clockwise or negative direction. From eq. (125.4), we find the circulation about a circle of radius ρ with center at the origin to be

$$\int_C v_s \, ds = -\frac{\Gamma}{2\pi} \int_0^{2\pi} \frac{1}{\rho} \rho \, d\varphi = -\Gamma \tag{125.14}$$

so that Γ is the circulation or vortex strength in the positive or counterclockwise direction of rotation.

126. Flow of an ideal fluid around a cylinder. The method of the previous section may readily be applied to the flow of an ideal fluid around a cylinder whose axis is at right angles to the direction of flow. Let the trace of this cylinder in the z plane be a circle with radius a and center at $z = 0$, as shown in Fig. 126.1. At very great distances from the circle, it will be supposed that the fluid has a velocity v_0 in the x direction. The stream lines will be symmetrical with respect to the real axis. Since for an ideal fluid, the boundary of an immersed object must coincide with a stream line, the appropriate mapping function must map the line $ABCDE$ in Fig. 126.1 into a line which is parallel to the real axis in the w plane.

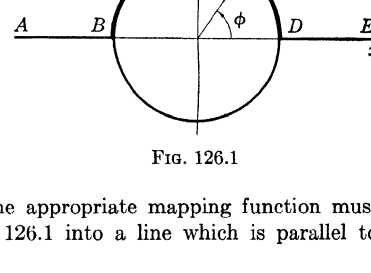


FIG. 126.1

The semicircular path BCD can be made to map on the real axis in the w plane by letting $w = K \cos \varphi$ for values of z on the circle. Since

$$\cos \varphi = \frac{e^{i\varphi} + e^{-i\varphi}}{2} = \frac{1}{2}\left(\frac{z}{a} + \frac{a}{z}\right)$$

for z on the circle, the function w can be written

$$w = \frac{K}{2}\left(\frac{z}{a} + \frac{a}{z}\right) = \frac{K}{2}\left(\frac{\rho}{a} e^{i\varphi} + \frac{a}{\rho} e^{-i\varphi}\right) \tag{126.1}$$

when z lies on the negative axis ($x < -a$), $\varphi = \pi$, and

$$w = -\frac{K}{2}\left(\frac{\rho}{a} + \frac{a}{\rho}\right)$$

Consequently, if K is real and positive, the negative real z axis, $x < -a$, maps into the negative real axis in the w plane. For $x > a$, then

$$w = \frac{K}{2}\left(\frac{\rho}{a} + \frac{a}{\rho}\right)$$

so that as the entire curve $ABCDE$ is traversed, w traverses the real axis. The constant K may be determined by the velocity condition at $z = \infty$. From the relation (125.10),

$$v = \left(\frac{dw}{dz}\right)^* = \frac{K}{2}\left(\frac{1}{a} - \frac{a}{z^2}\right)^*$$

Since $v = v_0$ for $z \to \infty$, the constant K is determined by the relation

$$v_0 = \frac{K}{2a}$$

and the complete mapping function is

$$w = v_0\left(z + \frac{a^2}{z}\right) \tag{126.2}$$

The effect of circulation around the cylinder can be taken into account [eq. (125.13)] by adding the term $-(i\Gamma/2\pi) \log z$, since this would not change the velocity $v = v_0$ at $z = \infty$.

The flow around any obstacle can be determined by using the function (126.2) if its boundary and interior in the t plane can be mapped on a circle in the z plane by means of a mapping function of the form

$$z = t + \frac{c_1}{t} + \frac{c_2}{t^2} + \cdots \tag{126.3}$$

A higher positive power of t cannot be allowed, because of the condition that at great distances from the obstacle the velocity v must be v_0. This may be readily seen from eqs. (126.2) and (126.3) by noting that

$$\lim_{|t|\to\infty} v = \lim_{|t|\to\infty} \left(\frac{dw}{dz}\frac{dz}{dt}\right)^*$$
$$= \lim_{|t|\to\infty} v_0\left(1 - \frac{a^2}{z^2}\right)^*\left(1 - \frac{c_1}{t^2} - \frac{2c_2}{t^2} - \cdots\right)^* = v_0$$

The determination of the functions for mapping polygons on circles was discussed in Sec. 124.

127. Complex force and moment on immersed obstacles. The force on an obstacle around which a fluid is flowing can be expressed very conveniently in terms of the mapping function w when the flow is two-dimensional. The x component of force acting on an element of fluid of dimension dx, dy, and unity will be $(\rho X - \partial p/\partial x)dx\,dy$, where p is the pressure, ρ is the density, and X is the x component of the

external force per unit mass acting on the element of mass $\rho\, dx\, dy$. Since this force must equal the x component of the acceleration,

$$\rho X - \frac{\partial p}{\partial x} = \rho \frac{dv_x}{dt} = \rho\left(\frac{\partial v_x}{\partial t} + \frac{\partial v_x}{\partial x} v_x + \frac{\partial v_x}{\partial y} v_y\right)$$
$$= \rho \frac{\partial v_x}{\partial t} + \frac{\partial}{\partial x}(\rho v_x^2) + \frac{\partial}{\partial y}(\rho v_x v_y) - v_x\left[\frac{\partial}{\partial x}(\rho v_x) + \frac{\partial}{\partial y}(\rho v_y)\right] \tag{127.1}$$

The last term in brackets can be replaced by $v_x(\partial\rho/\partial t)$ because when no sources or sinks are present in the element the continuity equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho v_x) + \frac{\partial}{\partial y}(\rho v_y) = 0$$

must be satisfied. For a steady-state condition all partial derivatives with respect to the time are zero, so that eq. (127.1) becomes

$$\rho X - \frac{\partial p}{\partial x} = \frac{\partial}{\partial x}(\rho v_x^2) + \frac{\partial}{\partial y}(\rho v_x v_y) \tag{127.2}$$

This differential equation must be satisfied throughout the region of flow, and in particular can be integrated over a volume of fluid located between the boundary of the immersed obstacle and an arbitrary boundary surrounding it of unit length in a direction at right angles to the xy plane. Let the surface of the obstacle be S_0 and that of the arbitrary

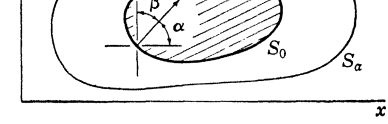


FIG. 127.1

boundary be S_a, as shown in Fig. 127.1. An integration of the above equation over the volume bounded by the surface S_0 and S_a gives

$$\iiint_V \rho X\, dV = \iiint \left[\frac{\partial}{\partial x}(p + \rho v_x^2) + \frac{\partial}{\partial y}(\rho v_x v_y)\right] dV$$
$$= \iint_{S_0} [(p + \rho v_x^2)\cos\alpha + \rho v_x v_y \cos\beta]\, dS$$
$$+ \iint_{S_a} [(p + \rho v_x^2)\cos\alpha + \rho v_x v_y \cos\beta]\, dS$$

where the volume integral on the right has been expressed as a surface

integral with the help of Green's theorem (cf. Sec. 67) and $\cos\alpha$ and $\cos\beta$ are the direction cosines of the normal outward from the fluid. The above equation can be conveniently rearranged as follows.

$$\iiint_V \rho X\, dV = \iint_{S_0} p\cos\alpha\, dS + \iint_{S_0} \rho v_x(v_x\cos\alpha + v_y\cos\beta)dS$$
$$+ \iint_{S_a} p\cos\alpha\, dS + \iint_{S_a} \rho v_x(v_x\cos\alpha + v_y\cos\beta)dS$$

In the second integral on the right, the quantity $v_x\cos\alpha + v_y\cos\beta$ is the component of velocity normal to the surface of the obstacle, which must be zero since the trace of this surface must be a stream line. Furthermore the first integral on the right is just the x component of force F_x acting on the obstacle, so that finally,

$$F_x = \iiint_V \rho X\, dV - \iint_{S_a} \rho v_x(v_x\cos\alpha + v_y\cos\beta)dS$$
$$- \iint_{S_a} p\cos\alpha\, dS \tag{127.3}$$

When the external forces may be neglected, the first term on the right vanishes and the pressure p can be expressed in terms of the kinetic energy by means of the Bernoulli equation

$$p = p_0 - \frac{1}{2}\rho v^2 = p_0 - \frac{1}{2}\rho(v_x^2 + v_y^2) \tag{127.4}$$

where p_0 is a constant. The expression (127.3) then becomes

$$F_x = \frac{\rho}{2}\iint_{S_0}[(v_x^2 + v_y^2)\cos\alpha - 2v_x(v_x\cos\alpha + v_y\cos\beta)]dS$$

Since the surface under consideration is generated by a line perpendicular to the xy plane, the above integral is really a line integral, and F_x is the force per unit length of the obstacle.

$$F_x = \frac{\rho}{2}\int_{C_a}[(v_x^2 + v_y^2)\cos\alpha - 2v_x(v_x\cos\alpha + v_y\cos\beta)]ds$$

Then $$\cos\alpha = \frac{dy}{ds} \quad \text{and} \quad \cos\beta = -\frac{dx}{ds}$$

so that $$F_x = \frac{\rho}{2}\int_{C_a}(v_y^2 - v_x^2)dy + 2v_xv_y\, dx$$

Similar reasoning shows that

$$F_y = -\frac{\rho}{2}\int_{C_a}(v_x^2 - v_y^2)dx + 2v_xv_y\, dy$$

From these expressions we can express the complex force as

$$F^* = F_x - iF_y = \frac{i\rho}{2}\int_{C_a}(v_x^2 - v_y^2 - 2iv_xv_y)(dx + i\, dy)$$

or from expression (125.10),

$$F^* = F_x - iF_y = \frac{i\rho}{2}\int_{C_a}\left(\frac{dw}{dz}\right)^2 dz \tag{127.5}$$

The turning moment on the obstacle can be derived in much the same way, and can be conveniently expressed as

$$M_z = -\text{real part of } \frac{\rho}{2}\int_{C_a} z\left(\frac{dw}{dz}\right)^2 dz \tag{127.6}$$

The above formulas for the force and turning moment on an immersed obstacle are particularly convenient to use, because the path C_a can be taken to be a circle enclosing the object, and in many cases the method of residues (cf. Sec. 94) can be used to evaluate the integrals.

128. Fluid flow with free surfaces. When a part of the boundary of a region of flow is free, so that the fluid is not in contact with the surface of any object and the gravitational or other external forces can be neglected, it is advantageous to use the following hodographic method in determining the appropriate mapping functions. Under the above conditions the trace of the free surface on the complex z plane must be a stream line along which the velocity is constant. This fact follows from Bernoulli's equation [cf. eq. (127.4)]. Since part of the fluid boundary is to be a stream line with constant velocity, it is convenient to make use of an intermediate mapping function $s = f(w)$ in which stream lines of constant velocity in the z plane map into arcs of circles in the s plane. In order to obtain such a function, we start with the usual mapping function

$$w = F(z) = \Phi + i\Psi$$

(cf. Sec. 125) in which Ψ = constant is the equation of a stream line, and solve for z as a function of w, so that

$$z = x + iy = z(w)$$

By means of eq. (125.10), we may write

$$\frac{dz}{dw} = \frac{1}{dw/dz} = \frac{1}{v_x - iv_y}$$

or

$$\frac{dz}{dw} = \frac{v_x + iv_y}{|\,v\,|^2}$$

This is the intermediate mapping function described, and we therefore write

$$s = \frac{dz}{dw} = \frac{v_x + iv_y}{|\,v\,|^2} = \frac{|\,v\,|e^{+i\varphi}}{|\,v\,|^2} = \frac{e^{+i\varphi}}{|\,v\,|} \tag{128.1}$$

As is evident from this equation, s represents a vector whose length is $1/|\,v\,|$ and whose direction is that of the velocity. The curve described

by the end of this vector s as a stream line is traversed is called an hodograph.

The next step is to map the region in the s plane, which in turn maps into the flow region in the z plane onto a strip of the w plane lying between two lines parallel to the real axis. When this mapping function $s = s(w)$ is found, it is readily seen from eq. (128.1) that

$$z = \int s\,dw = z(w) = x(\Phi,\Psi) + iy(\Phi,\Psi) \tag{128.2}$$

The procedure outlined above for finding $z = z(w)$ in the case of free flow may be illustrated by calculating the flow of a fluid from a semi-infinite reservoir through a jet. In Fig. 128.1a, the fluid occupies the

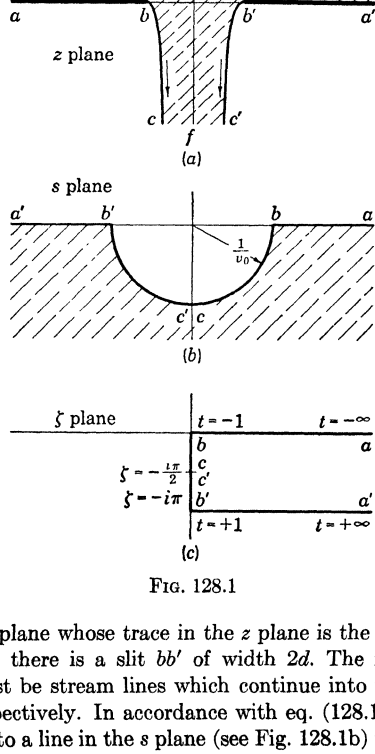


FIG. 128.1

region above a plane whose trace in the z plane is the real axis. In the bounding plane there is a slit bb' of width $2d$. The fixed boundaries ab and $a'b'$ must be stream lines which continue into the free surfaces bc and $b'c'$, respectively. In accordance with eq. (128.1), the boundary ab must map into a line in the s plane (see Fig. 128.1b) for which $\varphi = 0$,

since this is the direction of the velocity on the boundary ab, while the free surface bc must map into the arc of a circle whose radius is $1/v_0$, where v_0 is the constant velocity over the free surface stream line bc. Taking the velocity of the fluid in the reservoir to be zero at a great distance from the gap bb', the absolute value of s at $a(z = -\infty)$ is $s = \infty$. Corresponding to b, $|s| = 1/v_0$. Consequently the boundary ab maps into ab in the s plane, and bc maps into the quarter circle bc in the s plane. Likewise the boundary and free surface $a'b'c'$ maps into $a'b'c'$ in the s plane. The crosshatched region of the s plane must now be mapped in a strip between two lines parallel to the real axis in the w plane.

The mapping of the s plane onto the w plane can be carried out by a straightforward application of the methods discussed in Sec. 121. To apply these methods, the s plane is mapped inside a polygonal region in the ζ plane by means of the relation

$$\zeta = \log v_0 s \tag{128.3}$$

which maps the boundary abc in the s plane into the boundary abc in the ζ plane shown in Fig. 128.1c. The boundary $a'b'c'$ in the s plane goes over into the boundary $a'b'c'$ in the ζ plane. The mapping of the ζ plane on a strip in the w plane can be done by mapping each of these polygons on an auxiliary t plane and then eliminating the variable t. The parallel strip in the w plane is mapped on the auxiliary t plane [see eq. (122.5)] by the function

$$w = C \log t + D \tag{128.4}$$

The polygon in the ζ plane can be mapped on the t plane by employing the Schwarz-Christoffel formula [see eq. (121.1)]. Applying this to the polygon shown in Fig. 128.1c,

$$\begin{aligned} \zeta &= A \int (t + 1)^{-1/2}(t - 1)^{-1/2}\, dt + B \\ &= A \log (t + \sqrt{t^2 - 1}) + B \end{aligned}$$

The fact that $\zeta = 0$ when $t = -1$ requires that $B = -Ai\pi$, and when $t = +1$, then $\zeta = -i\pi$, so that A must be 1. The complete function is then

$$\zeta = \log (t + \sqrt{t^2 - 1}) - i\pi \tag{128.5}$$

A little care is required in order that the correct value be assigned to this function when $t < -1$. The proper value of the above function for $t < -1$ can readily be found by referring to the t^2 plane shown in Fig. 128.2. On writing $t = |t| e^{i\theta}$, it is readily seen from the dashed circle ($|t| < 1$) that

$$\frac{\pi}{2} < \arg (t^2 - 1) < \frac{3\pi}{2} \quad \text{when} \quad 0 \leq \theta \leq \pi$$

and $$\frac{\pi}{4} < \arg\sqrt{t^2 - 1} < \frac{3\pi}{4}$$

On the other hand, when $|t| > 1$ (solid circle) it may be seen that

$$0 \leq \arg(t^2 - 1) \leq 2\pi \quad \text{for} \quad 0 \leq \theta \leq \pi$$

and $$0 \leq \arg\sqrt{t^2 - 1} \leq \pi$$

Consequently, for $t < -1$, $\arg\sqrt{t^2 - 1} = \pi$, and the function (128.5) becomes

$$\zeta = \log(|t|e^{i\pi} + e^{i\pi}\sqrt{t^2 - 1}) - i\pi$$
$$= \log(|t| + \sqrt{|t|^2 - 1}) \tag{128.6}$$

This maps the part of the negative real axis in the t plane for which $t < -1$ into the positive real axis ba in the ζ plane as it should.

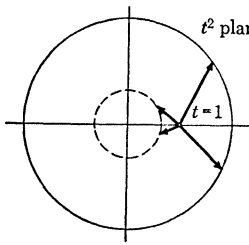


FIG. 128.2

The auxiliary variable t may be eliminated from (128.6) by solving for t from eq. (128.4). This gives

$$t = e^{(w-D)/C} \tag{128.7}$$

Since the points $t = -1$ and $t = 1$ lie on the boundary stream lines, i.e., at points b and b', respectively, we may determine the width of the strip in the w plane by requiring

$$-1 = e^{(w_2 - D)/C} \qquad +1 = e^{(w_1 - D)/C}$$

or $$\frac{w_2 - D}{C} = i\pi \qquad \frac{w_1 - D}{C} = 0$$

from which

$$w_2 - w_1 = iC\pi \quad \text{and} \quad \Phi_2 + i\Psi_2 - \Phi_1 - i\Psi_1 = iC\pi$$

If we make the velocity potential curve $\Phi = 0$ pass through the points (b,b'), then $\Phi_1 = \Phi_2 = 0$ in the above equation, and C is real. For convenience, we may take $\Psi_1 = 0$, when $\Psi_2 = C\pi$, and the fluid region in the z plane maps into the strip between the lines $w = \Phi$ and $w = \Phi + iC\pi$, with the point b lying on the bounding stream line $w = \Phi + iC\pi$, and b' lying on the line $w = \Phi$.

From eqs. (128.3), (128.5), and (128.7),

$$s = \frac{e^{-i\pi}}{v_0}\left[e^{(w-D)/C} + \sqrt{e^{2(w-D)/C} - 1}\right] \tag{128.8}$$

The w plane is finally mapped onto the z plane by making use of the

above result in eq. (128.2), whereupon

$$z = \frac{e^{-i\pi}}{v_0}\int [e^{(w-D)/C} + \sqrt{e^{2(w-D)/C} - 1}]dw + \text{constant}$$

Since the real constant C merely changes the scale of length in the w plane, and the constant D produces a translation of the plane, we may for convenience change the scale in the z plane and translate it instead of the w plane. Then

$$\frac{z}{C'} - D' = \frac{e^{-i\pi}}{v_0}\int (e^w + \sqrt{e^{2w} - 1})dw$$

$$= \frac{e^{-i\pi}}{v_0}\left(e^w + \sqrt{e^{2w} - 1} - i \log \frac{\sqrt{e^{2w} - 1} + i}{e^w}\right) \qquad (128.9)$$

Now the boundary stream lines in the w plane are given by $\Psi = 0$ and $\Psi = i\pi$, with the point b' lying on the former and the point b on $\Psi = i\pi$. The appropriate value of the many-valued function $\sqrt{e^{2w} - 1}$ must be found from the same considerations as those leading to eq. (128.6). The constants C' and D' in the above equation can be determined by locating the points b and b' in the z plane. It is simplest to let b and b' correspond to $z = -d$ and $z = +d$, respectively, which points must, as mentioned above, map into $w = i\pi$ and $w = 0$, thus requiring that

$$\frac{-d}{C'} - D' = \frac{1}{v_0}\left(1 + \frac{\pi}{2}\right)$$

$$\frac{d}{C'} - D' = -\frac{1}{v_0}\left(1 + \frac{\pi}{2}\right)$$

From these two equations, it is readily found that

$$D' = 0 \quad \text{and} \quad C' = -\frac{v_0 d}{(1 + \pi/2)}$$

The values of these constants used in eq. (128.9) yield the following final mapping function.

$$z = \frac{d}{1 + \pi/2}\left(e^w + \sqrt{e^{2w} - 1} - i \log \frac{\sqrt{e^{2w} - 1} + i}{e^w}\right) \qquad (128.10)$$

which provides the solution of the flow problem in question.

The contraction in the jet can be found by determining the real part of z at the point c' which corresponds to $w = -\infty$. From the above equation,

$$z = \frac{\pi d}{\pi + 2} - i\infty \quad \text{at} \quad c'$$

The coefficient of contraction is therefore $\pi/(\pi + 2) = 0.611$.

SUGGESTED REFERENCES

A. Hurwitz and R. Courant, *Allgemeine Funktionentheorie und elliptische Funktionen*, Berlin, Julius Springer, 1925.

R. Rothe, F. Ollendorff, and K. Pohlhausen, *Theory of Functions as Applied to Engineering Problems*, Cambridge, Mass., Technology Press, Mass. Inst. of Tech., 1933.

L. Lichtenstein, *Neuere Entwicklung der Potentialtheorie, Encyklopedie der mathematischen Wissenschaften:* Vol. II.3, "Analysis," B. G. Teuber, Leipzig, p. 218. 1919.

H. Jeffreys and Bertha S. Jeffreys, *Method of Mathematical Physics*, London, Cambridge University Press, 1946.

C. Caratheodory, *Conformal Representation*, London, Cambridge University Press, 1932.

PROBLEMS

119.1. Find the equipotentials, and the electric field between the two cylinders whose traces are the hyperbolas $(x/a)^2 - (y/b)^2 = 1$, when the potentials of the cylinders are $\pm V_0$. Find the charge density on the surface of the cylinders. Is it proportional to the curvature of the surface?

119.2. Use the transformation $z = w^{3/2}$ to find the electric field outside a wedge-shaped conductor whose dihedral angle is a right angle.

119.3. Show that the function $z = ia \cot w/2$ maps the line $V = V_0$ into the circle

$$(x - a \coth V_0)^2 + y^2 = a^2 \operatorname{csch}^2 V_0.$$

Hence show that the lines of force between two right circular cylinders are circles.

119.4. Find the curves into which the function $z = e^w + w$ maps the lines $V = \pi$ and $V = -\pi$. Hence find and sketch the equipotentials in the region near the edge of a parallel plate capacitor. Find the electric field as a function of position at the surface of one of the plates, and on the plane midway between the plates.

119.5. In the parallel plate capacitor above, the field at the edge of the plates becomes infinite. Since this effect is undesirable in capacitors used for high voltages, suppose that instead the plates are made in the shape of one of the equipotentials in the problem above. Discuss the field at the surface of such a plate and find that equipotential which has the "most uniform" surface field.

120.1. Find the potential inside a circular cylinder of radius R, when the potential on the surface is given by $V_0 \sin n\theta$.

120.2. The linear fractional transformation

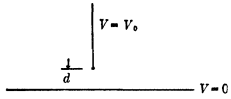
$$w = \frac{(\alpha z + \beta)}{(\alpha z + \delta)}$$

maps circles into circles, provided a straight line is considered a degenerate circle. Furthermore, the transformation is completely determined when the mapping of three points is given. Find the linear fractional transformation which maps the points $z = -1, 1, \infty$ into the points $w = -1, 1, i$, respectively. Of what problem does this give a solution?

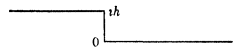
122.1. Use the Schwarz-Christoffel formula to find the mapping function of Problem 119.2.

122.2. Use the Schwarz-Christoffel formula to find the mapping function of Problem 119.4.

122.3. Find the electric potential around the configuration shown.



PROB. 122.3



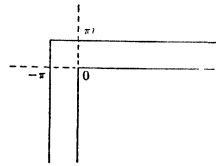
PROB. 122.4

122.4. Map the step function into the upper half plane.

122.5. Map the right-angle strip into the upper half plane.

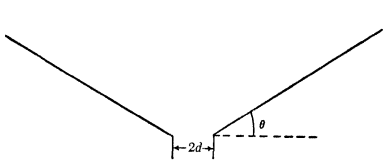
126.1. Find the velocity potential and stream function for flow around a cylinder when the velocity at infinity is $v_0 e^{i\varphi}$.

127.1. Using the function $z = w + a^2/w$ which maps the circle $|w| = a$ into the line segment between $z = -2a$ and $z = 2a$, find the velocity potential and stream function for flow around a strip when the velocity at infinity is $v_0 e^{i\varphi}$. Find the torque on the strip and show that the strip tends to be perpendicular to the line of flow. Find the torque if v_0 is replaced by a sinusoidal function of time. This is the principle of the Rayleigh disk used to measure the strength of sound.



PROB. 122.5

127.2. Show that a circle through the point $z = -1$ enclosing $z = 1$ is mapped into an airfoil-shaped figure by the function $w = z + (1/z)$. This is the Joukowski airfoil. Find the lines of flow around the airfoil for a general angle of incidence. Plot a few of the flow lines for a convenient choice of the parameters. Set up the integral giving the force on the airfoil and carry it as far as you can.



PROB. 128.1

128.1. Determine the flow pattern through a slit of width $2d$ in the bottom of a semi-infinite reservoir with the shape shown above.

Chapter 10

ALGEBRA OF LINEAR EQUATIONS, TRANSFORMATIONS AND QUADRATIC FORMS

129. Permutations and determinants. When a collection of n things which may be designated by $a_1, a_2, a_3, \ldots, a_n$ is arranged in any order, the arrangement is called a permutation. If they are arranged in order of increasing subscripts, one may consider this the natural order. In one and only one permutation are the objects arranged in their natural order. With n objects there are $n!$ permutations or distinct arrangements.

Any two elements of a collection arranged in their natural order is called a permanence. Any two elements arranged in the reverse of their natural order is called an inversion. According to this, the permutation $a_5\ a_1\ a_4\ a_3\ a_6\ a_2$ is made up of seven permanences, $a_5\ a_6$, $a_1\ a_4$, $a_1\ a_3$, $a_1\ a_6$, $a_1\ a_2$, $a_4\ a_6$, $a_3\ a_6$, and eight inversions. All permutations of n things may be divided into two main classes, one made up of all permutations with an odd number of inversions, the other containing all permutations having an even number of inversions. A permutation changes its class whenever any two of its members are interchanged. Furthermore, of all possible permutations of n objects, one-half have an even number of inversions and one-half have an odd number.

A determinant of order n can conveniently be defined in terms of an algebraic sum of all permutations of n quantities. The definition can be made clearer by first defining a determinant of order three in terms of the nine quantities

$$\begin{matrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{matrix}$$

The determinant of order three is defined as the algebraic sum of all the permutations that can be formed from the product $a_{11}\ a_{22}\ a_{33}$ by rearrangement of the second subscripts, a positive sign accompanying those permutations with an even number of inversions and a negative

sign accompanying those with an odd number. The six possible permutations in this case are given in the following table.

Permutation	*Number of Inversions*	*Algebraic Sign*
$a_{11}\ a_{22}\ a_{33}$	0 even	+
$a_{11}\ a_{23}\ a_{32}$	1 odd	−
$a_{12}\ a_{21}\ a_{33}$	1 odd	−
$a_{12}\ a_{23}\ a_{31}$	2 even	+
$a_{13}\ a_{21}\ a_{32}$	2 even	+
$a_{13}\ a_{22}\ a_{31}$	3 odd	−

In accordance with the above definition, the value of the determinant D, often written

$$D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

is $$D = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$$

The general definition of the determinant of n^2 quantities may be expressed by the equation

$$D = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \sum (-1)^h a_{1j_1} a_{2j_2} \dots a_{nj_n} \qquad (129.1)$$

where the summation is to be taken over all possible permutations $j_1, j_2, \dots, j_n$ of the integers $1, 2, \dots, n$, and h is the number of inversions occurring in each permutation.

130. Certain properties of determinants. Several useful facts concerning determinants are listed below. Proofs of their validity can be found in any standard treatment of algebra.

1. The value of a determinant is not changed when its rows and columns are interchanged.

2. The algebraic sign of the determinant is changed when any two rows (or columns) are interchanged.

3. When the elements in any two rows (or columns) are identical, the determinant has the value zero.

4. If each element in any row (or column) is multiplied by the same

factor, the resulting determinant is equivalent to the original determinant multiplied by that factor.

5. From (3) and (4) it follows that if one row (or column) is proportional to another row (or column) the determinant is zero.

6. If every element in any row (or column) can be resolved into the sum of two quantities, the determinant can be expressed as the sum of two determinants. This means that

$$\begin{vmatrix} a_{11}+\alpha_{11} & a_{12} & \dots & a_{1n} \\ a_{21}+\alpha_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{n1}+\alpha_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} + \begin{vmatrix} \alpha_{11} & a_{12} & \dots & a_{1n} \\ \alpha_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \alpha_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \tag{130.1}$$

7. If the elements of one row (or column) are equal to the sums of the corresponding elements in any number of other rows (or columns) each multiplied by a factor, the resulting determinant will have the value zero. This means that

$$\begin{vmatrix} \alpha a_{11}+\beta a_{12} & a_{11} & a_{12} & \dots & a_{1n-1} \\ \alpha a_{21}+\beta a_{22} & a_{21} & a_{22} & \dots & a_{2n-1} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \alpha a_{n1}+\beta a_{n2} & a_{n1} & a_{n} & \dots & a_{nn-1} \end{vmatrix} = \alpha \begin{vmatrix} a_{11} & a_{11} & a_{12} & \dots & a_{1n-1} \\ a_{21} & a_{21} & a_{22} & \dots & a_{2n-1} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n1} & a_{n2} & \dots & a_{nn-1} \end{vmatrix} + \beta \begin{vmatrix} a_{12} & a_{11} & a_{12} & \dots & a_{1n-1} \\ a_{22} & a_{21} & a_{22} & \dots & a_{2n-1} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{n2} & a_{n1} & a_{n2} & \dots & a_{nn-1} \end{vmatrix} = 0 \tag{130.2}$$

8. A determinant is unchanged when to each element of any row (or column) are added those of several other rows (or columns) multiplied, respectively, by constant factors. Consequently, we have the identity

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11}+\alpha a_{12}+\beta a_{13} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21}+\alpha a_{22}+\beta a_{23} & a_{22} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1}+\alpha a_{n2}+\beta a_{n3} & a_{n2} & a_{n3} & \dots & a_{nn} \end{vmatrix} \tag{130.3}$$

This property of a determinant is especially useful in facilitating the evaluation of certain determinants, because it often allows one to obtain a row or column with a considerable number of zeros in it.

9. The value of a determinant is unchanged by augmenting it with rows or columns of zeros, provided the new terms on the principal diagonal are ones. The general situation concerning the identity of a determinant and the augmented determinant is clearly indicated by the equation

$$\begin{vmatrix} & & & \\ a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ X & 1 & 0 & 0 & \dots & 0 \\ X & X & a_{11} & a_{12} & \dots & a_{1n} \\ X & X & a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X & X & a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \tag{130.4}$$

where positions marked X can be filled with arbitrary elements.

10. The product of two determinants of order n can be expressed as a single determinant of the same order. In order to find the element that occurs in the ith row and jth column in the product determinant, one multiplies each element in the ith row of the first determinant by the corresponding element in the jth column of the second, and adds them all together. This process thus gives

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \cdot \begin{vmatrix} b_{11} & \dots & b_{1j} & \dots & b_{1n} \\ b_{21} & \dots & b_{2j} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ b_{n1} & \dots & b_{nj} & \dots & b_{nn} \end{vmatrix}$$

$$= \begin{vmatrix} \sum_k a_{1k}b_{k1} & \sum_k a_{1k}b_{k2} & \dots & \sum_k a_{1k}b_{kn} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \sum_k a_{ik}b_{kj} & \dots \\ \dots & \dots & \dots & \dots \\ \sum_k a_{nk}b_{k1} & \sum_k a_{nk}b_{k2} & \dots & \sum_k a_{nk}b_{kn} \end{vmatrix} \tag{130.5}$$

Determinants obey the commutative law of multiplication.

131. Laplace's development by minors. The evaluation of determinants of order higher than three can be conveniently carried out by a method due to Laplace. This method requires the use of minors of a determinant. A *minor* of a determinant of order n is defined as the determinant formed from the elements left over after striking out the elements in any p rows and columns of the original determinant. The minor is therefore a determinant of order $n - p$, and is referred to as a pth minor. A second minor of the determinant

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{vmatrix} \tag{131.1}$$

is

$$\begin{vmatrix} a_{31} & a_{34} & \dots & a_{3n} \\ a_{41} & a_{44} & \dots & a_{4n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n4} & \dots & a_{nn} \end{vmatrix} \tag{131.2}$$

This particular second minor may be referred to as the second minor with respect to the determinant

$$\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \tag{131.3}$$

On the other hand, the determinant (131.3) is the $(n-2)$ minor with respect to the determinant (131.2). The minors (131.2) and (131.3) are called *complementary minors.* When a minor contains the elements from the same columns as rows it is called a *principal minor.*

If the determinant (131.1) is denoted by A, then a minor will be denoted by $A_{l_1 l_2 \dots l_j}^{k_1 k_2 \dots k_j}$ when it contains the elements from rows $k_1, k_2, \dots, k_j$ and columns $l_1, l_2, \dots, l_j$. The complement of the minor $A_{l_1 l_2 \dots l_j}^{k_1 k_2 \dots k_j}$ may be denoted by $A_{l_{j+1} \dots l_n}^{k_{j+1} \dots k_n}$. In this notation the minor (131.3) is A_{23}^{12} and its complement (131.2) is A_{14}^{34}. The *algebraic complement* of the minor $A_{l_1 \dots l_j}^{k_1 \dots k_j}$ is defined by the relation

$$\text{algebraic complement of } A_{l_1 \dots l_j}^{k_1 \dots k_j} = (-1)^h A_{l_{j+1} \dots l_n}^{k_{j+1} \dots k_n}$$

where h is the sum of the number of inversions in the permutation $l_1, l_2, \dots, l_n$ and the permutation $k_1, k_2, \dots, k_n$ (cf. Sec. 129) or h is the number of transpositions of the indexes $l_1, \dots, l_n$ required to make the order $l_1, l_2, \dots, l_n$ coincide with the order $k_1, k_2, \dots, k_n$.

According to the Laplace method the value of a determinant of order n can be found as follows. Select any j rows and any j columns in a determinant A and form all the different minors of order j that can be formed from the elements in the j rows and columns. The value of A is then the sum of the products of each of these minors and its algebraic complement. There will be $n!/j!(n-j)!$ minors of order j, and consequently this number of terms in the sum.

The Laplace development of the determinant A just explained can be conveniently expressed in terms of the minor notation given above by writing

$$A = \sum (-1)^{k_1+k_2+\ldots+k_l+l_1+\ldots+l_l} A^{k_1 \ldots k_l}_{l_1 \ldots l_l} \times A^{k_{l+1} \ldots k_n}_{l_{l+1} \ldots l_n} = \sum (-1)^h A^{k_1 \ldots k_l}_{l_1 \ldots l_l} A^{k_{l+1} \ldots k_n}_{l_{l+1} \ldots l_n} \tag{131.4}$$

where h is defined above and the sum is to be taken over all the combinations of the k integers $l_1, l_2, \ldots, l_k$ which can be formed from the n integers $1, 2, \ldots, n$.

The Laplace method of evaluating the determinant

$$A = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}$$

will be illustrated by developing it with respect to the first two rows. The development (131.4) can now be written in the explicit form

$$A = A^{12}_{12}A^{34}_{34} - A^{12}_{13}A^{34}_{24} + A^{12}_{14}A^{34}_{23} + A^{12}_{23}A^{34}_{14} - A^{12}_{24}A^{34}_{13} + A^{12}_{34}A^{34}_{12}$$

or

$$A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \cdot \begin{vmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{vmatrix} - \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \cdot \begin{vmatrix} a_{32} & a_{34} \\ a_{42} & a_{44} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{14} \\ a_{21} & a_{24} \end{vmatrix} \cdot \begin{vmatrix} a_{32} & a_{33} \\ a_{42} & a_{43} \end{vmatrix} + \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \cdot \begin{vmatrix} a_{31} & a_{34} \\ a_{41} & a_{44} \end{vmatrix} - \begin{vmatrix} a_{12} & a_{14} \\ a_{22} & a_{24} \end{vmatrix} \cdot \begin{vmatrix} a_{31} & a_{33} \\ a_{41} & a_{43} \end{vmatrix} + \begin{vmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{vmatrix} \cdot \begin{vmatrix} a_{31} & a_{32} \\ a_{41} & a_{42} \end{vmatrix}$$

In actually evaluating a determinant, it is often advantageous to use a combination of the properties given in Secs. 130 and 131. For example, suppose we are required to evaluate the determinant

$$D = \begin{vmatrix} 3 & 1 & 7 & 4 & 2 \\ -2 & 3 & 0 & 2 & 1 \\ 5 & -1 & 3 & 8 & 3 \\ 6 & 4 & 1 & -2 & 5 \\ 11 & -5 & -1 & 6 & -3 \end{vmatrix}$$

According to rule 8, Sec. 130, the value of the determinant is unchanged if we multiply the fifth column by 2 and subtract the result from column 4. This gives

$$D = \begin{vmatrix} 3 & 1 & 7 & 0 & 2 \\ -2 & 3 & 0 & 0 & 1 \\ 5 & -1 & 3 & 2 & 3 \\ 6 & 4 & 1 & -12 & 5 \\ 11 & -5 & -1 & 12 & -3 \end{vmatrix}$$

The object of this procedure was to introduce two zeros in the fourth column. Two more zeros can be introduced by adding row 5 to row 4, thereby obtaining

$$D = \begin{vmatrix} 3 & 1 & 7 & 0 & 2 \\ -2 & 3 & 0 & 0 & 1 \\ 5 & -1 & 3 & 2 & 3 \\ 17 & -1 & 0 & 0 & 2 \\ 11 & -5 & -1 & 12 & -3 \end{vmatrix}$$

Because of the zeros in rows 2 and 4, it is advantageous to develop the determinants according to minors based on the elements of the second and fourth rows. These nonzero minors are A_{12}^{24}, A_{15}^{24}, and A_{25}^{24}. From the relation (131.4) we may write

$$D = \begin{vmatrix} -2 & 3 \\ 17 & -1 \end{vmatrix} \cdot \begin{vmatrix} 7 & 0 & 2 \\ 3 & 2 & 3 \\ -1 & 12 & -3 \end{vmatrix} + \begin{vmatrix} -2 & 1 \\ 17 & 2 \end{vmatrix} \cdot \begin{vmatrix} 1 & 7 & 0 \\ -1 & 3 & 2 \\ -5 & -1 & 12 \end{vmatrix} - \begin{vmatrix} 3 & 1 \\ -1 & 2 \end{vmatrix} \cdot \begin{vmatrix} 3 & 7 & 0 \\ 5 & 3 & 2 \\ 11 & -1 & -3 \end{vmatrix}$$

The above determinants of orders 2 and 3 can be readily evaluated.

132. Linear equations. Determinants, discussed in the preceding two sections, play an important role in the theory of linear equations. When a system of m linear equations

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= k_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= k_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= k_m \end{aligned}\right\} \quad (132.1)$$

involving n unknowns $x_1, \ldots, x_n$ and the known constants a_{ij} and $k_1, \ldots, k_m$ is given, the following three situations may arise.

1. No solution exists, in which case they are said to be inconsistent.
2. The system of equations possesses just one solution.
3. The system may have more than one solution, in which case they necessarily have an infinite number of solutions.

In order to ascertain which of these situations obtains we must consider the following two arrays

$$\mathbf{A} = \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots\ldots & \ldots\ldots & \ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots & \ldots\ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{Vmatrix}; \quad \mathbf{A}' = \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1n} & k_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & k_2 \\ \ldots\ldots & \ldots\ldots & \ldots & \ldots\ldots & \ldots \\ \ldots\ldots & \ldots\ldots & \ldots & \ldots\ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & k_m \end{Vmatrix} \quad (132.2)$$

of the constants and coefficients which occur in the system (132.1). Each array is called a matrix (cf. Sec. 18). In particular, $\mathbf{A}$ is called the matrix of the system of equations, and $\mathbf{A}'$ is called the augmented matrix.

As explained in detail in Sec. 18, a matrix is said to be of rank r if, among its elements, it contains one r-rowed determinant which is not zero, while all determinants of higher order which can be formed from the elements of the matrix are zero.

Which of the three situations mentioned above obtains can be concisely given in terms of the ranks of $\mathbf{A}$ and $\mathbf{A}'$.

1. The system of eqs. (132.1) is consistent if and only if the rank of $\mathbf{A}$ equals the rank of $\mathbf{A}'$.

2. The system has only one solution if rank $\mathbf{A}$ = rank $\mathbf{A}'$ = n. This requires that $m = n$, i.e., there must be the same number of equations as unknowns, and the determinant of the n^2 coefficients must not vanish.

3. If rank $\mathbf{A}$ = rank $\mathbf{A}'$ = $r \leq n$, then $n - r$ unknowns can be assigned arbitrary values, and the remaining r unknowns will then be uniquely determined. The $n - r$ unknowns can be assigned in any way, provided that the matrix of the coefficients of the remaining unknowns is of rank r.

The actual solution of a system of eqs. (132.1) for which rank $\mathbf{A}$ = rank $\mathbf{A}'$ = n can be conveniently found either by Cramer's rule or by the elimination procedure of Gauss.

Cramer's Rule. If

$$A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \tag{132.3}$$

is the determinant of the coefficients of the system of equations

$$\sum_{j=1}^{n} a_{ij}x_j = k_i \qquad i = 1, 2, \dots, n \tag{132.4}$$

for which rank $\mathbf{A}$ = rank $\mathbf{A}'$ = n, Cramer's rule states that

$$x_j = \frac{A_j}{A} \qquad j = 1, 2, \dots, n \tag{132.5}$$

where A_j is the determinant

$$A_j = \begin{vmatrix} a_{11} & \dots & a_{1j-1} & k_2 & a_{1j+1} & \dots & a_{1n} \\ a_{21} & \dots & a_{2j-1} & k_2 & a_{2j+1} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{nj-1} & k_n & a_{nj+1} & \dots & a_{nn} \end{vmatrix} \tag{132.6}$$

Gauss's Elimination Procedure. When the coefficients in the system of eqs. (132.4) are numerical it is convenient to use a slide rule or computing machine in connection with the following elimination procedure. In the system of equations

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ldots + a_{1n}x_n &= k_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \ldots + a_{2n}x_n &= k_2 \\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \ldots + a_{3n}x_n &= k_3 \\
\ldots\ldots\ldots

1. If $m \geq n$ and rank $\mathbf{A} = n$, the homogeneous system will have no solution other than $x_1 = x_2 = \ldots = x_n = 0$.

2. The homogeneous system will have a solution other than $x_1 = x_2 = \ldots = x_n = 0$ if and only if rank $\mathbf{A} < n$.

3. If rank $\mathbf{A} = r < n$, values of $n - r$ of the unknowns can be assigned arbitrarily, after which r unknowns will be uniquely determined. Consequently, when the homogeneous system has a solution different from the trivial one $x_1 = x_2 = \ldots = x_n = 0$, it necessarily has an infinite number, since the value of at least one unknown can be arbitrarily assigned. This is equivalent to saying that at best only the ratios $x_1/x_j, x_2/x_j, \ldots, x_n/x_j$ can be uniquely determined.

134. Vectors and vector spaces. It is useful to have a definition of a vector which will permit the use of geometrical concepts and language when the number of dimensions is greater than three. With this in mind, we shall define the vector $\boldsymbol{x}$, to mean the set of n ordered numbers $(x_1, x_2, \ldots, x_n)$. The vector $\boldsymbol{x}$ is not completely specified unless all these n numbers are given and given in their proper order. The vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is different from the vector $\boldsymbol{x}' = (x_2, x_1, \ldots, x_n)$. The subject of vector analysis deals with the calculus of these ordered number sets. The numbers $x_1, x_2, \ldots, x_n$ are called the components of the vector.

Two fundamental operations in the calculus of vectors are the multiplication of a vector by a number c and the addition of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. If

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \tag{134.1}$$

we define $c\boldsymbol{x}$ to be

$$c\boldsymbol{x} = (cx_1, cx_2, \ldots, cx_n) \tag{134.2}$$

where c may be a real or a complex number. The definition of the addition of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is contained in the equation

$$\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \tag{134.3}$$

Thus $\boldsymbol{x} + \boldsymbol{y}$ is a new vector whose components are found by adding the components of each of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Equation (134.3) also contains the law of subtraction of two vectors.

The product of two vectors can be defined in several ways. The most useful product for the present exposition is the scalar product, denoted by $(\boldsymbol{x} \cdot \boldsymbol{y})$, and defined by the equation

$$(\boldsymbol{x} \cdot \boldsymbol{y}) = x_1y_1 + x_2y_2 + \ldots + x_ny_n = (\boldsymbol{y} \cdot \boldsymbol{x}) \tag{134.4}$$

When, in particular, the scalar product $(\boldsymbol{x} \cdot \boldsymbol{y})$ is zero, the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be *perpendicular* to one another or merely *orthogonal.*

An important special case of the scalar product arises when $\boldsymbol{y} = \boldsymbol{x}$, for which

$$(\boldsymbol{x}\cdot\boldsymbol{x}) = x_1^2 + x_2^2 + \ldots + x_n^2 \tag{134.5}$$

The quantity $+\sqrt{(\boldsymbol{x}\cdot\boldsymbol{x})}$ is called the length or magnitude of the vector $\boldsymbol{x}$. The similarity of the expression (134.5) to the square of the length of the radius vector in three dimensions is evident.

It may readily be shown that

$$(\boldsymbol{x}\cdot\boldsymbol{y})^2 \leq (\boldsymbol{x}\cdot\boldsymbol{x})(\boldsymbol{y}\cdot\boldsymbol{y}) \tag{134.6}$$

This is called the Schwarz inequality. In terms of the components of x and y, this relation becomes

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right)$$

We shall next consider linear combinations of a number of vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$. It will now be necessary to label the components of these vectors with two subscripts; thus x_{ij} denotes the jth component of the vector $\boldsymbol{x}_i$, so that

$$\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \qquad i = 1, 2, \ldots, m$$

These m vectors are said to be *linearly independent*, provided no m numbers $c_1, c_2, \ldots, c_m$ not all zero exist for which

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \ldots + c_m\boldsymbol{x}_m = 0 \tag{134.7}$$

This means that the zero vector cannot be formed by a linear combination of linearly independent vectors except in the trivial case $c_1 = c_2 = \ldots c_m = 0$. When m numbers $c_1, c_2, \ldots, c_m$ do exist which satisfy (134.7), the m vectors are *linearly dependent*.

The aggregate of vectors formed by all possible linear combinations of n linearly independent vectors constitutes an *n-dimensional linear vector space*. One may also say that n such vectors span an n-dimensional vector space. Thus any n linearly independent set of vectors forms a representative of the space in the sense that any other vector in that space can be represented by a linear combination of them. Such a group of vectors representing the space are sometimes called basic vectors to indicate that all other vectors are to be expressed in terms of them. The selection of the basic vectors is equivalent to a selection of a coordinate system in the space.

It is often convenient to choose the basic vectors representing a given space so that they have special properties which make them well adapted to the problem in hand. For instance, it is often desirable that the basic system of vectors be an *orthonormal set*, that is to say, vectors which have unit length and are mutually perpendicular. This property is reflected in the equations

$$(\boldsymbol{u}_i\cdot\boldsymbol{u}_j) = \begin{matrix} 0 & i \neq j \\ 1 & i = j \end{matrix} \tag{134.8}$$

Unless otherwise specified, the letters $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, ... , etc., will denote an orthonormal set of basic vectors.

It is easy to demonstrate that vectors of an orthonormal set are necessarily linearly independent, that is the relation

$$c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \ldots + c_j\boldsymbol{u}_j + \ldots + c_n\boldsymbol{u}_n = 0$$

will be satisfied only when all the constants c are zero. The fact that the vectors $\boldsymbol{u}$ are orthonormal provides a simple means of actually calculating the values of the coefficients. The coefficient c_j may be calculated by forming the scalar product of the linear combination and $\boldsymbol{u}_j$. This gives

$$\begin{aligned}[\boldsymbol{u}_j \cdot (c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \ldots + c_j\boldsymbol{u}_j + \ldots + c_n\boldsymbol{u}_n)] \\ = c_j(\boldsymbol{u}_j \cdot \boldsymbol{u}_j) = c_j\end{aligned}$$

Therefore

$$c_j = 0 \qquad j = 1, 2, \ldots, n$$

135. Orthogonalization and completeness. If the components of a unit vector $\boldsymbol{u}_i$ are u_{ij} where $j = 1, 2, \ldots, n$, the group of vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ form an orthonormal set if

$$u_{ij} = \begin{matrix} 0 & i \neq j \\ 1 & i = j \end{matrix}$$

as is at once obvious from the relations

$$\begin{aligned}(\boldsymbol{u}_i \cdot \boldsymbol{u}_j) &= u_{i1}u_{j1} + u_{i2}u_{j2} + \ldots + u_{in}u_{jn} \\ &= 0 \quad \text{when} \quad i \neq j \\ &= 1 \quad \text{when} \quad i = j\end{aligned}$$

On the other hand, if the vector space under consideration is represented by the linearly independent set of basic vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, which however do not form an orthonormal set, an orthonormal set can be obtained from them by the following process. First let

$$\boldsymbol{u}_1 = \frac{\boldsymbol{x}_1}{\sqrt{(\boldsymbol{x}_1 \cdot \boldsymbol{x}_1)}} \tag{135.1}$$

thereby making $(\boldsymbol{u}_1 \cdot \boldsymbol{u}_1) = 1$. A vector $\boldsymbol{u}_2$ must be formed which is orthogonal to $\boldsymbol{u}_1$ and of unit length. This can readily be done by letting

$$\boldsymbol{u}_2 = c_1\boldsymbol{u}_1 + c_2\boldsymbol{x}_2$$

and determining the constants c_1 and c_2 so that

$$(\boldsymbol{u}_1 \cdot \boldsymbol{u}_2) = c_1(\boldsymbol{u}_1 \cdot \boldsymbol{u}_1) + c_2(\boldsymbol{u}_1 \cdot \boldsymbol{x}_2) = 0$$

and

$$\begin{aligned}(\boldsymbol{u}_2 \cdot \boldsymbol{u}_2) &= [(c_1\boldsymbol{u}_1 + c_2\boldsymbol{x}_2) \cdot (c_1\boldsymbol{u}_1 + c_2\boldsymbol{x}_2)] \\ &= |\, c_1\boldsymbol{u}_1 + c_2\boldsymbol{x}_2 \,|^2 = 1\end{aligned}$$

From the last two equations it follows that

$$c_1 = -(\boldsymbol{u}_1 \cdot \boldsymbol{x}_2)c_2 \qquad c_2 = \frac{1}{|\,\boldsymbol{x}_2 - (\boldsymbol{u}_1 \cdot \boldsymbol{x}_2)\boldsymbol{u}_1\,|}$$

Consequently,

$$\boldsymbol{u}_2 = \frac{\boldsymbol{x}_2 - (\boldsymbol{u}_1 \cdot \boldsymbol{x}_2)\boldsymbol{u}_1}{|\,\boldsymbol{x}_2 - (\boldsymbol{u}_1 \cdot \boldsymbol{x}_2)\boldsymbol{u}_1\,|} \tag{135.2}$$

has unit length and is orthogonal to $\boldsymbol{u}_1$. The process can be continued by letting $\boldsymbol{u}_3$ be

$$\boldsymbol{u}_3 = c_1'\boldsymbol{u}_1 + c_2'\boldsymbol{u}_2 + c_3'\boldsymbol{x}_3$$

and determining the three constants c_1', c_2', and c_3' so that $(\boldsymbol{u}_3 \cdot \boldsymbol{u}_3) = 1$ and $\boldsymbol{u}_3$ will be orthogonal to both $\boldsymbol{u}_2$ and $\boldsymbol{u}_1$.

Since any linearly independent set of basic vectors can be converted into an orthonormal basic system, one may always regard a vector space as spanned by an orthonormal set of basic vectors.

When the number of orthonormal basic vectors is less than the dimension number of the space, one speaks of an incomplete orthonormal system of basic vectors. Even if the orthonormal system is incomplete, the inequality,

$$(\boldsymbol{x} \cdot \boldsymbol{x}) \geq x_1^2 + x_2^2 + \ldots + x_m^2 \tag{135.3}$$

known as Bessel's inequality, is satisfied, where $\boldsymbol{x}$ is an arbitrary vector in an n-dimensional space, and x_1, x_2, etc., are the components of $\boldsymbol{x}$ with respect to the orthonormal system $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m$, where $m \leq n$. Thus

$$x_i = (\boldsymbol{u}_i \cdot \boldsymbol{x})$$

The inequality (135.3) is the generalization of the theorem of Pythagorus so familiar in a space of three dimensions.

The relation (135.3) forms a convenient criterion of completeness of the orthonormal basic system of vectors in the sense that the orthonormal set is complete only if the equality in the relation (135.3) holds; otherwise the set is incomplete. An advantage of this criterion is that it can be extended to spaces having an infinite number of dimensions where the dimension number of the space looses its physical significance.

136. Orthogonal and unitary spaces. An orthogonal vector space is spanned by an orthonormal system of basic vectors, so that an arbitrary vector $\boldsymbol{x}$ is represented by

$$\boldsymbol{x} = x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2 + \ldots + x_n\boldsymbol{u}_n \tag{136.1}$$

where the components $x_1, x_2, \ldots, x_n$ may assume all values in the continuum of real numbers, and the square of the length or absolute magnitude of the vector is given by

$$\boldsymbol{x}^2 = (\boldsymbol{x}\cdot\boldsymbol{x}) = x_1^2 + x_2^2 + \ldots + x_n^2 \tag{136.2}$$

In various applications it is necessary to work with a somewhat more general space than the orthogonal space, by permitting the components $(z_1, z_2, \ldots, z_n)$ of an arbitrary vector in this space to assume the continuum of all complex numbers $z_k = x_k + iy_k$. The space spanned by

$$z_1\boldsymbol{u}_1 + z_2\boldsymbol{u}_2 + \ldots + z_n\boldsymbol{u}_n$$

is called the *unitary space* of n dimensions. The fundamental laws of addition and multiplication by a number are defined by eqs. (134.2) and (134.3), when the indicated operations are carried out, using the laws of addition and multiplication of complex numbers (cf. Sec. 71) instead of real numbers.

The definition of the scalar product must be modified in the unitary space. This arises from the fact that if the square of the length or the square of the absolute magnitude of the vector is defined as the sum of the squares of the components, in accordance with eq. (136.2), the quantity thus obtained may assume negative values, or indeed could vanish without implying that each component of the vector was zero. This difficulty is entirely obviated by defining the square of the absolute magnitude of the vector $\boldsymbol{z}$ by the equation

$$|\,\boldsymbol{z}\,|^2 = (\boldsymbol{z}^*\cdot\boldsymbol{z}) = z_1^*z_1 + z_2^*z_2 + \ldots + z_n^*z_n \tag{136.3}$$

where $\qquad z_k = x_k + iy_k \quad \text{and} \quad z_k^* = x_k - iy_k$

Since $z_k^*z_k = x_k^2 + y_k^2$, the right-hand side of eq. (136.3) is always positive and can vanish only if all the individual components z_k are zero.

In line with the above definition of the absolute magnitude of a vector, the definition of the *scalar product* in the unitary space is

$$(\boldsymbol{z}\cdot\boldsymbol{s}) = z_1^*s_1 + z_2^*s_2 + \ldots + z_n^*s_n \tag{136.4}$$

As in the orthogonal space, we say that two vectors are perpendicular if their scalar product as defined by eq. (136.4) is zero.

137. Linear transformations. The subject of linear transformations may be interpreted in two ways. First we may represent a given vector $\boldsymbol{x}$, in a space of n dimensions in terms of two different systems of orthonormal basic vectors, $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$, and $\boldsymbol{u}_1', \boldsymbol{u}_2', \ldots, \boldsymbol{u}_n'$, so that

$$\boldsymbol{x} = x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2 + \ldots + x_n\boldsymbol{u}_n \tag{137.1}$$

$$= x_1'\boldsymbol{u}_1' + x_2'\boldsymbol{u}_2' + \ldots + x_n'\boldsymbol{u}_n' \tag{137.2}$$

The transition from one basic system of vectors $\boldsymbol{u}_i$ to another system $\boldsymbol{u}_i'$ constitutes a linear transformation because the transition is actually accomplished by expressing the new basic vectors $\boldsymbol{u}_i'$ in terms of the old by the linear relations

$$\left.\begin{aligned} \boldsymbol{u}_1' &= a_{11}\boldsymbol{u}_1 + a_{21}\boldsymbol{u}_2 + \ldots + a_{n1}\boldsymbol{u}_n \\ \boldsymbol{u}_2' &= a_{12}\boldsymbol{u}_1 + a_{22}\boldsymbol{u}_2 + \ldots + a_{n2}\boldsymbol{u}_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \boldsymbol{u}_n' &= a_{1n}\boldsymbol{u}_1 + a_{2n}\boldsymbol{u}_2 + \ldots + a_{nn}\boldsymbol{u}_n \end{aligned}\right\} \tag{137.3}$$

A rotation of the coordinate system in a given space is a special case of the linear transformation (137.3).

A different interpretation of a linear transformation comes about when we regard the system of basic vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ as fixed, but we wish to carry out an operation on a vector $\boldsymbol{x}$, expressed in terms of the system $\boldsymbol{u}_i$, which converts $\boldsymbol{x}$ into a new vector $\boldsymbol{x}'$ in the same space, expressed in the same system of basic vectors $\boldsymbol{u}_i$.

$$\boldsymbol{x} = x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2 + \ldots + x_n\boldsymbol{u}_n \tag{137.4}$$

$$\boldsymbol{x}' = x_1'\boldsymbol{u}_1 + x_2'\boldsymbol{u}_2 + \ldots + x_n'\boldsymbol{u}_n \tag{137.5}$$

The transition $\boldsymbol{x} \to \boldsymbol{x}'$ is accomplished by expressing the components x_i' of the new vector $\boldsymbol{x}'$ in terms of the components x_i of the old vector $\boldsymbol{x}$ by means of the linear equations

$$\left.\begin{aligned} x_1' &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ x_2' &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n' &= a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n \end{aligned}\right\} \tag{137.6}$$

Thus the transformation of one vector into another in a given space is accomplished by means of a linear relationship between the components of the two vectors.

The two points of view mentioned above can be related by converting the vector $\boldsymbol{x}$ specified by eq. (137.4) into a new vector $\boldsymbol{x}''$ expressed in terms of the basic system $\boldsymbol{u}_i'$ but retaining the same components $x_1, x_2, \ldots, x_n$ as $\boldsymbol{x}$ had in the system $\boldsymbol{u}_i$, so that

$$\boldsymbol{x}'' = x_1\boldsymbol{u}_1' + x_2\boldsymbol{u}_2' + \ldots + x_n\boldsymbol{u}_n' \tag{137.7}$$

If $\boldsymbol{x}''$ is now expressed in terms of the basic system $\boldsymbol{u}_i$ by means of eqs. (137.3),

$$\left.\begin{aligned} \boldsymbol{x}'' &= x_1\sum_{k=1}^{n} a_{k1}\boldsymbol{u}_k + x_2\sum_{k=1}^{n} a_{k2}\boldsymbol{u}_k + \ldots + x_n\sum_{k=1}^{n} a_{kn}\boldsymbol{u}_k \\ &= \left(\sum_{j=1}^{n} a_{1j}x_j\right)\boldsymbol{u}_1 + \left(\sum_{j=1}^{n} a_{2j}x_j\right)\boldsymbol{u}_2 + \ldots + \left(\sum_{j=1}^{n} a_{nj}x_j\right)\boldsymbol{u}_n \\ &= x_1'\boldsymbol{u}_1 + x_2'\boldsymbol{u}_2 + \ldots + x_n'\boldsymbol{u}_n = \boldsymbol{x}' \end{aligned}\right\} \tag{137.8}$$

The last two equalities follow from eqs. (137.6) and (137.5). This shows

that we may convert the vector $\boldsymbol{x}$ into a new vector $\boldsymbol{x}'$ either by retaining the same basic vector system $\boldsymbol{u}_i$ and transforming the components according to eqs. (137.6), or by retaining the same components and transforming the basic system of vectors according to eqs. (137.3).

It is evident that any linear transformation can be completely determined when the matrix of the coefficients a_{ij} is known (cf. Sec. 138). In particular, the transformations (137.6) and (137.3) are associated, respectively, with the two matrices

$$\mathbf{A} = \begin{Vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{Vmatrix}; \qquad \tilde{\mathbf{A}} = \begin{Vmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{Vmatrix} \tag{137.9}$$

The matrix $\tilde{\mathbf{A}}$ differs from $\mathbf{A}$ only in the fact that corresponding elements across the principal diagonal have been transferred. Consequently, $\tilde{\mathbf{A}}$ is said to be the transposed matrix of $\mathbf{A}$.

138. Matrix calculus. Operations with matrices are customarily defined to facilitate the theory of linear transformations. Because a linear transformation which converts a vector $\boldsymbol{x}$ in a space of n dimensions into another vector $\boldsymbol{x}'$ can be completely defined when the matrix $\mathbf{A}$ of the transformation coefficients is known, it is convenient to think of the vector $\boldsymbol{x}'$ as being obtained by operating on the vector $\boldsymbol{x}$ with the matrix $\mathbf{A}$. This may be indicated by writing

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{x}' \tag{138.1}$$

We may also say that "multiplication" of the vector $\boldsymbol{x}$ with the matrix $\mathbf{A}$ yields a new vector $\boldsymbol{x}'$. This equation can mean only that the components of $\boldsymbol{x}'$ are completely determined by means of eqs. (137.6). The law of multiplication of matrices can be determined by extending the formalism adopted in eqs. (138.1). To do this, let us transform the vector $\boldsymbol{x}'$ into still another vector $\boldsymbol{x}''$ with the help of the matrix

$$\mathbf{B} = \begin{Vmatrix} b_{11} & \dots & b_{1n} \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ b_{n1} & \dots & b_{nn} \end{Vmatrix}$$

so that

$$\mathbf{B}\boldsymbol{x}' = \boldsymbol{x}'' \tag{138.2}$$

The vector $\boldsymbol{x}''$ could, however, have been obtained directly from the vector $\boldsymbol{x}$ by a suitable transformation associated with the matrix $\mathbf{C}$, so that

$$\mathbf{C}\boldsymbol{x} = \boldsymbol{x}'' \tag{138.3}$$

From eqs. (138.1), (138.2), and (138.3), we must have the identity

$$\mathbf{x}'' = \mathbf{C}x = \mathbf{B}x' = \mathbf{B}(\mathbf{A}x)$$

The matrix **C** will be defined as the product of the two matrices **BA**, and will be written

$$\mathbf{C} = \mathbf{BA} \tag{138.4}$$

The actual rules of multiplication are found by determining the elements c_{ik} of the matrix **C** in terms of the known elements b_{ij} and a_{jk} of the matrices **B** and **A**. In terms of the components of x, x', x'', we have

$$x'' = \sum_{j=1}^{n} b_{ij}x_j' = \sum_{j=1}^{n} b_{ij}\left(\sum_{k=1}^{n} a_{jk}x_k\right) \qquad i = 1, 2, \ldots, n$$

$$= \sum_{k=1}^{n}\left(\sum_{j=1}^{n} b_{ij}a_{jk}\right)x_k = \sum_{k=1}^{n} c_{ik}x_k$$

Consequently, the law of multiplication must be such as to yield a matrix with the elements

$$c_{ik} = \sum_{j=1}^{n} b_{ij}a_{jk} \tag{138.5}$$

so that

$$\begin{Vmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ b_{n1} & b_{n2} & \ldots & b_{nn} \end{Vmatrix} \cdot \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{Vmatrix}$$

$$= \begin{Vmatrix} \sum_j b_{1j}a_{j1} & \sum_j b_{1j}a_{j2} & \ldots & \sum_j b_{1j}a_{jn} \\ \sum_j b_{2j}a_{j1} & \sum_j b_{2j}a_{j2} & \ldots & \sum_j b_{2j}a_{jn} \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \sum_j b_{nj}a_{j1} & \sum_j b_{nj}a_{j2} & \ldots & \sum_j b_{nj}a_{jn} \end{Vmatrix} \tag{138.6}$$

In order to find the element that occurs in the ith row and kth column of the product matrix, it is necessary only to multiply each element in the ith row of the first matrix by the corresponding element in the kth column of the second matrix and add them all together. It must be emphasized that the order of the matrix factors in a product is important, since **B·A** is not in general equal to **A·B**.

Although the commutative law of multiplication does not hold in general, the associative law

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \tag{138.7}$$

and the distributive law

$$(\mathbf{A} + \mathbf{B})\cdot\mathbf{C} = \mathbf{AC} + \mathbf{BC} \tag{138.8}$$

do hold. The last law requires the law of addition of two matrices, or more generally, the way in which a linear combination of two matrices is to be formed. The matrix representing the linear combination $\lambda\mathbf{A} + \mu\mathbf{B}$, where λ and μ are real or complex numbers, is defined as that matrix whose elements are $\lambda a_{ik} + \mu b_{ik}$.

The zero matrix is a matrix all of whose elements are zero.

The unit matrix $\mathbf{U}$ is defined so that it will play the same roll in matrix calculus as the number 1 plays in ordinary calculus. Thus

$$\mathbf{UA} = \mathbf{AU} = \mathbf{A} \tag{138.9}$$

This requirement, together with the law of multiplication (138.6), requires $\mathbf{U}$ to be

$$\mathbf{U} = \begin{Vmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & .. & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{Vmatrix} \tag{138.10}$$

Another important matrix is the inverse matrix $\mathbf{A}^{-1}$. If the matrix $\mathbf{A}$ carries a vector $\boldsymbol{x}$ over into a vector $\boldsymbol{x}'$, the matrix $\mathbf{A}^{-1}$ is that matrix which takes the vector $\boldsymbol{x}'$ back into the vector $\boldsymbol{x}$. Consequently, the successive operations $\mathbf{A}^{-1}\mathbf{A}$ or $\mathbf{AA}^{-1}$ must leave the vector $\boldsymbol{x}$ unchanged and the single matrix which, on operating on the vector $\boldsymbol{x}$ leaves it unchanged, is the unit matrix $\mathbf{U}$.

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{U} \tag{138.11}$$

The inverse $\mathbf{A}^{-1}$ of the matrix $\mathbf{A}$ always exists if the determinant A of the elements of $\mathbf{A}$ is different from zero.

The actual determination of the inverse matrix $\mathbf{A}^{-1}$, when $\mathbf{A}$ is known, rests on the determination of the coefficients which are required to express the components $x_1, \dots, x_n$ in terms of the components $x_1', \dots, x_n'$ or merely the solution of linear eqs. (137.6), regarding the components $x_1, \dots, x_n$ as unknowns. From Sec. 132, we know that a solution of these equations exists when the determinant of the coefficients does not vanish, i.e., $A = \det \mathbf{A} \neq 0$. From eq. (132.5), the solution is

$$x_i = \frac{\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1i-1}x_1' & a_{1i+1} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2i-1}x_2' & a_{2i+1} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{ni-1}x_n' & a_{ni+1} & \dots & a_{nn} \end{vmatrix}}{\det \mathbf{A}}$$

When the determinant in the numerator is expanded,

$$x_i = \frac{\det \mathbf{A}_{1i}}{\det \mathbf{A}} x_1' + \frac{\det \mathbf{A}_{2i}}{\det \mathbf{A}} x_2' + \ldots + \frac{\det \mathbf{A}_{ni}}{\det \mathbf{A}} x_n'$$

where $\mathbf{A}_{jk}$ is the matrix left after the jth row and kth column have been suppressed.

Consequently, the inverse of the matrix $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \begin{Vmatrix} \frac{\det \mathbf{A}_{11}}{\det \mathbf{A}} & \frac{\det \mathbf{A}_{21}}{\det \mathbf{A}} & \cdots & \frac{\det \mathbf{A}_{n1}}{\det \mathbf{A}} \\ \frac{\det \mathbf{A}_{12}}{\det \mathbf{A}} & \frac{\det \mathbf{A}_{22}}{\det \mathbf{A}} & \cdots & \frac{\det \mathbf{A}_{n2}}{\det \mathbf{A}} \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ \frac{\det \mathbf{A}_{1n}}{\det \mathbf{A}} & \frac{\det \mathbf{A}_{2n}}{\det \mathbf{A}} & \cdots & \frac{\det \mathbf{A}_{nn}}{\det \mathbf{A}} \end{Vmatrix} \tag{138.12}$$

On the basis of the computational laws given above, a definite meaning can be given to matrices raised to powers, polynomials whose terms contain matrices, power series of matrices, and equations involving matrices. As an example, we may be given the matrix equation

$$\mathbf{J}(\mathbf{U} - \mathbf{A}) = \mathbf{U}$$

to solve for the unknown matrix $\mathbf{J}$ in terms of the known matrices $\mathbf{A}$ and $\mathbf{U}$. One may obtain a formal solution of this equation by employing a method of successive approximations starting from the equation

$$\mathbf{J} = \mathbf{U} + \mathbf{J}\mathbf{A} \tag{138.13}$$

and calling the zeroth order approximation $\mathbf{J}_0 = \mathbf{U}$; then the first-order approximation may be found by letting $\mathbf{J} = \mathbf{J}_0 = \mathbf{U}$ on the right-hand side, so that

$$\mathbf{J}_1 = \mathbf{U} + \mathbf{U}\mathbf{A} = \mathbf{U} + \mathbf{A}$$

The second approximation $\mathbf{J}_2$ may be found by substituting $\mathbf{J}_1$ for $\mathbf{J}$ on the right side of (138.13). This process may be repeated indefinitely, so that

$$\mathbf{J} = \mathbf{U} + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \ldots \tag{138.14}$$

If this series converges, the expression (138.14) is the solution of eq. (138.13). It should be noted that the series (138.14) is the formal power series expansion of

$$\frac{\mathbf{U}}{\mathbf{U} - \mathbf{A}}$$

and would define this matrix. It is beyond the scope of this book to enter into a discussion of these more complicated manipulations with

matrices involving questions of convergence of series whose terms are matrices.

139. Operational point of view in linear transformations. When the laws of matrix calculus are defined in accordance with the previous section, all linear transformations may be indicated by matrix equations. The operational equation (138.1), which converts the vector x into a vector x', may be conveniently written as a matrix equation in which the vector x is represented by the matrix

$$\mathbf{x} = \begin{Vmatrix} 0 & 0 & \cdots & x_1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & x_2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & x_n & \cdots & 0 & 0 \end{Vmatrix} = \begin{Vmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ x_n \end{Vmatrix} \tag{139.1}$$

whose elements are zeros and the components of x. The multiplication law (138.6) gives

$$\mathbf{Ax} = \begin{Vmatrix} 0 & 0 & \cdots & \sum_{i=1}^{n} a_{1i}x_i & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \sum_{i=1}^{n} a_{2i}x_i & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sum_{i=1}^{n} a_{ni}x_i & \cdots & 0 & 0 \end{Vmatrix} = \begin{Vmatrix} x_1' \\ x_2' \\ x_3' \\ \cdot \\ x_n' \end{Vmatrix} = \mathbf{x}' \tag{139.2}$$

The representation of a vector by a matrix of one column containing the components of the vector as elements is advantageous in practical computations.

From an operational point of view, matrices having only one row of elements different from zero are of importance in representing linear and bilinear forms. The matrix with one row is really defined by the matrix

$$\begin{Vmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ x_1 & x_2 & \cdots & x_n \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{Vmatrix}$$

When it is not important to distinguish which row has elements different from zero, we may use the symbol $\tilde{\mathbf{x}}$ of the transposed matrix for the

single row matrix. Then

$$\mathbf{x} = \begin{Vmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{Vmatrix} \quad \text{and} \quad \tilde{\mathbf{x}} = \| \, x_1 x_2 \ldots x_n \, \| \tag{139.3}$$

Matrices represented by small letters will always have one column or one row. With the help of this notation, the ordinary scalar products of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ (see Sec. 134) can be expressed as a product of matrices.

$$(\mathbf{x} \cdot \mathbf{y}) = \tilde{\mathbf{x}}\mathbf{y} = \| \, x_1 x_2 \ldots x_n \, \| \cdot \begin{Vmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{Vmatrix}$$

$$= \sum_{i=1}^{n} x_i y_i = \tilde{\mathbf{y}}\mathbf{x} \tag{139.4}$$

Other more general forms can be conveniently expressed in terms of matrices.

The bilinear form is

$$\sum_{ik} y_i a_{ik} x_k = \tilde{\mathbf{y}}\mathbf{A}\mathbf{x} \tag{139.5}$$

and is built on the matrix $\mathbf{A}$, whose elements are a_{ik}. The scalar product (139.4) is a special case of the above bilinear form built on the unit matrix $\mathbf{U}$.

The quadratic form is

$$\sum_{ik} x_i a_{ik} x_k = \tilde{\mathbf{x}}\mathbf{A}\mathbf{x} \tag{139.6}$$

A quadratic form is said to be positive definite when it has no negative values. One may ascertain whether a quadratic form is positive definite or not by noting whether the matrix $\mathbf{A}$ upon which it is built can be expressed as the product of matrix $\mathbf{B}$ and its transposed matrix $\tilde{\mathbf{B}}$. If $\mathbf{A} = \mathbf{B}\tilde{\mathbf{B}} = \tilde{\mathbf{B}}\mathbf{B}$, the quadratic form (139.6) is positive definite. This follows since

$$\tilde{\mathbf{x}}\mathbf{B}\tilde{\mathbf{B}}\mathbf{x} = \sum_{ik} x_i \left(\sum_j b_{ji} b_{jk} \right) x_k = \sum_j \sum_{ik} x_i b_{ji} b_{jk} x_k$$

$$= \sum_j \left(\sum_i x_i b_{ji} \right) \left(\sum_k b_{jk} x_k \right) = \sum_j \left(\sum_k b_{jk} x_k \right)^2$$

The Hermitian form is a generalization of the scalar product of two vectors in a unitary space (see Sec. 136), and is defined to be

$$\sum_{ik} y_i^* h_{ik} x_k = \bar{\mathbf{y}}^* \mathbf{H} \mathbf{x} \tag{139.7}$$

and is built on the Hermitian matrix $\mathbf{H}$, whose elements h_{ik} are complex numbers obeying the relation $h_{ik} = h_{ki}^*$. This makes $\mathbf{H} = \mathbf{H}^*$.

The Hermitian quadratic form

$$\sum_{ik} x_i^* h_{ik} = \bar{\mathbf{x}}^* \mathbf{H} \mathbf{x} \tag{139.8}$$

has only real values.

Transformation in a New Coordinate System. Here we shall suppose that in a given system of basic vectors $\boldsymbol{u}_1, \ldots \boldsymbol{u}_n$ two vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ are related by the transformation equation

$$\mathbf{A}\mathbf{x} = \mathbf{x}' \tag{139.9}$$

and that the vectors $\boldsymbol{y}$ and $\boldsymbol{y}'$ are the same vectors as $\boldsymbol{x}$ and $\boldsymbol{x}'$ except that they are expressed in a new system of basic vectors $\boldsymbol{u}_1' \ldots \boldsymbol{u}_n'$. The operational point of view provides a simple way of determining the matrix which converts $\boldsymbol{y}$ into $\boldsymbol{y}'$. If $\mathbf{P}$ is the matrix which transforms the old system of vectors into the new system,

$$\mathbf{x} = \mathbf{P}\mathbf{y} \quad \text{and} \quad \mathbf{x}' = \mathbf{P}\mathbf{y}'$$

When these equations are substituted into eq. (139.9),

$$\mathbf{A}\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{y}'$$

or

$$(\mathbf{P}^{-1}\mathbf{A}\mathbf{P})\mathbf{y} = \mathbf{y}' \tag{139.10}$$

This means that the matrix which performs the same transformation in the new coordinates as the matrix $\mathbf{A}$ did in the old system is $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$.

140. Transformation of bilinear, quadratic, and Hermitian forms. According to eq. (139.5), a bilinear form $P(y,x)$ is really characterized by, or built on, a given matrix $\mathbf{P}$. When a transformation of coordinates is carried out, the same bilinear form is characterized by a new matrix $\mathbf{P}'$. By expressing the transformations in operational or matrix form, as indicated in Sec. 139, the resulting new matrix P' may easily be arrived at. As in eq. (139.5), let the bilinear form $P(\boldsymbol{y},\boldsymbol{x})$ be

$$P(y,x) = \sum_{i,k} y_i p_{ik} x_k = \bar{\mathbf{y}}\mathbf{P}\mathbf{x} \tag{140.1}$$

We may now transform the vector $\boldsymbol{x}$ with components $x_1, x_2, \ldots, x_n$ into a new vector $\boldsymbol{x}'$ with components $x_1', x_2', \ldots x_n'$ in accordance with eq. (139.1), so that

$$\mathbf{x} = \mathbf{A}\mathbf{x}' \tag{140.2}$$

Similarly, we may at the same time transform the vector $\boldsymbol{y}$ according to the matrix equation

$$\mathbf{y} = \mathbf{B}\mathbf{y}' \tag{140.3}$$

In expression (140.1), we require $\tilde{\mathbf{y}}$ instead of $\mathbf{y}$. This may be obtained from eq. (140.3) by noting that

$$\tilde{\mathbf{y}} = (\widetilde{\mathbf{B}\mathbf{y}'}) = \tilde{\mathbf{y}}'\tilde{\mathbf{B}} \tag{140.4}$$

When the expressions (140.2) and (140.4) are substituted in (140.1), we have the transformed bilinear form $P'(y',x')$ given by

$$P'(y',x') = \tilde{\mathbf{y}}'\tilde{\mathbf{B}}\mathbf{P}\mathbf{A}\mathbf{x}' = \tilde{\mathbf{y}}'(\tilde{\mathbf{B}}\mathbf{P}\mathbf{A})\mathbf{x}' = \tilde{\mathbf{y}}'\mathbf{P}'\mathbf{x}' \tag{140.5}$$

Consequently, the matrix for the new transformed bilinear form is

$$\mathbf{P}' = \tilde{\mathbf{B}}\mathbf{P}\mathbf{A} \tag{140.6}$$

Having found how the matrix of a bilinear form is altered by a linear transformation of the variables x_i and y_i, it is necessary only to apply the result [eq. (140.6)] to find how the quadratic form

$$\sum_{ik} x_i p_{ik} x_k = \tilde{\mathbf{x}}\mathbf{P}\mathbf{x}$$

is altered when the variables x_i are subjected to a linear transformation indicated by the matrix equation $\mathbf{x} = \mathbf{A}\mathbf{x}'$. Using eq. (139.6), the transformed quadratic form is

$$\tilde{\mathbf{x}}'\tilde{\mathbf{A}}\mathbf{P}\mathbf{A}\mathbf{x}' \tag{140.7}$$

In the special case that $\mathbf{P} = \mathbf{U}$ the quadratic form is just the square of the length of the vector $\boldsymbol{x}$, since

$$(\boldsymbol{x}\cdot\boldsymbol{x}) = \sum_i x_i x_i = \tilde{\mathbf{x}}\mathbf{U}\mathbf{x} \tag{140.8}$$

If the vector $\boldsymbol{x}$ be converted into a new vector by the linear transformation whose matrix is $\mathbf{A}$, then (140.8) becomes

$$\tilde{\mathbf{x}}'\tilde{\mathbf{A}}\mathbf{U}\mathbf{A}\mathbf{x}'$$

A special kind of transformation of coordinates is defined when the linear transformation is such that the matrix of the transformed quadratic form is again the unit matrix; that is to say

$$\tilde{\mathbf{A}}\mathbf{U}\mathbf{A} = \mathbf{U}$$

or

$$\tilde{\mathbf{A}} = \mathbf{A}^{-1} \tag{140.9}$$

When the quadratic form built on the unit matrix remains invariant under a linear transformation, the transformation is said to be an *orthogonal transformation*, and the matrix $\mathbf{A}$ defining the transformation and conforming to the conditions (140.9) is called an *orthogonal matrix*. The determinant A of an orthogonal matrix has some interesting properties. On writing the determinant equation corresponding to (140.9),

$$\tilde{A}A = 1$$

and since

$$\tilde{A} = A$$

$$A^2 = 1$$

or

$$A = \det \mathbf{A} = \pm 1$$

From the expression (140.7) and the above result, it follows that when a quadratic form is subjected to an orthogonal transformation, the determinant of the matrix of the quadratic forms remains invariant. Thus if the quadratic form built on the matrix $\mathbf{P}$ is transformed orthogonally by means of the orthogonal matrix $\mathbf{A}$, the new matrix of the quadratic form, according to (140.7), is $\tilde{\mathbf{A}}\mathbf{P}\mathbf{A}$. But taking determinants

$$\det \tilde{\mathbf{A}}\mathbf{P}\mathbf{A} = \tilde{A}PA = P\tilde{A}A = \det \mathbf{P}$$

From the above discussion, it is apparent that whenever the coordinate system is subjected to an orthogonal transformation, the square of the length of a vector in the new coordinates is still given by the sum of the squares of the new components.

The transformation of the Hermitian form (139.7) can be carried out in a similar manner, as that for the bilinear form (139.5). If the variables y_i and x_i in eqs. (138.7) are transformed in accordance with the matrix equations

$$\mathbf{x} = \mathbf{A}\mathbf{x}' \quad \text{and} \quad \mathbf{y} = \mathbf{B}\mathbf{y}'$$

we have

$$\bar{\mathbf{y}}^*\mathbf{H}\mathbf{x} = \widetilde{(\mathbf{B}\mathbf{y}')}^*\mathbf{H}\mathbf{A}\mathbf{x}' = \bar{\mathbf{y}}'^*\tilde{\mathbf{B}}^*\mathbf{H}\mathbf{A}\mathbf{x}'$$

so that the new Hermitian form is built on the matrix

$$\tilde{\mathbf{B}}^*\mathbf{H}\mathbf{A} \tag{140.10}$$

The special transformation which leaves the Hermitian form

$$\sum_i x_i^* x_i = \bar{\mathbf{x}}^*\mathbf{U}\mathbf{x}$$

invariant is called a *unitary transformation.* The conditions which must hold for the matrix of the transformation can readily be determined in the same way as for the orthogonal transformation discussed above. Thus if

$$\bar{\mathbf{x}} = \mathbf{A}\mathbf{x}'$$

then

$$\bar{\mathbf{x}}^*\mathbf{U}\mathbf{x} = \bar{\mathbf{x}}'^*\tilde{\mathbf{A}}^*\mathbf{U}\mathbf{A}\mathbf{x}'$$

and the invariance requires that

$$\tilde{\mathbf{A}}^*\mathbf{A} = \mathbf{U} \tag{140.11}$$

A matrix satisfying the condition (140.11) is called a *unitary matrix.*

If the determinant equation is formed from eq. (140.11),

$$\det \tilde{\mathbf{A}}^*\mathbf{A} = 1 \quad \text{or} \quad |\det \mathbf{A}| = 1$$

The determinant of the matrix $\mathbf{H}$ of the Hermitian form $\mathfrak{x}^*\mathbf{H}\mathbf{x}$ remains invariant under a unitary transformation. The line of argument here is the same as that leading to the invariance of the matrix of a quadratic form under the orthogonal transformations discussed above.

141. Characteristic vectors and the transformation to principal axes. In the theory of vibrations, the kinetic and potential energies of the vibrating system are expressed as quadratic forms in the velocities and coordinates. These quadratic forms according to eq. (139.6) may be written

$$\sum x_i q_{ik} x_k = \mathfrak{x}\mathbf{Q}\mathbf{x} \tag{141.1}$$

It is characteristic of these particular quadratic forms that $q_{ik} = q_{ki}$, which in turn makes the matrix $\mathbf{Q}$ symmetric in the sense that $\tilde{\mathbf{Q}} = \mathbf{Q}$. The determination of the equations of motion is enormously simplified if the above quadratic form can be reduced to a sum of squares, that is

$$\mathfrak{x}\mathbf{Q}\mathbf{x} = \mathfrak{x}'\mathbf{D}\mathbf{x}' = \sum_{i=1}^{n} d_i x_i'^2 \tag{141.2}$$

thereby requiring the matrix $\mathbf{D}$ to be the diagonal matrix

$$\begin{Vmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & d_n \end{Vmatrix}$$

This reduction can always be accomplished when the matrix $\mathbf{Q}$ of the original quadratic form is symmetric by a suitable linear transformation $\mathbf{x} = \mathbf{A}\mathbf{x}'$. The coordinates $x_1', x_2', \dots x_n'$ are called the normal coordinates.

The problem before us, therefore, is to determine the matrix $\mathbf{A}$ and the matrix $\mathbf{D}$. This may be done in a number of different ways, several of which will be explained below. In the first place, the elements of the matrix $\mathbf{A}$ may be determined algebraically in a straightforward manner. We shall attempt to determine an orthogonal transformation (see Sec. 140), $\mathbf{x} = \mathbf{A}\mathbf{x}'$, which will satisfy eq. (141.2). From expression (140.7), it is known that the quadratic form $\mathfrak{x}\mathbf{Q}\mathbf{x}$ transforms into the quadratic form $\mathfrak{x}'\tilde{\mathbf{A}}\mathbf{Q}\mathbf{A}\mathbf{x}'$, so that according to eq. (141.2), $\mathbf{A}$ must satisfy the matrix equation

$$\tilde{\mathbf{A}}\mathbf{Q}\mathbf{A} = \mathbf{D} \tag{141.3}$$

That $\mathbf{Q}$ must be a symmetric matrix ($\mathbf{Q} = \tilde{\mathbf{Q}}$) is obvious, since $\mathbf{D} = \tilde{\mathbf{D}}$ and

$$\tilde{\mathbf{D}} = \widetilde{\tilde{\mathbf{A}}\mathbf{Q}\mathbf{A}} = \tilde{\mathbf{A}}\tilde{\mathbf{Q}}\mathbf{A} = \mathbf{D} = \tilde{\mathbf{A}}\mathbf{Q}\mathbf{A}$$

Furthermore, since $\mathbf{A}$ is to be an orthogonal matrix, $\tilde{\mathbf{A}} = \mathbf{A}^{-1}$, and eq. (141.3) may be written

$$\mathbf{QA} = \mathbf{AD} \tag{141.4}$$

When the elements of the above product matrices are equated, we have the system of homogeneous linear equations

$$\sum_{k=1}^{n} q_{ik}a_{kj} = a_{ij}d_j \qquad \begin{matrix} i = 1, 2, \ldots, n \\ j = 1, 2, \ldots, n \end{matrix} \tag{141.5}$$

which, with the requirement that $\det \mathbf{A} = 1$ (see Sec. 140) for an orthogonal transformation, serve to determine the n^2 elements a_{ij} of the transformation matrix $\mathbf{A}$ when the elements d_j are known. The above system of equations will have a solution other than $a_{ij} = 0$ for all i and j only if the determinant of the coefficients vanishes (see Sec. 133). The determinant of the above system of equations has the following special form.

(141.6)

$$\begin{vmatrix}
(q_{11}-d_1) & q_{12} & \ldots & q_{1n} & 0 & 0 & \ldots & 0 & . & . & . \\
q_{21} & (q_{22}-d_1) & \ldots & q_{2n} & 0 & 0 & \ldots & 0 & . & . & . \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
q_{n1} & q_{n2} & \ldots & (q_{nn}-d_1) & 0 & 0 & \ldots & 0 & . & . & . \\
0 & 0 & \ldots & 0 & (q_{11}-d_2) & q_{12} & \ldots & q_{1n} & 0 & . & . \\
0 & 0 & \ldots & 0 & q_{21} & (q_{22}-d_2) & \ldots & q_{2n} & 0 & . & . \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & 0 & \ldots & 0 & q_{n1} & q_{n2} & \ldots & q_{nn} & 0 & . & . \\
0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & (q_{11}-d_3) & . & . \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & q_{21} & . & . \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots
\end{vmatrix}$$

In accordance with the rules for evaluating determinants (see Sec. 131), the value of the above determinant is equal to the product of the n determinants of order n which lie along the principal diagonal of the determinant (141.6). Since it is required that this determinant vanish, it is necessary that at least one of these factors vanish, so that

$$\begin{vmatrix}
(q_{11} - \lambda) & q_{12} & \ldots & q_{1n} \\
q_{21} & (q_{22} - \lambda) & \ldots & q_{2n} \\
\ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots \\
q_{n1} & q_{n2} & \ldots & (q_{nn} - \lambda)
\end{vmatrix} = 0 \tag{141.7}$$

where λ denotes any of the as yet unknown quantities d_j. Equation (141.7) is usually referred to as the secular equation. The determinant is a polynomial of degree n in λ so that eq. (141.7) may be written

$$f(\lambda) = -(\lambda^n - s_1\lambda^{n-1} + \ldots \pm s_n) = 0 \tag{141.8}$$

where $s_1, s_2, \ldots, s_n$ are rational functions of the elements q_{ij}. In particular, the coefficient s_1 of λ^{n-1} is

$$s_1 = q_{11} + q_{22} + \ldots + q_{nn} \tag{141.9}$$

and is called the trace of the determinant. It has the property of remaining invariant when the system of basic vectors is subjected to a linear transformation.

The polynomial (141.8) will have n real roots $\lambda = d_1$, $\lambda = d_2$, $\ldots$, $\lambda = d_n$ which satisfy the secular equation (141.7) and furnish the elements of the diagonal matrix D in eq. (141.2).

The normal coordinate x' may be expressed in terms of the old coordinate x by means of the matrix equation

$$\mathbf{x}' = \mathbf{A}^{-1}\mathbf{x} = \tilde{\mathbf{A}}\mathbf{x} \tag{141.10}$$

This completes the algebraic reduction of the quadratic form $\tilde{\mathbf{x}}\mathbf{Q}\mathbf{x}$ to the sum of squares $\sum_{i=1}^{n} d_i x_i^2$.

The reduction of the Hermitian quadratic form $\sum_{ik} x_i^* h_{ik} x_k = \tilde{\mathbf{x}}^*\mathbf{H}\mathbf{x}$ can be reduced to the form $\sum_k d_k |\, x_k \,|^2$ in a similar way when $\tilde{H}^* = H$, by a transformation whose matrix is unitary.

The elements of the transformation matrix $\mathbf{A}$ obtained above can be used to construct a system of vectors which have special properties with respect to the matrix $\mathbf{Q}$ of the quadratic form. Consider the vector $\boldsymbol{x}_1'$ which, when expressed in terms of a system of unit vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ has the components $a_{11}, a_{21}, \ldots, a_{n1}$, so that

$$\boldsymbol{x}_1' = \sum_j a_{j1}\boldsymbol{u}_j \tag{141.11}$$

If we operate on this vector with the matrix $\mathbf{Q}$ given above in accordance with the matrix eq. (139.1), we have

$$\mathbf{Q}\mathbf{x}_1' = \mathbf{Q}\begin{Vmatrix} a_{11} \\ a_{21} \\ \cdot \\ \cdot \\ \cdot \\ a_{n1} \end{Vmatrix} = \begin{Vmatrix} \sum_j q_{1j}a_{j1} \\ \sum_j q_{2j}a_{j1} \\ \cdots\cdots \\ \cdots\cdots \\ \cdots\cdots \\ \sum_j q_{nj}a_{j1} \end{Vmatrix} = \begin{Vmatrix} d_1 a_{11} \\ d_1 a_{21} \\ \cdots \\ \cdots \\ \cdots \\ d_1 a_{n1} \end{Vmatrix} = d_1\mathbf{x}_1' \tag{141.12}$$

The relation (141.5) was used to obtain the matrix on the right. It is thus evident that the vector $\boldsymbol{x}_1'$ given by the relation (141.11) is transformed into a multiple of itself by means of the matrix $\mathbf{Q}$, which means

that the length and not the direction of $\mathbf{x}_1'$ is altered by the transformation. The vector $\mathbf{x}_1'$ is called a principal axis vector, or a characteristic vector of the matrix $\mathbf{Q}$, and d_1 is called a characteristic value of $\mathbf{Q}$ belonging to $\mathbf{x}_1'$.

Following the line of reasoning given above, n characteristic vectors $\mathbf{x}_i'$ $(i = 1, 2, \ldots, n)$ can be found when there are n distinct characteristic values d_i of the matrix $\mathbf{Q}$. These vectors are

$$\mathbf{x}_i' = \sum_j a_{ji}\mathbf{u}_j \qquad i = 1, 2, \ldots, n \tag{141.13}$$

The characteristic vectors $\mathbf{x}_i'$ are mutually orthogonal. This may readily be shown for any two vectors $\mathbf{x}_i'$ and $\mathbf{x}_j'$, which by hypothesis satisfy the equations

$$\mathbf{Q}\mathbf{x}_i' = d_i\mathbf{x}_i'$$
$$\mathbf{Q}\mathbf{x}_j' = d_j\mathbf{x}_j'$$

Multiply the first equation by $\tilde{\mathbf{x}}_j'$, and the second equation transposed by $\mathbf{x}_i'$, so that

$$\tilde{\mathbf{x}}_j'\mathbf{Q}\mathbf{x}_i' = d_i\tilde{\mathbf{x}}_j'\mathbf{x}_i'$$
$$\widetilde{\mathbf{Q}\mathbf{x}_j'}\mathbf{x}_i' = d_j\tilde{\mathbf{x}}_j'\mathbf{x}_i'$$

and subtract, thereby obtaining

$$\begin{aligned}(d_i - d_j)\tilde{\mathbf{x}}_j'\mathbf{x}_i' &= \tilde{\mathbf{x}}_j'\mathbf{Q}\mathbf{x}_i' - \widetilde{\mathbf{Q}\mathbf{x}_j'}\mathbf{x}_i' \\ &= \tilde{\mathbf{x}}_j'\mathbf{Q}\mathbf{x}_i' - \tilde{\mathbf{x}}_j'\tilde{\mathbf{Q}}\mathbf{x}_i' = 0\end{aligned}$$

The right side vanishes because of the symmetry of the matrix Q. Since, according to eq. (139.4), $\tilde{\mathbf{x}}_j' \cdot \mathbf{x}_j'$ is just the scalar product of the vectors $\mathbf{x}_j'$ and $\mathbf{x}_i'$, it follows that if $d_i \neq d_j$, the characteristic vectors in question are orthogonal.

As a result of the above discussion, it is apparent that one can proceed to find the principal axes transformation by determining the characteristic vectors of the matrix $\mathbf{Q}$. Looking at the problem from this point of view, we seek a solution of the equations

$$\mathbf{Q}\mathbf{x}_k' = d_k\mathbf{x}_k' \qquad k = 1, 2, \ldots, n \tag{141.14}$$

thereby determining n characteristic vectors,

$$\mathbf{x}_k' = \sum_j a_{jk}\mathbf{u}_j \qquad k = 1, 2, \ldots, n \tag{141.15}$$

and n characteristic values $d_1, d_2, \ldots, d_n$.

Considering any one of eqs. (141.14), it may be written in the form

$$(\mathbf{Q} - \lambda\mathbf{U})\mathbf{x}_i' = 0 \cdot$$

where U is a unit matrix. This requires that all elements of the matrix on the left vanish; hence

$$\sum_j (q_{ij} - \lambda\delta_{ij})a_{jk} = 0 \qquad i = 1, 2, \ldots, n \tag{141.16}$$

where $$\delta_{ij} = \begin{matrix} 0 & i \neq j \\ 1 & i = j \end{matrix}$$

The above equations must also hold for all $k = 1, 2, \ldots, n$. The n^2 eqs. (141.16) for the determination of the components a_{jk} of the vectors $\mathbf{x}_k'$ are identical with the equations (141.5) already determined. The characteristic values $\lambda = d_1$, $\lambda = d_2$, $\ldots$, $\lambda = d_n$ are found as before to be the roots of the secular eq. (141.7), which also may be written

$$\det(\mathbf{Q} - \lambda\mathbf{U}) = 0 \tag{141.17}$$

Still another method of approach to the problem of determining the principal axes has as its basis a certain maximal principle. This principle asserts that, given a quadratic form $\mathfrak{x}\mathbf{Q}\mathbf{x}$ built on the matrix $\mathbf{Q}$, there exists a unit vector $\mathbf{x}_1'$ with components $a_{11}, a_{21}, \ldots, a_{n1}$, such that the quadratic form $\mathfrak{x}_1'\mathbf{Q}\mathbf{x}'$ assumes its maximum value d_1. This means that the components of the characteristic vector $\mathbf{x}_1'$ are found by choosing those for which the quadratic form $\mathfrak{x}_1'\mathbf{Q}\mathbf{x}_1'$ is a maximum, subject to the auxiliary condition

$$\sum_{j=1}^{n} (a_{1j})^2 = 1 \tag{141.18}$$

The maximum value of $\mathfrak{x}_1'\mathbf{Q}\mathbf{x}_1$ is just the value of the characteristic value d_1 belonging to the principal axes vector $\mathbf{x}_1'$. After having found $\mathbf{x}_1'$, another characteristic vector $\mathbf{x}_2'$ orthogonal to $\mathbf{x}_1'$ can be found by determining its components $a_{21}, a_{22}, \ldots, a_{2n}$, so that the quadratic form $\mathfrak{x}_2'\mathbf{Q}\mathbf{x}_2'$ is a maximum subject to the auxiliary conditions

$$\sum_{j=1}^{n} (a_{2j})^2 = 1 \tag{141.19}$$

$$\sum_{j=1}^{n} a_{1j}a_{2j} = 0 \tag{141.20}$$

The maximum value of $\mathfrak{x}_2'\mathbf{Q}\mathbf{x}_2' = d_2$. The number of auxiliary conditions increases for the determination of successive characteristic vectors because each of these vectors is required to be orthogonal to each of the preceding characteristic vectors. For further details concerning this method, the reader is referred to *Methoden der mathematischen Physik*, Vol. 1, by R. Courant and D. Hilbert.

As an example of the reduction to principal axes or the determination of the characteristic vectors of a matrix, let it be required to transform the quadratic form

$$2x_1^2 + 2x_2^2 + 2x_3^3 - 2x_1x_2 - 2x_2x_3 \tag{141.21}$$

to principal axes.

The matrix upon which this quadratic form is built is obviously

$$\mathbf{Q} = \begin{Vmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{Vmatrix} \tag{141.22}$$

The characteristic values are found by solving the secular eq. (141.7), which in this case is

$$\begin{vmatrix} 2-\lambda & -1 & 0 \\ -1 & 2-\lambda & -1 \\ 0 & -1 & 2-\lambda \end{vmatrix} = 0$$

or

$$(2-\lambda)(\lambda^2 - 4\lambda + 2) = 0$$

The three roots of this equation are $\lambda = d_1, \lambda = d_2, \lambda = d_3$, where $d_1 = 2$, $d_2 = 2 - \sqrt{2}, d_3 = 2 + \sqrt{2}$. The principal axis vector $\mathbf{x}_1' = \sum_{j=1}^{3} a_{j1}\mathbf{u}_j$ belonging to d_1 is such that the components a_{j1} must satisfy the matrix equation

$$\mathbf{Q}\mathbf{x}_1' = d_1\mathbf{x}_1' \tag{141.23}$$

or

$$\begin{Vmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{Vmatrix} \cdot \begin{Vmatrix} a_{11} \\ a_{21} \\ a_{31} \end{Vmatrix} = 2 \begin{Vmatrix} a_{11} \\ a_{21} \\ a_{31} \end{Vmatrix}$$

or

$$\begin{Vmatrix} (2a_{11} - a_{21} + 0) \\ (-1a_{11} + 2a_{21} - a_{31}) \\ (0 - a_{21} + 2a_{31}) \end{Vmatrix} = 2 \begin{Vmatrix} a_{11} \\ a_{21} \\ a_{31} \end{Vmatrix}$$

Equating corresponding matrix elements we have

$$\begin{aligned} 2a_{11} - a_{21} &= 2a_{11} \\ -a_{11} + 2a_{21} - a_{31} &= 2a_{21} \\ - a_{21} + 2a_{31} &= 2a_{31} \end{aligned}$$

This gives

$$\begin{aligned} a_{21} &= 0 \\ a_{31} &= -a_{11} \end{aligned}$$

Since eq. (140.23) is linear, we may make the length of the vector $\mathbf{x}_1'$ unity, thus requiring that

$$a_{11}^2 + a_{31}^2 = 1$$

thereby uniquely determining the components a_{11}, a_{21}, a_{31} to be

$$a_{11} = \frac{1}{\sqrt{2}} \qquad a_{21} = 0 \qquad a_{31} = -\frac{1}{\sqrt{2}}$$

When the same procedure is carried out for the equations

$$\mathbf{Q}\mathbf{x}_2' = d_2\mathbf{x}_2' \qquad \mathbf{Q}\mathbf{x}_3' = d_3\mathbf{x}_3'$$

it is found that

$$a_{12} = \frac{1}{2} \qquad a_{13} = \frac{1}{2}$$

$$a_{22} = \frac{1}{\sqrt{2}} \qquad a_{23} = -\frac{1}{\sqrt{2}}$$

$$a_{32} = \frac{1}{2} \qquad a_{33} = \frac{1}{2}$$

The matrix $\mathbf{A}$ of the transformation $\mathbf{x} = \mathbf{A}\mathbf{x}'$ is

$$\mathbf{A} = \begin{Vmatrix} \frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \end{Vmatrix} \tag{141.24}$$

It may be noted that $\det \mathbf{A} = 1$, a property possessed by the determinant of an orthogonal transformation, as shown in Sec. 140.

In accordance with eq. (140.7), the matrix of the transformed quadratic form is

$$\tilde{\mathbf{A}}\mathbf{Q}\mathbf{A} = \begin{Vmatrix} 2 & 0 & 0 \\ 0 & 2-\sqrt{2} & 0 \\ 0 & 0 & 2+\sqrt{2} \end{Vmatrix}$$

and the quadratic form (141.21) has been transformed to principal axes, so that

$$\begin{aligned} &2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3 \\ &= 2x_1'^2 + (2-\sqrt{2})x_2'^2 + (2+\sqrt{2})x_3'^2 \end{aligned}$$

142. Simultaneous reduction of a pair of quadratic forms. In mechanics, it sometimes happens that both the kinetic energy and potential energy of the system can be expressed as quadratic forms, and it is desirable to be able to reduce these quadratic forms to sums of squares of the coordinates. In practice this is usually possible. Given two quadratic forms $\mathfrak{x}\mathbf{P}\mathbf{x}$ and $\mathfrak{x}\mathbf{Q}\mathbf{x}$, we seek an orthogonal transformation $\mathbf{x} = \mathbf{A}\mathbf{x}'$ with matrix $\mathbf{A}$ such that the matrices of the transformed quadratic forms are simultaneously diagonal. That is,

$$\tilde{\mathbf{A}}\mathbf{P}\mathbf{A} = \mathbf{D} = \begin{Vmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ 0 & 0 & & . \\ 0 & 0 & \dots & d_n \end{Vmatrix} \tag{142.1}$$

and

$$\tilde{\mathbf{A}}\mathbf{Q}\mathbf{A} = \boldsymbol{E} = \begin{Vmatrix} e_1 & 0 & \dots & 0 \\ 0 & e_2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & e_n \end{Vmatrix} \tag{142.2}$$

The necessary and sufficient condition for the existence of matrix **A** which will satisfy eqs. (142.1) and (142.2) simultaneously is that $\mathbf{PQ} = \mathbf{QP}$; that is, the matrices **P** and **Q** must commute. More generally, any number of quadratic forms can be simultaneously reduced if and only if their matrices commute with one another.

A convenient method of determining the matrix **A** is to determine the vectors $\boldsymbol{x}_k'$ which will be the characteristic vectors of both matrices **P** and **Q** simultaneously. According to the discussion in the preceding section, which applied to the characteristic vector problem for one matrix, we must determine a set of vectors $\boldsymbol{x}_k'$ such that

$$\mathbf{P}\mathbf{x}_k' = d_k\mathbf{x}_k' \tag{142.3}$$

and

$$\mathbf{Q}\mathbf{x}_k' = e_k\mathbf{x}_k'$$

If the second equation is multiplied by d_k/e_k and subtracted from the first,

$$\left(\mathbf{P} - \frac{d_k}{e_k}\mathbf{Q}\right)\mathbf{x}_k' = 0$$

By setting $d_k/e_k = \lambda_k$, this equation becomes

$$(\mathbf{P} - \lambda_k\mathbf{Q})\mathbf{x}_k' = 0 \tag{142.4}$$

When the vector $\boldsymbol{x}_k'$ is expressed in terms of its components

$$\boldsymbol{x}_k' = \sum_j a_{jk}\mathbf{u}_j$$

As in the preceding section, the matrix eq. (142.4) may be written

$$\begin{Vmatrix} p_{11} - \lambda_k q_{11} & p_{12} - \lambda_k q_{12} & \dots & p_{1n} - \lambda_k q_{1n} \\ p_{21} - \lambda_k q_{21} & p_{21} - \lambda_k q_{22} & \dots & p_{2n} - \lambda_k q_{2n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ p_{n1} - \lambda_k q_{n1} & p_{n2} - \lambda_k q_{n2} & \dots & p_{nn} - \lambda_k q_{nn} \end{Vmatrix} \cdot \begin{Vmatrix} a_{1k} \\ a_{2k} \\ \cdot \\ \cdot \\ \cdot \\ a_{nk} \end{Vmatrix} = \begin{Vmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{Vmatrix} \tag{142.5}$$

Corresponding elements of the product matrix on the left and the matrix on the right may be equated, yielding n linear homogeneous equations for the components of the vector $\boldsymbol{x}_k'$

$$\sum_j (p_{ij} - \lambda_k q_{ij})a_{jk} = 0 \qquad i = 1, 2, \dots, n \tag{142.6}$$

These equations possess a nontrivial solution when

$$\det (\mathbf{P} - \lambda \mathbf{Q}) = 0$$

This secular equation serves to determine the characteristic values $\lambda_1, \lambda_2, \ldots, \lambda_n$. Since λ_k gives the ratio d_k/e_k there is some arbitrariness in the choice of d_k and e_k. It is customary to take $e_k = 1$. Then, in terms of the new variables x'_k, the two quadratic forms become

$$\mathfrak{x}\mathbf{P}\mathfrak{x} = \mathfrak{x}'\mathbf{\Lambda}\mathfrak{x}' = \sum_k \lambda_k (x'_k)^2$$

$$\mathfrak{x}\mathbf{Q}\mathfrak{x} = \mathfrak{x}'\mathbf{U}\mathfrak{x}' = \sum_k (x'_k)^2$$

This completes the simultaneous reduction of two quadratic forms to principal axes.

143. Operations in a space of a countably infinite number of dimensions. In certain fields of physics such as the quantum theory it is necessary to deal with a space having an infinite number of dimensions. Such a space is made up of the aggregate of all vectors

$$\boldsymbol{x} = (x_1, x_2, x_3, \ldots)$$

having a countably infinite sequency of components x_i for which the series

$$(\boldsymbol{x} \cdot \boldsymbol{x}) = x_1^1 + x_2^2 + \ldots \tag{143.1}$$

converges. Under these conditions all the operations and theory given in the foregoing sections for an orthogonal space of finite dimensions still hold for the infinite-dimensional space, with the exception of certain completeness theorems.

For an infinite-dimensional unitary space, the components x_i are complex, and the requirement that the series (143.1) converge must be replaced by the requirement that the series

$$(\boldsymbol{x}^* \cdot \boldsymbol{x}) = x_1^* x_1 + x_2^* x_2 + \ldots \tag{143.2}$$

must converge.

SUGGESTED REFERENCES

M. Bôcher, *Introduction to Higher Algebra*, New York, The Macmillan Company, 1924.

R. Courant and D. Hilbert, *Methoden der mathematischen Physik* I., Berlin, Julius Springer, 1931.

M. H. Stone, *Linear Transformations in Hilbert Space*, New York, American Mathematical Society, 1932.

Handbuch der Physik, Vol. 3, Berlin, Julius Springer, 1928.

R. A. Frazer, W. J. Duncan, and A. R. Collar, *Elementary Matrices*, London, Cambridge University Press, 1938.

PROBLEMS

130.1. If the elements of a determinant D are functions of a parameter t, find dD/dt.

130.2. Find the product of a Jacobian and its inverse by direct multiplication. *Hint.* Use rule 1.

131.1. Evaluate the determinant

$$\begin{vmatrix} 5 & 3 & 0 & 1 \\ 4 & 7 & 7 & 2 \\ 3 & 1 & 4 & 2 \\ 1 & 3 & 5 & 0 \end{vmatrix}$$

(a) by minors of its first row; (b) by minors of its first two rows.

Test the following sets of equations for consistency and give a complete solution in each case.

132.1. $4x - y + z = -5$, $x + y - 2z = -2$,
$2x - 3y + 5z = -1$, $5x - z = -2$.

132.2. $2x - y + 3z = 1$, $2x - y - 4z = -4$,
$4x - 2y - z = -3$, $10x - 5y - 6z = -10$.

132.3. $5x - 3y + z = 2$,
$x + 4y = -1$,
$7x - 18y + 2z = 8$.

135.1. Construct an orthogonal set of vectors from the set

$x_1 = (-4, -3, 2, 1)$ $x_3 = (-1, 2, -1, 0)$
$x_2 = (5, -2, 5, -4)$ $x_4 = (3, 2, -1, -1)$

Are these vectors independent?

136.1. If r and s are two vectors in a unitary space, show that $(r \cdot s) = (s \cdot r)^*$.

138.1. Find the inverse of

$$\mathbf{A} = \begin{Vmatrix} 1 & 3 & 0 \\ -1 & 2 & 1 \\ 0 & 0 & 2 \end{Vmatrix}$$

Verify that $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A}$.

138.2. Find the inverse of

$$\begin{Vmatrix} 5 & 0 & 0 & 0 \\ 0 & 2 & 6 & -4 \\ 0 & 0 & -3 & 1 \\ 0 & 4 & -1 & 2 \end{Vmatrix}$$

139.1. Determine the matrix which represents a rotation through the angle θ about the z axis, in three dimensional space.

139.2. Determine the matrix which represents a rotation through an angle θ about the axis whose direction cosines are α, β, γ.

Determine the characteristic values and characteristic vectors for the following matrices.

141.1. $\begin{Vmatrix} \frac{7}{4} & -\frac{1}{4} & \frac{1}{2\sqrt{2}} \\ -\frac{3}{4} & \frac{3}{4} & -\frac{1}{2\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{7}{2} \end{Vmatrix}$

141.2. $\begin{Vmatrix} 3 & 2 & 1 \\ 2 & 2 & 0 \\ 1 & 0 & 1 \end{Vmatrix}$

141.3. If $\mathbf{A}$ is a Hermitian matrix, show that as n increases $\mathbf{A}^n u$ approaches the characteristic vector belonging to the largest (magnitude) characteristic value of $\mathbf{A}$, where u is almost any vector. *Hint.* Write u as a linear combination of the characteristic vectors of $\mathbf{A}$.

141.4. Let $u = x_i + \sum_j \epsilon_j x_j$, where x_i are the characteristic vectors of a Hermitian matrix $\mathbf{A}$, and ϵ_i are infinitesimals. Show that

$$\frac{\bar{u}^* \mathbf{A} u}{\bar{u}^* u} = d_i + \text{terms in } \epsilon_i^2$$

where d_i is the characteristic value belonging to x_i.

Use the two problems above to find approximate values for the largest characteristic values and vectors for the following matrices. Find characteristic values to 3 decimal places.

141.5. $\begin{Vmatrix} 5 & -2 & 1 \\ -2 & 3 & -4 \\ 1 & -4 & 0 \end{Vmatrix}$

141.6. $\begin{Vmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{Vmatrix}$

141.7. Find one-decimal values for all characteristic values of Problem 141.6.

141.8. Let $\mathbf{A}$ be an Hermitian matrix of n rows and columns, x be the vector of unit length

$$\begin{Vmatrix} 0 \\ 0 \\ x_3 \\ 1 \\ x_n \end{Vmatrix}, \quad u \text{ be the vector } \begin{Vmatrix} a \\ b \\ cx_3 \\ . \\ . \\ . \\ cx_n \end{Vmatrix}$$

Show that those values of a, b, and c which make $u^*\mathbf{A}u/u^*u$ a maximum are those values which make $\begin{Vmatrix} a \\ b \\ c \end{Vmatrix}$ a solution of the characteristic value problem for a matrix $\mathbf{B}$ with three rows and columns.

141.9. Use Problem 141.8 to find two-decimal characteristic values for the matrix

of Problem 141.6 by reducing that matrix to a 2 by 2 matrix. Check that the sum of the characteristic values is equal to the trace of the matrix.

141.10. Find three approximate characteristic values for the matrix

$$\begin{Vmatrix} 1 & 5 & -6 & 1 & 2 \\ 5 & 9 & 3 & 1 & 0 \\ -6 & 3 & 1 & -4 & 2 \\ 1 & 1 & -4 & -3 & 0 \\ 2 & 0 & 2 & 0 & 7 \end{Vmatrix}$$

Note that the method of Problem 141.3 may be used to find two characteristic values of a matrix $\mathbf{A}$ if one uses the fact that the characteristic values of the matrix $\mathbf{A} - \lambda\mathbf{U}$ are λ less than those of $\mathbf{A}$, where $\mathbf{U}$ is the unit matrix.

142.1. Reduce to normal coordinates the pair of forms

$$2T = m(x_1^2 + x_2^2 + x_3^2)$$
$$2V = \mu(x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_2x_3)$$

142.2. Reduce to normal coordinates the pair of forms

$$2T = mx_1^2 + 2mx_2^2 + mx_3^2$$
$$2V = 3x_1^2 + 4x_2^2 + 3x_3^2 + 2x_1x_3$$

Chapter 11

VECTOR AND TENSOR ANALYSIS

144. Vectors in three-dimensional space. In Sec. 134, a vector v in a space of n dimensions was defined to be the set of ordered numbers $v_1, v_2, \ldots, v_n$, which we called the components of the vector v. In three dimensions, the ordering of the numbers v_1, v_2, v_3, and their associations with the vector v are usually accomplished as follows. A right-handed Cartesian system of three mutually perpendicular axes meeting at a point is selected such that distances measured along each axis are specified by x_1, x_2, and x_3, respectively, as shown in Fig. 144.1. In such a system of coordinates a point P is completely specified by the ordered number triple (x_1, x_2, x_3). The same point P will be specified in another Cartesian system with identical origin 0 but rotated with respect to the first system by a different number triple (x_1', x_2', x_3'). From analytical geometry, the numbers x_1', x_2', x_3' are expressed in terms of the numbers x_1, x_2, x_3, by the linear equations

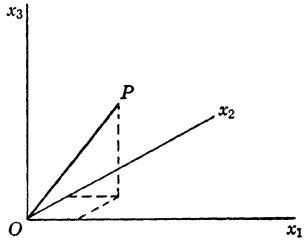


FIG. 144.1

$$\left.\begin{aligned} x_1' &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ x_2' &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ x_3' &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{aligned}\right\} \quad (144.1)$$

where a_{11}, a_{12}, a_{13} are the direction cosines of the new x_1' axis with respect to the x_1, x_2, x_3 axes, while a_{21}, a_{22}, a_{23} are the direction cosines of the x_2' axis with respect to the x_1, x_2, x_3 axes, and so on. As we have seen in Sec. 139, eqs. (144.1) can be conveniently expressed by the single matrix equation

$$\mathbf{x}' = \mathbf{A}\mathbf{x} \quad (144.2)$$

or, when written out more fully, by

$$\begin{Vmatrix} x_1' \\ x_2' \\ x_3' \end{Vmatrix} = \begin{Vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{Vmatrix} \cdot \begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix}$$

Because of the relationships satisfied by the direction cosines, the matrix $\mathbf{A}$ is an orthogonal matrix, so that $\tilde{\mathbf{A}} = \mathbf{A}^{-1}$ (see Sec. 140), and

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{x}' = \tilde{\mathbf{A}}\mathbf{x}' \tag{144.3}$$

In order that the number triple (v_1, v_2, v_3) constitute a vector $\boldsymbol{v}$, the numbers v_1, v_2, v_3 must assume the values v_1', v_2', v_3' when the system of reference axes is rotated, and be expressible in terms of v_1, v_2, v_3 in exactly the same way as the coordinates x_1', x_2', x_3' of the point P are expressed in terms of x_1, x_2, x_3, that is, by eq. (144.1) or eq. (144.2). Thus

$$\mathbf{v}' = \mathbf{A}\mathbf{v} \tag{144.4}$$

From the above definition, the three numbers specifying the coordinates of the point P constitute a vector. The point P can also be specified by the length of a line joining O and P, and the direction of this line. This leads to the geometric definition of a vector as a quantity having magnitude (distance between the points O and P) and direction. Thus, any pair of points located in space determines a vector, and two pairs of points which can be superimposed by a simple translation determine one and the same vector.

The magnitude or length of the vector OP is

$$+\sqrt{x_1^2 + x_2^2 + x_3^2}$$

In general, the length of a vector v, written $| \boldsymbol{v} |$, is

$$| \boldsymbol{v} | = \sqrt{v_1^2 + v_2^2 + v_3^2} \tag{144.5}$$

In Sec. 140, it was shown that an orthogonal transformation (such as the rotation of Cartesian axes) leaves the quadratic form $v_1^2 + v_2^2 + v_3^2$ invariant, which is to say

$$| \boldsymbol{v} | = \sqrt{v_1^2 + v_2^2 + v_3^2} = \sqrt{v_1'^2 + v_2'^2 + v_3'^2} = | \boldsymbol{v}' |$$

Consequently, the length of a vector is independent of the coordinate system used. Such a quantity is said to be an invariant or scalar quantity.

145. Algebra of vectors. The simplest operation to be performed on a vector $\boldsymbol{v}$ is that of multiplication with a scalar quantity c. If $\boldsymbol{v}$ is defined by the number triple (v_1, v_2, v_3), the vector $c\boldsymbol{v}$ is defined by the number triple (cv_1, cv_2, cv_3). Since v_1, v_2, v_3 satisfy the transformation eq. (144.4),

$$c\mathbf{v}' = c\mathbf{A}\mathbf{v} = \mathbf{A}c\mathbf{v} = \mathbf{A}c\mathbf{v} \tag{145.1}$$

so that the number triple (cv_1, cv_2, cv_3) defines a vector whose length is $c|\,\boldsymbol{v}\,|$ but whose direction is the same as that of $\boldsymbol{v}$.

Given two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$, we may form the linear combination

$$c\boldsymbol{v} + d\boldsymbol{w} \tag{145.2}$$

where c and d are scalars. The vector $c\boldsymbol{v} + d\boldsymbol{w}$ is defined by the number triple

$$cv_1 + dw_1,\ cv_2 + dw_2,\ cv_3 + dw_3$$

Since $$\mathbf{cv}' = \mathbf{Acv}$$

and $$\mathbf{dw}' = \mathbf{Adw}$$

then $$(\mathbf{cv}' + \mathbf{dw}') = \mathbf{A}(\mathbf{cv} + \mathbf{dw})$$

thereby showing that the number triple

$$cv_1 + dw_1,\ cv_2 + dw_2,\ cv_3 + dw_3$$

satisfies the transformation eq. (144.4), and actually defines a vector.

The expression (145.2) defines the processes of addition and subtraction of vectors. It is obvious that the geometric representation of the vector $c\boldsymbol{v} + d\boldsymbol{w}$ can be found according to Fig. 145.1.

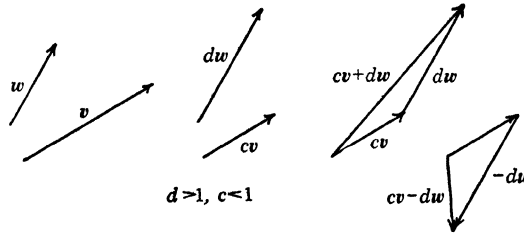


FIG. 145.1

The scalar product of two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ is written $\boldsymbol{v}\cdot\boldsymbol{w}$, and is defined to be

$$\boldsymbol{v}\cdot\boldsymbol{w} = v_1w_1 + v_2w_2 + v_3w_3 = \boldsymbol{w}\cdot\boldsymbol{v} \tag{145.3}$$

As indicated in Sec. 139, this bilinear form can be conveniently written in matrix form as

$$\tilde{\mathbf{v}}\mathbf{w} = \| v_1v_2v_3 \| \begin{Vmatrix} w_1 \\ w_2 \\ w_3 \end{Vmatrix} = \tilde{\mathbf{w}}\mathbf{v} \tag{145.4}$$

The quantity (145.3) is called the scalar product because the value of it is independent of the coordinate system. This may be readily verified, since $\boldsymbol{v}$ and $\boldsymbol{w}$ satisfy the transformation eq. (144.4), so that $\mathbf{v}' = \mathbf{Av}$ and $\mathbf{w}' = \mathbf{Aw}$, and

$$\begin{aligned} \boldsymbol{v}'\cdot\boldsymbol{w}' &= \mathfrak{v}'\mathfrak{w}' = (\widetilde{\mathbf{A}\mathfrak{v}})(\mathbf{A}\mathfrak{w}) \\ &= \tilde{\mathfrak{v}}\tilde{\mathbf{A}}\mathbf{A}\mathfrak{w} = \tilde{\mathfrak{v}}\mathfrak{w} = \boldsymbol{v}\cdot\boldsymbol{w} \end{aligned}$$

The property (140.9) of an orthogonal matrix was used, together with the fact that $\widetilde{\mathbf{A}\mathfrak{v}} = \tilde{\mathfrak{v}}\tilde{\mathbf{A}}$.

The geometric significance of the scalar product may be realized by noting that the quantities $v_1/|\,\boldsymbol{v}\,|$, $v_2/|\,\boldsymbol{v}\,|$, $v_3/|\,\boldsymbol{v}\,|$ are the direction cosines of the line segment representing the vector $\boldsymbol{v}$. From analytic geometry, it is known that the cosine of the angle θ between two straight lines is given by the sum of the products of their corresponding direction cosines. That is

$$\cos\theta = \frac{v_1w_1 + v_2w_2 + v_3w_3}{|\,\boldsymbol{v}\,|\cdot|\,\boldsymbol{w}\,|}$$

from which it follows that

$$\boldsymbol{v}\cdot\boldsymbol{w} = v_1w_1 + v_2w_2 + v_3w_3 = |\,\boldsymbol{v}\,|\cdot|\,\boldsymbol{w}\,|\cos\theta \tag{145.5}$$

Thus the scalar product gives the magnitude of the projection of the vector $\boldsymbol{w}$ in the direction of $\boldsymbol{v}$ times the magnitude of $\boldsymbol{v}$.

As a very simple example of the use of the scalar product, we may take a body acted upon by a constant force $\boldsymbol{F}$. If the body is moved in a direction specified by the direction of the vector $\boldsymbol{r}$ and a distance equal to the length of $\boldsymbol{r}$, the work done on the body is given by the scalar product of $\boldsymbol{F}$ and $\boldsymbol{r}$, or

$$W = \boldsymbol{F}\cdot\boldsymbol{r}$$

It follows immediately from eqs. (145.5) and (144.5), that the scalar product of a vector $\boldsymbol{v}$ with itself is equal to the square of its length.

$$\boldsymbol{v}\cdot\boldsymbol{v} = v_1^2 + v_2^2 + v_3^2 = |\,\boldsymbol{v}\,|^2 \tag{145.6}$$

If $\boldsymbol{v}\cdot\boldsymbol{w} = |\,\boldsymbol{v}\,|\,|\,\boldsymbol{w}\,|\cos\theta = 0$, either one of the vectors is zero, i.e., $v_1 = v_2 = v_3 = 0$, or they are perpendicular or orthogonal to each other.

The vector product of two vectors, $\boldsymbol{v}$ and $\boldsymbol{w}$, is written $\boldsymbol{v}\times\boldsymbol{w}$, and is a vector specified by the number triple

$$[(v_2w_3 - v_3w_2),\quad (v_3w_1 - v_1w_3),\quad (v_1w_2 - v_2w_1)]$$

In order to verify the assertion that these three numbers constitute a vector, we shall calculate $v_2'w_3' - v_3'w_2'$, which results from a rotation of coordinate axes. According to eq. (144.4), or eq. (144.1), we have

$$\begin{aligned} v_2'w_3' - v_3'w_2' &= \Big(\sum_j a_{2j}v_j\Big)\Big(\sum_k a_{3k}w_k\Big) - \Big(\sum a_{3j}v_j\Big)\Big(\sum_k a_{2k}w_k\Big) \\ &= \sum_{jk}(a_{2j}a_{3k} - a_{3j}a_{2k})v_jw_k \\ &= (a_{22}a_{33} - a_{32}a_{23})(v_2w_3 - v_3w_2) \\ &+ (a_{23}a_{31} - a_{33}a_{21})(v_3w_1 - v_1w_3) \\ &+ (a_{21}a_{32} - a_{31}a_{22})(v_1w_2 - v_2w_1) \end{aligned}$$

The coefficients of the terms $v_j w_k$ for which $j = k$ all vanish. Also the following relations between the various direction cosines,

$$\begin{aligned} a_{11} &= (a_{22}a_{33} - a_{32}a_{23}) \\ a_{12} &= (a_{23}a_{31} - a_{33}a_{21}) \\ a_{13} &= (a_{21}a_{32} - a_{31}a_{22}) \end{aligned}$$

allow us to write

$$\begin{aligned} v_2'w_3' - v_3'w_2' = a_{11}(v_2w_3 - v_3w_2) &+ a_{12}(v_3w_1 - v_1w_3) \\ &+ a_{13}(v_1w_2 - v_2w_1) \end{aligned}$$

This agrees with the first of the transformation eqs. (144.1). Similar equations for $v_3'w_1' - v_1'w_3'$ and $v_1'w_2' - v_1'w_3'$ can readily be obtained which agree with eqs. (144.1), showing that the number triple

$$\boldsymbol{v}\times\boldsymbol{w} = [(v_2w_3 - v_3w_2),\quad v_3w_1 - v_1w_3),\quad (v_1w_2 - v_2w_1)] \tag{145.7}$$

does indeed constitute a vector.

The geometric relation between the vector $\boldsymbol{v}\times\boldsymbol{w}$ and the two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ is partially indicated by evaluating the scalar products $\boldsymbol{v}\cdot(\boldsymbol{v}\times\boldsymbol{w})$ and $\boldsymbol{w}\cdot(\boldsymbol{v}\times\boldsymbol{w})$. By writing the numbers $v_2w_3 - v_3w_2$ etc., as determinants, the scalar products can be written

$$\boldsymbol{v}\cdot(\boldsymbol{v}\times\boldsymbol{w}) = v_1\begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} - v_2\begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} + v_3\begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix}$$

$$\boldsymbol{w}\cdot(\boldsymbol{v}\times\boldsymbol{w}) = w_1\begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} - w_2\begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} + w_3\begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix}$$

The right sides of these equations are just the expansions according to minors of the three-rowed determinants

$$\begin{vmatrix} v_1 & v_2 & v_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} w_1 & w_2 & w_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

respectively. Since each of these has two rows identical, their values are zero, and we have found that

$$\begin{aligned} \boldsymbol{v}\cdot(\boldsymbol{v}\times\boldsymbol{w}) &= 0 \\ \boldsymbol{w}\cdot(\boldsymbol{v}\times\boldsymbol{w}) &= 0 \end{aligned} \tag{145.8}$$

The vanishing of these two scalar products, according to eq. (145.5) shows that the vector $\boldsymbol{v}\times\boldsymbol{w}$ is perpendicular to both $\boldsymbol{v}$ and $\boldsymbol{w}$. This still does not determine the direction of $\boldsymbol{v}\times\boldsymbol{w}$ uniquely, since there are two directions which are at right angles to a plane containing $\boldsymbol{v}$ and $\boldsymbol{w}$. In order to determine the direction precisely, let us take two special vectors, $\boldsymbol{v} = (1, 0, 0)$ and $\boldsymbol{w} = (0, 1, 0)$, which are obviously unit vectors which are perpendicular to one another. In this case, $\boldsymbol{v}\times\boldsymbol{w}$ is, according to (145.7), represented by the number triple (0, 0, 1), so that $\boldsymbol{v}$, $\boldsymbol{w}$, and

$\boldsymbol{v}\times\boldsymbol{w}$ are unit vectors coinciding with the axes of the reference system. Since these were originally chosen to be a right-handed system, this means that $\boldsymbol{v}$, $\boldsymbol{w}$, and $\boldsymbol{v}\times\boldsymbol{w}$ form a right-handed system. Consequently, in the general case, if $\boldsymbol{v}$ is rotated so as to coincide with $\boldsymbol{w}$, the direction of $\boldsymbol{v}\times\boldsymbol{w}$ is that in which a right-handed screw would advance. This indicates that there would be a difference in direction between $\boldsymbol{v}\times\boldsymbol{w}$ and $\boldsymbol{w}\times\boldsymbol{v}$, and from eq. (145.7) it is apparent that

$$\boldsymbol{v}\times\boldsymbol{w} = -\boldsymbol{w}\times\boldsymbol{v} \tag{145.9}$$

so that $\boldsymbol{v}\times\boldsymbol{w}$ and $\boldsymbol{w}\times\boldsymbol{v}$ are oppositely directed.

The square of the magnitude of $\boldsymbol{v}\times\boldsymbol{w}$ is readily computed by evaluating the scalar product

$$\begin{aligned}(\boldsymbol{v}\times\boldsymbol{w})\cdot(\boldsymbol{v}\times\boldsymbol{w}) &= \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix}^2 + \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix}^2 + \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix}^2 \\ &= [v_1^2 + v_2^2 + v_3^2][w_1^2 + w_2^2 + w_3^2] \\ &\quad - [v_1w_1 + v_2w_2 + v_3w_3]^2 \\ &= |\,\boldsymbol{v}\,|^2\,|\,\boldsymbol{w}\,|^2 - |\,\boldsymbol{v}\,|^2\,|\,\boldsymbol{w}\,|^2\cos^2\theta \\ &= |\,v\,|^2\,|\,w\,|^2\sin^2\theta\end{aligned}$$

so that

$$|\,\boldsymbol{v}\times\boldsymbol{w}\,| = |\,\boldsymbol{v}\,|\,|\,\boldsymbol{w}\,|\sin\theta$$

where θ is the smallest angle between $\boldsymbol{v}$ and $\boldsymbol{w}$. The $\cos\theta$ was introduced in the above formula by using eq. (145.5).

As important examples of the practical use of the vector product, we may consider a particle of mass m moving with the velocity $\boldsymbol{v}$, whose position at any instant relative to the origin of a Cartesian system of axes is specified by the vector $\boldsymbol{r}$; the angular momentum of the particle is given by the vector $m\boldsymbol{r}\times\boldsymbol{v}$. Also, a particle with charge q moving with a velocity $\boldsymbol{v}$ through a magnetic field whose intensity is $\boldsymbol{H}$ experiences a force given by the vector product $q(\boldsymbol{v}\times\boldsymbol{H})$.

There are a number of useful combinations of the operations defined above, which are listed below for convenience.

$$\left.\begin{aligned}\boldsymbol{v}\cdot\boldsymbol{w} &= \boldsymbol{w}\cdot\boldsymbol{v} \\ \boldsymbol{v}\times\boldsymbol{w} &= -\boldsymbol{w}\times\boldsymbol{v} \\ \boldsymbol{u}\cdot(\boldsymbol{v}+\boldsymbol{w}) &= \boldsymbol{u}\cdot\boldsymbol{v} + \boldsymbol{u}\cdot\boldsymbol{w} \\ \boldsymbol{u}\times(\boldsymbol{v}+\boldsymbol{w}) &= \boldsymbol{u}\times\boldsymbol{v} + \boldsymbol{u}\times\boldsymbol{w} \\ \boldsymbol{u}\cdot\boldsymbol{v}\times\boldsymbol{w} &= \boldsymbol{v}\cdot\boldsymbol{w}\times\boldsymbol{u} = \boldsymbol{w}\cdot\boldsymbol{u}\times\boldsymbol{v} \\ \boldsymbol{u}\times(\boldsymbol{v}\times\boldsymbol{w}) &= (\boldsymbol{w}\times\boldsymbol{v})\times\boldsymbol{u} \\ &= (\boldsymbol{u}\cdot\boldsymbol{w})\boldsymbol{u} - (\boldsymbol{u}\cdot\boldsymbol{v})\boldsymbol{w}\end{aligned}\right\} \tag{145.10}$$

146. Vector fields. Many physical quantities such as electric field intensity, gravitational force, etc., have magnitudes and directions

which vary from point to point in a given region of space. Such a quantity can be represented by a vector whose components are functions of the coordinates specifying the position of the point in question. A region of space in which a vector $\boldsymbol{v}$ is associated with each point is called a vector field.

A vector $\boldsymbol{v}$ is said to be continuous if each element of the function triple (v_1, v_2, v_3) are continuous functions of one or more variables (see Secs. 5 and 6). If the vector $\boldsymbol{v}$ and consequently v_1, v_2, v_3 are continuous functions of a single variable t, the *derivative* of $\boldsymbol{v}$ with respect to t is written $d\boldsymbol{v}/dt$ and defined to be the function triple $(dv_1/dt, dv_2/dt, dv_3/dt)$. It is often convenient to represent the vector $d\boldsymbol{v}/dt$ by the matrix

$$\frac{d\boldsymbol{v}}{dt} = \left\| \begin{matrix} \dfrac{dv_1}{dt} \\ \dfrac{dv_2}{dt} \\ \dfrac{dv_3}{dt} \end{matrix} \right\| \tag{146.1}$$

That the function triple $(dv_1/dt, dv_2/dt, dv_3/dt)$ constitutes a vector for each value of t follows from the fact that v_1, v_2, v_3, satisfy the transformation eq. (144.4)

$$\mathbf{v}' = \mathbf{A}\mathbf{v}$$

for all values of t, so that

$$\frac{d\mathbf{v}'}{dt} = \frac{d}{dt}\mathbf{A}\mathbf{v} = \mathbf{A}\frac{d\mathbf{v}}{dt} = \mathbf{A}\left\| \begin{matrix} \dfrac{dv_1}{dt} \\ \dfrac{dv_2}{dt} \\ \dfrac{dv_3}{dt} \end{matrix} \right\|$$

Consequently, $dv_1/dt, dv_2/dt, dv_3/dt$ transform correctly under a rotation of the axes and thus form the components of a vector.

It is easily found that

$$\frac{d}{dt}\boldsymbol{v}\cdot\boldsymbol{w} = \frac{d\mathbf{v}}{dt}\cdot\boldsymbol{w} + \boldsymbol{v}\cdot\frac{d\mathbf{w}}{dt} \tag{146.2}$$

$$\frac{d}{dt}\boldsymbol{v}\times\boldsymbol{w} = \frac{d\mathbf{v}}{dt}\times\boldsymbol{w} + \boldsymbol{v}\times\frac{d\mathbf{w}}{dt} \tag{146.3}$$

The *integral* of the vector $\boldsymbol{v}$ is defined by the function triple

$$\int_0^t \boldsymbol{v}\,dt = \left\| \begin{matrix} \displaystyle\int_0^t v_1\,dt \\ \displaystyle\int_0^t v_2\,dt \\ \displaystyle\int_0^t v_3\,dt \end{matrix} \right\| \tag{146.4}$$

That the elements of the above matrix constitute a vector for all values of t can be readily shown.

147. The gradient, divergence, and curl. A vector quantity of considerable practical importance can be constructed from a scalar function φ of the three variables x_1, x_2, x_3. If x_1, x_2, x_3 are the coordinates of a point in one Cartesian system of axes, and x_1', x_2', x_3', are the coordinates of the same point in a Cartesian system of axes which has been rotated with respect to the first, so that a matrix equation of the form (144.2) connects the primed and unprimed coordinates,

$$\varphi'(x_1',x_2',x_3') = \varphi(x_1,x_2,x_3)$$

For a given point, the number triple $\partial\varphi/\partial x_1$, $\partial\varphi/\partial x_2$, $\partial\varphi/\partial x_3$ constitutes a vector. This assertion can be verified by showing that the matrix equation

$$\left\| \begin{matrix} \dfrac{\partial\varphi'}{\partial x_1'} \\ \dfrac{\partial\varphi'}{\partial x_2'} \\ \dfrac{\partial\varphi'}{\partial x_3'} \end{matrix} \right\| = \mathbf{A} \left\| \begin{matrix} \dfrac{\partial\varphi}{\partial x_1'} \\ \dfrac{\partial\varphi}{\partial x_2'} \\ \dfrac{\partial\varphi}{\partial x_3'} \end{matrix} \right\| \tag{147.1}$$

is satisfied where

$$\mathbf{x}' = \mathbf{A}\mathbf{x} \quad \text{or} \quad \mathbf{x} = \tilde{\mathbf{A}}\mathbf{x}' \tag{147.2}$$

From the rule for change of variable (see Sec. 13),

$$\frac{\partial\varphi'}{\partial x_i'} = \frac{\partial\varphi}{\partial x_1}\frac{\partial x_1}{\partial x_i'} + \frac{\partial\varphi}{\partial x_2}\frac{\partial x_2}{\partial x_i'} + \frac{\partial\varphi}{\partial x_3}\frac{\partial x_3}{\partial x_i'} \qquad i = 1, 2, 3$$

and by differentiating the second of eqs. (147.2) partially with respect to x_i', we find

$$\frac{\partial x_j}{\partial x_i'} = a_{ij}$$

From this result and the equation just above, it is at once obvious that eq. (147.1) is satisfied.

The vector whose components are $\partial\varphi/\partial x_1$, $\partial\varphi/\partial x_2$, $\partial\varphi/\partial x_3$, is called the *gradient* of $\varphi(x_1,x_2,x_3)$ and very often written grad φ.' From an operational standpoint it is convenient to use the operator matrix

$$\boldsymbol{\nabla} = \left\| \begin{matrix} \dfrac{\partial}{\partial x_1} \\ \dfrac{\partial}{\partial x_2} \\ \dfrac{\partial}{\partial x_3} \end{matrix} \right\| \tag{147.3}$$

in terms of which the vector we have called the gradient can be expressed as

$$\mathbf{grad}\ \varphi = \nabla\varphi = \begin{Vmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} \end{Vmatrix} \varphi = \begin{Vmatrix} \frac{\partial \varphi}{\partial x_1} \\ \frac{\partial \varphi}{\partial x_2} \\ \frac{\partial \varphi}{\partial x_3} \end{Vmatrix} \tag{147.4}$$

It has already been shown in Sec. 14, that *grad* φ is a vector whose direction is that of the greatest rate of increase of the function $\varphi(x_1,x_2,x_3)$ at a given point, and whose magnitude is the actual rate of increase or the directional derivative in the direction of maximum increase of φ.

In the theory of electricity, the importance of the gradient arises from the fact that the electric force $\boldsymbol{E}$ on a unit positive charge is expressible in terms of a scalar potential function V by the relation

$$\boldsymbol{E} = -\text{grad}\ V = \nabla V$$

Again in the theory of the flow of an ideal fluid, the velocity $\boldsymbol{v}$ at any point is expressible in terms of a velocity potential φ (see Sec. 125), by the relation

$$\boldsymbol{v} = -\mathbf{grad}\ \varphi$$

The total differential (see Sec. 15) $d\varphi$ of the scalar function φ can be expressed with the help of the gradient of φ. First, let the vector $\boldsymbol{dx}$ have components dx_1, dx_2, dx_3. Then

$$d\varphi = \mathbf{grad}\ \varphi \cdot \boldsymbol{dx} = \nabla\varphi \cdot \boldsymbol{dx} = \tilde{\nabla}\varphi\ \mathbf{dx} \tag{147.5}$$

$$= \begin{Vmatrix} \frac{\partial}{\partial x_1}, & \frac{\partial}{\partial x_2}, & \frac{\partial}{\partial x_3} \end{Vmatrix} \varphi \begin{Vmatrix} dx_1 \\ dx_2 \\ dx_3 \end{Vmatrix}$$

$$= \frac{\partial \varphi}{\partial x_1} dx_1 + \frac{\partial \varphi}{\partial x_2} dx_2 + \frac{\partial \varphi}{\partial x_3} dx_3$$

The *divergence* of a vector $\boldsymbol{v}$ can be defined most easily in terms of the operator ∇ by forming the scalar product of ∇ and $\boldsymbol{v}$.

$$\text{div}\ \boldsymbol{v} = \nabla \cdot \boldsymbol{v} = \tilde{\nabla}\mathbf{v} = \begin{Vmatrix} \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3} \end{Vmatrix} \begin{Vmatrix} v_1 \\ v_2 \\ v_3 \end{Vmatrix} \tag{147.6}$$

$$= \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \frac{\partial v_3}{\partial x_3}$$

This operation produces a scalar function of x_1, x_2, x_3 from the vector field $\boldsymbol{v}(x_1,x_2,x_3)$ in contrast to the gradient operation which produced a vector field from a scalar function.

The divergences of certain vector quantities provide a very convenient way of expressing important physical quantities. For example, if the vector $\boldsymbol{q}$ represents the mass of a fluid which flows through a unit area perpendicular to the direction of $\boldsymbol{q}$ in unit time, div $\boldsymbol{q}$ represents the decrease in the mass of fluid in a unit volume per unit time. This may readily be shown by computing the decrease in mass per unit time in a small parallelepiped of dimensions Δx_1, Δx_2, Δx_3 as shown in Fig. 147.1.

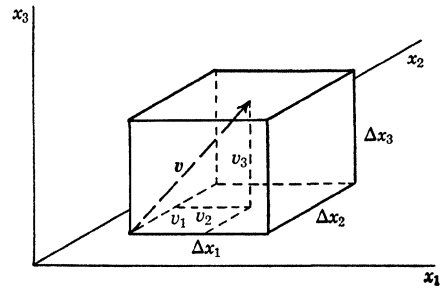


FIG. 147.1

If the velocity of the fluid is represented by the vector $\boldsymbol{v}$ with components v_1, v_2, v_3, the masses flowing per unit time through the faces of the parallelepiped at right angles to x_1 are

$$\rho v_1 \Delta x_2 \Delta x_3 = q_1 \Delta x_2 \Delta x_3 \tag{1}$$

$$\left[\rho v_1 + \frac{\partial}{\partial x_1}(\rho v_1)\Delta x_1\right] \Delta x_2 \Delta x_3 = \left(q_1 + \frac{\partial q_1}{\partial x_1}\Delta x_1\right) \Delta x_2 \Delta x_3 \tag{2}$$

where ρ is the fluid density. The first quantity is that flowing into the element of volume through the left face, and the second is the quantity flowing out through the right face. The decrease in mass due to flow through these two faces is the difference between (2) and (1), or

$$\frac{\partial q_1}{\partial x_1} \Delta x_1 \Delta x_2 \Delta x_3$$

When the decrease produced by flow through the remaining faces is computed, the total decrease of mass per unit time in the parallelepiped is

$$\left(\frac{\partial q_1}{\partial x_1} + \frac{\partial q_2}{\partial x_2} + \frac{\partial q_3}{\partial x_3}\right) \Delta x_1 \Delta x_2 \Delta x_3$$

or from eq. (147.6) the decrease in mass per unit volume per unit time is

$$\nabla \cdot \boldsymbol{q} = \operatorname{div} \boldsymbol{q} = \operatorname{div} \rho \boldsymbol{v}$$

This decrease in mass per unit time is just $-\partial\rho/\partial t$ so that

$$\operatorname{div} \boldsymbol{q} = \operatorname{div} \rho \boldsymbol{v} = -\frac{\partial \rho}{\partial t} \tag{147.7}$$

This equation is called the equation of continuity. The expression $\nabla \cdot \boldsymbol{q}$ is called the divergence, since it represents the net flow out of the unit volume.

An important vector function can be obtained with the help of the operator ∇. This is the vector function obtained by forming the vector product $\nabla \times \boldsymbol{v}$, where $\boldsymbol{v}$ is a vector function $\boldsymbol{v}(x_1,x_2,x_3)$. The vector $\nabla \times \boldsymbol{v}$ is called the curl of $\boldsymbol{v}$ and written curl $\boldsymbol{v}$. According to the definition of the vector product given in Sec. 145, the vector representing the curl $\boldsymbol{v}$ at any point (x_1,x_2,x_3) is specified by the number triple

$$\left[\left(\frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3}\right), \left(\frac{\partial v_1}{\partial x_3} - \frac{\partial v_3}{\partial x_1}\right), \left(\frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2}\right)\right]$$

The vector field curl $\boldsymbol{v}$ enters into many physical laws and equations, especially in connection with hydrodynamics and electromagnetic theory. The physical interpretation of curl $\boldsymbol{v}$ can be indicated as follows.

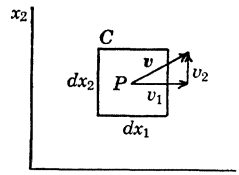


FIG. 147.2

Let us consider a flowing mass of fluid, the flow being described at all points in a given region by the velocity $\boldsymbol{v}$ of a fluid particle at the point in question. The velocity at a given point may or may not vary with the time. For simplicity let us suppose there is no flow in the direction x_3, so that $\boldsymbol{v}$ will be a function of x_1, x_2, and t only. At a given instant let the velocity at a point P be $(v)_P$. At this instant, we shall compute the line integral of $\boldsymbol{v}$ around the boundary C of the small element of area shown in Fig. 147.2. In hydrodynamics, this line integral $\int_C \boldsymbol{v} \cdot d\boldsymbol{s}$ is called the circulation because it furnishes a measure of the rotation of the fluid about an axis through the point P perpendicular to the element of area enclosed by C. For small dx_1 and dx_2, the component of velocity parallel to x_1 on the lower side of the element of area is

$$(v_1)_P + \left(\frac{\partial v_1}{\partial x_2}\right)_P \left(\frac{-dx_2}{2}\right)$$

The contribution to the line integral over this lower side will then be

$$\left[(v_1)_P + \left(\frac{\partial v_1}{\partial x_2}\right)_P \left(\frac{-dx_2}{2}\right)\right] dx_1$$

Carrying out this procedure over the whole boundary,

$$\begin{aligned}\int_C \boldsymbol{v}\cdot d\boldsymbol{s} &= \left[(v_1)_P + \left(\frac{\partial v_1}{\partial x_2}\right)_P \left(\frac{-dx_2}{2}\right)\right]dx_1 + \left[(v_2)_P + \left(\frac{\partial v_2}{\partial x_1}\right)_P \frac{dx_1}{2}\right]dx_2 \\ &+ \left[(v_1)_P + \left(\frac{\partial v_1}{\partial x_2}\right)_P \left(\frac{dx_2}{2}\right)\right](-dx_1) \\ &+ \left[(v_2)_P + \left(\frac{\partial v_2}{\partial x_1}\right)_P \left(\frac{-dx_1}{2}\right)\right](-dx_2) \\ &= \left[\left(\frac{\partial v_2}{\partial x_1}\right)_P - \left(\frac{\partial v_1}{\partial x_2}\right)_P\right]dx_1\,dx_2 \qquad (147.8)\end{aligned}$$

The term in the bracket on the right side of the last equation is just the component in the x_3 direction of the curl $\boldsymbol{v}$ at the point P. Hence the circulation in the neighborhood of a point about an axis through the point is measured by the component of curl $\boldsymbol{v}$ in the direction of that axis. In general, the rotation at a given instant of a fluid at a given point is given by curl $\boldsymbol{v}$. Consequently, a flow for which curl $\boldsymbol{v} = 0$ at all points is called irrotational.

Equation (147.8) is a special case of Stokes's theorem (see Sec. 70), which, expressed in vector form, becomes

$$\int_C \boldsymbol{v}\cdot d\boldsymbol{s} = \int_S \text{curl } \boldsymbol{v}\cdot d\boldsymbol{S}$$

where the surface integral is taken over the surface bounded by the curve C.

148. Useful vector formulas. A number of expansion formulas involving the gradient, divergence, and curl are listed below for the reader's convenience.

$$\begin{aligned}
&\text{grad } (\varphi+\psi) = \nabla(\varphi+\psi) &&= \nabla\varphi+\nabla\psi \\
&\text{div } (\boldsymbol{v}+\boldsymbol{w}) &&= \nabla\cdot(\boldsymbol{v}+\boldsymbol{w}) = \nabla\cdot\boldsymbol{v}+\nabla\cdot\boldsymbol{w} \\
&\text{curl } (\boldsymbol{v}+\boldsymbol{w}) &&= \nabla\times(\boldsymbol{v}+\boldsymbol{w}) = \nabla\times\boldsymbol{v}+\nabla\times\boldsymbol{w} \\
&\text{grad } \varphi\psi &&= (\nabla\varphi\psi) = \psi\nabla\varphi+\varphi\nabla\psi \\
&\text{div } \varphi\boldsymbol{v} &&= \nabla\cdot\varphi\boldsymbol{v} = \nabla\varphi\cdot\boldsymbol{v}+\varphi\nabla\cdot\boldsymbol{v} \\
&\text{curl } \varphi\boldsymbol{v} &&= \nabla\times\varphi\boldsymbol{v} = \nabla\varphi\times\boldsymbol{v}+\varphi\nabla\times\boldsymbol{v} \\
&\text{grad } \boldsymbol{v}\cdot\boldsymbol{w} &&= \nabla(\boldsymbol{v}\cdot\boldsymbol{w}) = (\boldsymbol{v}\cdot\nabla)\boldsymbol{w}+(\boldsymbol{w}\cdot\nabla)\boldsymbol{v} \\
&&&\quad +\boldsymbol{v}\times(\nabla\times\boldsymbol{w})+\boldsymbol{w}\times(\nabla\times\boldsymbol{v}) \\
&\text{div } (\boldsymbol{v}\times\boldsymbol{w}) &&= \nabla\cdot(\boldsymbol{v}\times\boldsymbol{w}) = \boldsymbol{w}\cdot\nabla\times\boldsymbol{v}-\boldsymbol{v}\cdot\nabla\times\boldsymbol{w} \\
&\text{curl } (\boldsymbol{v}\times\boldsymbol{w}) &&= \nabla\times(\boldsymbol{v}\times\boldsymbol{w}) = \boldsymbol{v}(\nabla\cdot\boldsymbol{w})-\boldsymbol{w}(\nabla\cdot\boldsymbol{v}) \\
&&&\quad +(\boldsymbol{w}\cdot\nabla)\boldsymbol{v}-(\boldsymbol{v}\cdot\nabla)\boldsymbol{w} \\
&\text{curl curl } \boldsymbol{v} &&= \nabla\times(\nabla\times\boldsymbol{v}) = \nabla\nabla\cdot\boldsymbol{v}-(\nabla\cdot\nabla)\boldsymbol{v} \\
&\text{div grad } \varphi &&= \nabla\cdot\nabla\varphi = (\nabla\cdot\nabla)\varphi \\
&\text{curl grad } \varphi &&= \nabla\times\nabla\varphi = 0 \\
&\text{div curl } \boldsymbol{v} &&= \nabla\cdot\nabla\times\boldsymbol{v} = 0
\end{aligned}$$

where φ and ψ are scalar functions, and $\boldsymbol{v}$ and $\boldsymbol{w}$ are vector point functions.

149. The tensor concept and definition. All relations between physical quantities which constitute physical laws or are derivable from them must be independent of the measuring scale used to measure the quantities involved and the nature of the frame of reference used. As we shall see below it is the most important property of tensors that when the measuring scale or frame of reference is changed the tensor is transformed in such a way that its magnitude in a kind of absolute sense remains unchanged. If one expresses the quantities which are related in a physical law as tensors, as a consequence one can be sure that the relation holds in an absolute sense, that is, the relation will not change when the units of measurement are changed or the coordinate frame of reference is changed. Thus when it is important to emphasize the logical and philosophical nature of the laws of physics it is convenient to express the relations between the physical quantities in the form of a tensor equation.

If in Sec. 144, the components v_1, v_2, v_3 of a vector $\boldsymbol{v}$ are considered to be functions of (x_1, x_2, x_3), i.e., functions of the position of a point in the coordinate system C, the vector $\boldsymbol{v}$ is a special case of a contravariant tensor of order one. The reason arises from the fact that the coordinates of the point and the components of the vector transform in the same manner when referred to a new coordinate system C'. Upon writing the transformation equations in matrix form [see eq. (144.2)], this means that

$$\mathbf{x}' = \mathbf{A}\mathbf{x} \tag{149.1}$$

$$\mathbf{v}' = \mathbf{A}\mathbf{v} \tag{149.2}$$

where $\mathbf{A}$ is the matrix of the transformation.

In order to proceed to a general definition of a contravariant tensor of order one it will be supposed that the coordinates of a point in the n-dimensional coordinate system C are designated by $x^1, x^2, x^3, \ldots, x^n$*, and that the transformation to another coordinate system C' is accomplished by the matrix equation

$$\mathbf{x}' = \mathbf{A}\mathbf{x} \tag{149.3}$$

and that the inverse matrix exists [see eq. (138.11)], so that

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{x}' \tag{149.4}$$

Given a set of n functions $T^1(x^1,x^2,\ldots,x^n),\ldots,T^n(x^1,x^2,\ldots,x^n)$ of the coordinates $(x^1,x^2,\ldots,x^n)$ in system C, then if the coordinate system is

* The use of superscripts to enumerate the coordinates instead of subscripts as in previous chapters is done in order to conform to customary usage in tensor analysis.

transformed to a new system C' in accordance with eq. (149.1) and a new system of functions

$$T'^1(x'^1,x'^2,\ldots,x'^n),\ldots,T'^n(x'^1,x'^2,\ldots,x'^n)$$

is obtained, each of which is related to those in the system C by the transformation law

$$T'^\alpha(x'^1,x'^2,\ldots,x'^n) = \sum_{i=1}^{n} T^i(x^1,x^2,\ldots,x^n)\frac{\partial x'^\alpha}{\partial x^i} \qquad \alpha = 1, 2, \ldots, n \qquad (149.5)$$

then $T^1, \ldots, T^n$ are the components of a *contravariant tensor* or order 1 in the coordinate system C.

Since the transformation (149.2) is a special case of (149.5), a contravariant tensor of order one is sometimes called a contravariant vector.

Equation (149.5) suggests that in an n-dimensional coordinate system C the differentials $dx^1, dx^2, \ldots, dx^n$ considered as defined at each point in C form the components of a contravariant tensor of order one. That this is so can be seen from the fact that under a transformation of coordinates to C' the new differentials (see Sec. 15) are given by

$$dx'^\alpha = \sum_{i=1}^{n} \frac{\partial x'^\alpha}{\partial x^i} dx^i \qquad \alpha = 1, 2, \ldots, n \qquad (149.6)$$

so that the differentials dx^i transform as they should according to eq. (149.5).

Another example of a particular contravariant tensor of order one is furnished by the velocity of an element of fluid which is a function of the coordinates x^1, x^2, x^3 in a system C. The components of velocity of the element in C are dx^1/dt, dx^2/dt, dx^3/dt. If a transformation is made to a new system of coordinates C', the new velocities are given by

$$\frac{dx'^\alpha}{dt} = \sum_{i=1}^{3} \frac{\partial x'^\alpha}{\partial x^i}\frac{dx^i}{dt} \qquad \alpha = 1, 2, 3 \qquad (149.7)$$

Comparing this transformation with (149.5), it is seen that the components of velocity of an element of fluid in motion form the components of a contravariant tensor of order one.

Another important tensor of order one is the *covariant tensor.*

Given a set of n functions

$$T_1(x^1,x^2,\ldots,x^n),\ldots,T_n(x^1,x^2,\ldots,x^n)$$

of the coordinates $(x^1,x^2,\ldots,x^n)$ of system C, then if the coordinate system is transformed to a new system C' in accordance with eq. (149.1) and a new system of functions

$$T'_1(x'^1,x'^2,\ldots,x'^n),\ldots,T'^1_n(x'^1,x'^2,\ldots,x'^n)$$

is obtained, each of which is related to those in system C by the transformation law,

$$T'_\alpha(x'^1,x'^2,\ldots,x'^n) = \sum_{i=1}^{n} T_i(x^1,x^2,\ldots,x^n)\frac{\partial x^i}{\partial x'^\alpha}$$

$$\alpha = 1, 2, \ldots, n \qquad (149.8)$$

Then $T_1, \ldots, T_n$ are the components of a covariant tensor of order one.

The partial derivatives of a function $f(x^1,x^2,x^3)$ form the components of a covariant tensor of order one. This can be readily seen as follows. If a transformation of coordinates is effected, $f(x^1,x^2,x^3)$ goes over into a new function $f'(x'^1, x'^2, x'^3)$, and

$$\frac{\partial f'(x'^1,x'^2,x'^3)}{\partial x'^\alpha} = \sum_{i=1}^{3} \frac{\partial f}{\partial x^i}\frac{\partial x^i}{\partial x'^\alpha} \qquad (149.9)$$

Thus $\partial f/\partial x^i$ etc. transform in accordance with (149.8), and hence $\partial f/\partial x^1$, $\partial f/\partial x^2$, $\partial f/\partial x^3$ form the components of a covariant tensor.

If $T^1, T^2, \ldots, T^n$ and $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_n$ are the components of any contravariant and covariant tensors, respectively, $\sum_i^n T^i\mathcal{T}_i$ is a scalar point function or an invariant with respect to transformations of the coordinate system.

150. Definition of tensors of higher order. The concept of a tensor as given in the preceding section can readily be generalized as follows.

Given a set of n^2 functions

$$T^{11}(x^1,x^2,\ldots,x^n)\ldots,T^{nn}(x^1,x^2,\ldots,x^n)$$

of the coordinates $x^1, x^2, \ldots, x^n$ in system C, then if the coordinate system is transformed to a new system C' in accordance with eq. (149.1) and if the new set of functions T' is obtained, each of which is related to those in the system C by the transformation law,

$$T'^{\alpha\beta} = \sum_{i,j}^{n} T^{ij}\frac{\partial x'^\alpha}{\partial x^i}\frac{\partial x'^\beta}{\partial x^j} \qquad \alpha, \beta = 1, 2, \ldots, n \qquad (150.1)$$

the n^2 functions T^{ij} are the components of a contravariant tensor of order two.

Also if the set of n^2 functions T_{ij} transform into the n^2 functions $T'_{\alpha\beta}$ according to the law

$$T'_{\alpha\beta} = \sum_{i,j}^{n} T_{ij}\frac{\partial x^i}{\partial x'^\alpha}\frac{\partial x^j}{\partial x'^\beta} \qquad (150.2)$$

the functions T_{ij} form the components of a covariant tensor of order two.

Also if the n^2 functions T^i_j transform into the n^2 functions T'^{α}_{β} according to the law

$$T'^{\alpha}_{\beta} = \sum_{i,j}^{n} T^i_j \frac{\partial x'^{\alpha}}{\partial x^i} \frac{\partial x^j}{\partial x'^{\beta}} \tag{150.3}$$

the n^2 functions T^i_j form the components of a mixed tensor of order two, contravariant of order one, and covariant of order one.

Tensors of higher orders can be defined in a similar way. Such tensors would have components designated T^{ijk}, T_{ijk}, T^{ij}_{klm}, etc. The tensor with components T^{ij}_{klm} would be a mixed tensor of order five, having n^5 components, contravariant of order two, and covariant of order three.

A simple example of a tensor of order two in a space of three dimensions is afforded by the stresses in a solid material or viscous fluid. The nine components of this tensor are usually designated P_{xx}, P_{xy}, P_{xz}, P_{yz}, $\ldots$, P_{zz}, where P_{xx} is the force per unit area in the x direction exerted on the interface perpendicular to the x axis, and P_{xy} is the shearing stress in the y direction at an interface perpendicular to the x direction. The reader may readily determine whether this is a contravariant or covariant tensor of order two.

151. Addition, multiplication and contraction of tensors. It is obvious from the definitions of tensors that the *sum* or *difference* of two tensors of the same type and order is again a tensor of the same type and order. For example, if T_{ij} and T'_{ij} are the components of two covariant tensors of order two, the $\tau_{ij} = T_{ij} + T'_{ij}$ are the components of a covariant tensor of order two.

The product of *any* two tensors is a tensor. For example if T^{ij} are the components of a contravariant tensor of order two and T_{klm} are the components of a covariant tensor of order three,

$$\tau^{ij}_{klm} = T^{ij} T_{klm}$$

where the τ^{ij}_{klm} is a mixed tensor of order five.

If a transformation of coordinates is performed, the above components become

$$\begin{aligned}
\tau'^{\alpha\beta}_{\gamma\delta\epsilon} &= T'^{\alpha\beta} T'_{\gamma\delta\epsilon} \\
&= \left(\sum_{ij}^{n} T^{ij} \frac{\partial x'^{\alpha}}{\partial x^i} \frac{\partial x'^{\beta}}{\partial x^j}\right)\left(\sum_{klm}^{n} T_{klm} \frac{\partial x^k}{\partial x'^{\gamma}} \frac{\partial x^l}{\partial x'^{\delta}} \frac{\partial x^m}{\partial x'^{\epsilon}}\right) \\
&= \sum_{ijklm}^{n} \tau^{ij}_{klm} \frac{\partial x'^{\alpha}}{\partial x^i} \frac{\partial x'^{\beta}}{\partial x^j} \frac{\partial x^k}{\partial x'^{\gamma}} \frac{\partial x^l}{\partial x'^{\delta}} \frac{\partial x^m}{\partial x'^{\epsilon}}
\end{aligned}$$

This shows that the product of two tensors defined above is also a tensor.

The *contraction* of a tensor takes place when the components of a

tensor of order p over any contravariant and covariant index are summed. The resulting functions are components of a tensor of order $p - 2$. Thus if T^{ij}_{klm} are the n^5 components of a tensor of order five,

$$\tau^{i}_{kl} = \sum_{j=1}^{n} T^{ij}_{klj}$$

gives a tensor with n^3 components of order three. If a transformation of coordinates is made, the above will transform into

$$\begin{aligned}
\tau'^{\alpha}_{\gamma\delta} &= \sum_{\beta} T'^{\alpha\beta}_{\gamma\delta\beta} \\
&= \sum_{\beta} \left(\sum_{ijklm} T^{ij}_{klm} \frac{\partial x'^{\alpha}}{\partial x^{i}} \frac{\partial x'^{\beta}}{\partial x^{j}} \frac{\partial x^{k}}{\partial x'^{\gamma}} \frac{\partial x^{l}}{\partial x'^{\delta}} \frac{\partial x^{m}}{\partial x'^{\beta}} \right) \\
&= \sum_{ijkm} T^{ij}_{klm} \delta^{m}_{j} \frac{\partial x'^{\alpha}}{\partial x^{i}} \frac{\partial x^{k}}{\partial x'^{\gamma}} \frac{\partial x^{l}}{\partial x'^{\delta}} \\
&= \sum_{ikl} \left(\sum_{j} T^{ij}_{klj} \right) \frac{\partial x'^{\alpha}}{\partial x^{i}} \frac{\partial x^{k}}{\partial x'^{\gamma}} \frac{\partial x^{l}}{\partial x'^{\delta}} \\
&= \sum_{ikl} \tau^{i}_{kl} \frac{\partial x'^{\alpha}}{\partial x^{i}} \frac{\partial x^{k}}{\partial x'^{\gamma}} \frac{\partial x^{l}}{\partial x'^{\delta}}
\end{aligned}$$

Use has been made of the fact that

$$\sum_{\beta} \frac{\partial x'^{\beta}}{\partial x^{j}} \frac{\partial x^{m}}{\partial x'^{\beta}} = \delta^{m}_{j} = \begin{cases} 1 & \text{if } j = m \\ 0 & \text{if } j \neq m \end{cases}$$

The above operation shows that a tensor which is contravariant of order one and covariant of order two can be obtained by contracting a tensor which is contravariant of order two and covariant of order three.

The operations of multiplication and contraction can be combined to obtain new tensors. Thus

$$\tau_{kl} = \sum_{i} T^{i} T'_{ikl}$$

or

$$\tau^{i}_{lm} = \sum_{jk} T^{ijk} T'_{jklm}$$

The combined process of multiplication and contraction is sometimes called *inner* multiplication.

152. Fundamental and associate tensors. In accordance with eq. (51.1) the square of the line element in a given coordinate system can be written in the form

$$(ds)^2 = \sum_{ij} g_{ij} dx^{i} dx^{j} \tag{152.1}$$

Since an element of length along a given curve cannot depend on the particular coordinate system, $(ds)^2$ is an invariant or scalar quantity.

From eq. (149.6) we know that the dx^i's are the components of a contravariant tensor. From the statement at the end of Sec. 149 we can conclude that $\sum_i g_{ij}\,dx^i$ must be a covariant tensor in order that $(ds)^2$ be a scalar. Furthermore, since dx^i are the components of a contravariant tensor, the g_{ij} must be the component of a covariant tensor of order two. The tensor with components g_{ij} is symmetric and is usually called the fundamental metric tensor for n-dimensional Euclidean space, and ds is the magnitude of the tensor whose components are dx^i.

If we form the determinant

$$g = \begin{vmatrix} g_{11} & \cdots & g_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ g_{n1} & \cdots & g_{nn} \end{vmatrix} \tag{152.2}$$

from the components of the fundamental metric tensor and let

$$g^{ij} = \frac{1}{g}\cdot(\text{the cofactor of } g_{ij} \text{ in } g) \tag{152.3}$$

the g^{ij} turn out to be the components of a symmetric fundamental contravariant tensor of order two, and

$$\sum_k g_{jk}g^{ik} = \delta^i_j \tag{152.4}$$

If T^i are the components of a contravariant tensor of order one,

$$\sum_i g_{ij}T^i = \sum_i g_{ji}T^i = T_j \tag{152.5}$$

The tensor with components T_j is said to be the tensor associated with the tensor with components T^i, and the T_j are said to be the covariant components of the tensor having contravariant components T^i.

Also, if we associate the tensor with components T^i with that with components T_i with the help of a fundamental contravariant tensor,

$$\begin{aligned} T^i &= \sum_j g^{ij}T_j \\ &= \sum_j g^{ij} \sum_k g_{jk}T^k \\ &= \sum_k \delta^i_k T^k = T^i \end{aligned}$$

Thus contractings with respect to g_{ij} and g^{ij} represent inverse operations.

The *magnitude* of a tensor can be defined with the help of the fundamental contravariant tensor g^{ij} or in terms of the associate tensor. If T_i are the components of a covariant tensor, the square of its magnitude is defined to be

$$T^2 = \sum_{ij} g^{ij} T_i T_j \tag{152.6}$$

It is readily seen that T^2 is an invariant and gives the magnitude of the tensor with components T_i in every coordinate system.

153. The covariant derivative. It was shown in Sec. 149 that the derivatives of a scalar formed the components of a covariant vector or tensor of order one. However, if the T_i are the components of a covariant tensor of order one, the n^2 derivatives $\partial T_i/\partial x^j$ do not form the components of a tensor, for it may easily be shown that the $\partial T_i/\partial x^j$ do not transform in the correct way when a transformation of coordinates is effected. However, there is a useful quantity involving the $\partial T_i/\partial x^j$ that does transform correctly and thus constitutes a tensor. The quantities

$$T_{i,j} \equiv \frac{\partial T_i}{\partial x^j} - \sum_k T_k \{ij,k\} \tag{153.1}$$

constitute the components of a second-order covariant tensor. In the above equation the Christoffel 3-index symbol $\{ij,k\}$ has been used for convenience. The Christoffel 3-index symbols are defined as follows.

$$[ij,k] = \frac{1}{2}\left(\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}\right) \tag{153.2}$$

$$\{ij,k\} = \sum_l g^{kl}[ij,l] \tag{153.3}$$

The tensor with components $T_{i,j}$ is called the covariant derivative of the tensor with components T_i with respect to the fundamental tensor g_{ij}.

Rules for Performing Covariant Differentiations.

1. To obtain the covariant derivative of any tensor with components $T_{\cdots}^{\cdots}$ with respect to x^i,

(a) Take the ordinary derivative $\dfrac{\partial T_{\cdots}^{\cdots}}{\partial x^i}$.

(b) Add to this a term $-\sum_\gamma \{ij,\gamma\} T_{\cdots;\gamma}^{\cdots}$ for each covariant suffix $T_{\cdots;i}^{\cdots}$.

(c) Add to all of the above a term $\sum_\gamma \{\gamma j,i\} T_{\cdots}^{\cdots;\gamma}$ for each contravariant index $T_{\cdots}^{\cdots;i}$. For example, the covariant derivative of T^{ij} with respect to x^k is

$$T^{ij}_{k} = \frac{\partial T^{ij}}{\partial x^k} + \sum_{\gamma} \{\gamma k,i\} T^{\gamma j} + \sum_{\gamma} \{\gamma k,j\} T^{i\gamma}$$

2. The covariant derivative of a product is found by covariant differentiation of each factor just as in ordinary differentiation.

3. The fundamental tensors are regarded as constants during the covariant differentiation.

4. The covariant derivative of a salar or invariant is its ordinary derivative.

5. In performing successive covariant differentiations, the order of differentiation is not interchangeable.

154. The nature of the covariant derivative. The peculiar importance of the covariant derivative arises from the fact that it can be used to express the rate of change of a physical quantity in a kind of absolute sense regardless of the curvilinear nature of the frame of reference or whether the frame of reference is undergoing acceleration, etc.

Referring to eq. (153.1), which defines the covariant derivative of T_i, it is seen that the first term is the ordinary rate of change of T_i with respect to the coordinate x^j. In order to obtain the true rate of change of T_i in all coordinate systems one must subtract from this term the term $\sum_k T_k \{ij,k\}$. This latter is a kind of spurious term which takes into account the curvilinear nature of the coordinate system used. In ordinary Cartesian coordinates every $\{ij,k\}$ vanishes, so that the entire term to be subtracted is zero. Hence in this case the covariant derivative reduces to the ordinary rate of change of the rectangular components of T.

From this we infer that whenever physical laws are expressed as tensor equations in the spirit of Sec. 149 and the law relates such things as force to the time rate of change of momentum or the space rate of change of potential, these rates of change are to be represented by the covariant derivative of the quantity in question.

SUGGESTED REFERENCES

E. B. Wilson, *Vector Analysis*, New Haven, Yale University Press, 1901.

J. H. Taylor, *Vector Analysis*, New York, Prentice-Hall, Inc., 1939.

Handbuch der Physik Vol. 3, Chap. 5, Berlin, Julius Springer, 1928.

H. B. Phillips, *Vector Analysis*, New York, John Wiley & Sons, Inc., 1933.

H. V. Craig, *Vector and Tensor Analysis*, New York, McGraw-Hill Book Company, Inc., 1943.

PROBLEMS

145.1. Let C be a coordinate system at rest, C' be a coordinate system fixed in a rotating rigid body. Let (x_1,x_2,x_3) and (x_1',x_2',x_3') be the coordinates of a point on the rigid body in the two systems, and let $\boldsymbol{x}' = \mathbf{A}\boldsymbol{x}$. Show that one can write $\dot{\boldsymbol{x}} = \boldsymbol{\omega}\times\boldsymbol{x}$ and express $\boldsymbol{\omega}$ in terms of $\mathbf{A}$.

145.2. Let $\mathbf{A}(t + dt) - \mathbf{A}(t)$ be the matrix for a rotation through an infinitesimal angle $d\theta$ around the axis whose direction cosines are α, β, γ. Find the components of $\boldsymbol{\omega}$.

145.3. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be the vectors $(1,1,0)$ and $(1,-1,0)$ and let $\boldsymbol{w} = \boldsymbol{u}\times\boldsymbol{v}$. Show that under the transformation $x_1' = -x_1$, $x_2' = -x_2$, $x_3' = -x_3$; $\boldsymbol{w}' = \boldsymbol{w}'\times\boldsymbol{v}'$. Sketch the systems and vectors. Hence under such an inversion, the vector product is not a vector. A quantity transforming like $\boldsymbol{w}$ above is called a pseudovector, a relative vector, or a vector density. Show that $\boldsymbol{w}\cdot\boldsymbol{u}\times\boldsymbol{v}$ is a pseudoscalar.

145.4 Prove that

$$(\boldsymbol{u}\times\boldsymbol{v})\times\boldsymbol{w} = (\boldsymbol{u}\cdot\boldsymbol{w})\boldsymbol{v} - (\boldsymbol{v}\cdot\boldsymbol{w})\boldsymbol{u}$$

147.1. If a fluid is rotating like a rigid body, find the curl of its velocity at each point.

148.1. Prove $\nabla\cdot\nabla\times\boldsymbol{v} = 0$.

148.2. Prove $\nabla\times\nabla\varphi = 0$.

148.3. Prove $\nabla\times(\nabla\times\boldsymbol{v}) = \nabla(\nabla\cdot\boldsymbol{v}) - (\nabla\cdot\nabla)\boldsymbol{v}$.

148.4. Prove

$$\nabla\times(\boldsymbol{u}\times\boldsymbol{v}) = \boldsymbol{u}(\nabla\cdot\boldsymbol{v}) - \boldsymbol{v}(\nabla\cdot\boldsymbol{u}) + (\boldsymbol{v}\cdot\nabla)\boldsymbol{u} - (\boldsymbol{u}\cdot\nabla)\boldsymbol{v}.$$

148.5. If $\nabla\cdot\boldsymbol{B} = 0$, find a vector $\boldsymbol{A}$ such that $\boldsymbol{B} = \nabla\times\boldsymbol{A}$. If it is given that $A_1 = f(x,y,z)$, find $\boldsymbol{A}$.

150.1. Show that the stress components

$$p_{xx},\ p_{xy},\ p_{xz},\ p_{yx},\ p_{yy},\ p_{yz},\ p_{zx},\ p_{zy},\ p_{zz}$$

are the components of a tensor of order two. Also determine whether the tensor is covariant or contravariant.

152.1. Show that the functions g^{ij} defined by eq. (152.3) are the components of a contravariant tensor of order two.

152.2. Show that the tensors with components g_{ij} and g^{ij} are symmetric.

152.3. If T_j are the components of a covariant tensor, show that $T^2 = \sum_{i,j} g^{ij}T_iT_j$ is invariant for all transformations of coordinates.

153.1. Show that the covariant derivatives of the fundamental tensor g^{ij} are zero.

153.2. Show that the covariant derivative of a scalar function is identical with its ordinary derivative.

153.3. LaPlace's equation in Cartesian coordinates is

$$\frac{\partial^2\varphi}{\partial t^2} - \frac{\partial^2\varphi}{\partial x^2} - \frac{\partial^2\varphi}{\partial y^2} - \frac{\partial^2\varphi}{\partial z^2} = 0$$

Write this equation in terms of the appropriate covariant derivatives. Show that the resulting equation would hold in all coordinate systems and that it can be written in the form

$$g^{ij}\left(\frac{\partial^2\varphi}{\partial x_i\,\partial x_j} - \{ij,k\}\frac{\partial\varphi}{\partial x_k}\right) = 0$$

Use this result to transform Laplace's equation from Cartesian to (a) cylindrical coordinates, (b) spherical coordinates, and (c) parabolic coordinates.

Chapter 12

ORTHONORMAL FUNCTION SYSTEMS

155. Definitions. Two functions $f(x)$ and $g(x)$ defined throughout the finite interval $a \leq x \leq b$ are said to be orthogonal to one another if

$$\int_a^b f(x)g(x)dx = 0 \tag{155.1}$$

and $f(x)$ is said to be normalized when

$$\int_a^b f^2(x)dx = 1 \tag{155.2}$$

These terms arise in a natural way from an analogy with vectors in a vector space (see Sec. 134) wherein a vector $\boldsymbol{f}$ with components $f_1, f_2, f_3, \ldots, f_n$ is perpendicular to or orthogonal to a vector $\boldsymbol{g}$ with components $g_1, g_2, \ldots, g_n$ when the scalar product $(\boldsymbol{f}\cdot\boldsymbol{g}) = f_1g_1 + f_2g_2 + \ldots + f_ng_n = 0$. When f and g are functions of x and this sum is replaced by an integral, one sees the formal analogy of the vector scalar product with the operation

$$\int_a^b f(x)g(x)dx$$

It is natural to use the geometrical terminology and call it the scalar product of the two functions and to say that $f(x)$ is orthogonal to $g(x)$ when the scalar product so defined vanishes.

This analogy can be carried much further and this will be done to a certain extent in the following paragraphs. For example, the scalar product of the vector $\boldsymbol{f}$ with itself, namely, $f_1^2 + f_2^2 + \ldots + f_n^2$ is the square of the length of the vector $\boldsymbol{f}$, so that we may call $\left[\int_a^b f^2(x)dx\right]^{1/2}$ the length of $f(x)$, and when (155.2) is satisfied, we may speak of a function of unit length in function space.

The theorem of Pythagoras which holds for ordinary vectors states that

$$(\boldsymbol{f}\cdot\boldsymbol{g})^2 \leq (\boldsymbol{f}\cdot\boldsymbol{f})(\boldsymbol{g}\cdot\boldsymbol{g})$$

The analogy to this theorem is to be found in Schwartz's inequality

$$(f\cdot g)^2 \leq (f\cdot f)(g\cdot g) \tag{155.3}$$

where
$$(f \cdot g) = \int_a^b f(x)g(x)dx \tag{155.4}$$

The concepts of scalar product and orthogonality, etc., are not restricted to real functions, but apply also to complex-valued functions provided the definition is modified so that the scalar product is formed from f and the complex conjugate of g, so that

$$(f \cdot g^*) = (f^* \cdot g) = \int_a^b f^*(x)g(x)dx \tag{155.5}$$

In this case if $(f \cdot g^*) = 0$, the function $g^*(x)$ is orthogonal to $f(x)$ instead of $g(x)$.

In practical applications it is convenient to deal with systems of orthonormal functions. In this case we have at hand a set of functions

$$\varphi_0(x), \varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x), \ldots \tag{155.6}$$

defined over a common interval, which satisfy the conditions

$$\int_a^b \varphi_n(x)\varphi_m(x)dx = \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases} \tag{155.7}$$

or in case of complex-valued functions,

$$\int_a^b \varphi_n^*(x)\varphi_m(x)dx = \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases} \tag{155.8}$$

One of the most useful orthonormal sets is composed of the trigonometric functions

$$\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos x, \frac{1}{\sqrt{\pi}}\sin x, \ldots, \frac{1}{\sqrt{\pi}}\cos nx, \frac{1}{\sqrt{\pi}}\sin nx, \ldots$$

defined over the interval $0 \leq x \leq 2\pi$. It is a simple matter to verify that

$$\begin{aligned} \int_0^{2\pi} \frac{1}{\sqrt{\pi}}\cos nx \frac{1}{\sqrt{\pi}}\cos mx\,dx &= \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases} \\ \int_0^{2\pi} \frac{1}{\sqrt{\pi}}\sin nx \frac{1}{\sqrt{\pi}}\cos mx\,dx &= 0 \text{ for all } m \text{ and } n \\ \int_0^{2\pi} \frac{1}{\sqrt{\pi}}\sin nx \frac{1}{\sqrt{\pi}}\sin mx\,dx &= \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases} \end{aligned} \tag{155.9}$$

The corresponding useful complex-valued orthonormal set is

$$\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{2\pi}}e^{\pm ix}, \ldots, \frac{1}{\sqrt{2\pi}}e^{\pm inx}, \ldots$$

These functions satisfy the orthogonality condition

$$\int_0^{2\pi} \frac{1}{\sqrt{2\pi}}e^{-inx}\cdot\frac{1}{\sqrt{2\pi}}e^{imx}\,dx = \begin{cases} 0, & n \neq m \\ 1, & n = m \end{cases} \tag{155.10}$$

156. Approximation of a function by a linear aggregate of orthonormal functions. One of the important practical uses of a system of orthonormal functions is their use in approximating an arbitrary function. Let the function $f(x)$ to be approximated and the orthonormal set $\varphi_0(x), \varphi_1(x) \ldots, \varphi_n(x)$ be defined over a common interval $a \leq x \leq b$. We seek to approximate $f(x)$ over the above interval by the linear combination

$$c_0\varphi_0(x) + c_1\varphi_1(x) + \ldots + c_n\varphi_n(x) = \sum_{\nu=0}^{n} c^\nu\varphi^\nu \tag{156.1}$$

of the members of the orthonormal set $\varphi_0, \varphi_1, \ldots, \varphi_n$ in such a way that the mean square error will be least.

The mean square error is

$$M = \int_a^b \left| f(x) - \sum_{\nu=0}^{n} c^\nu\varphi^\nu(x) \right|^2 dx \tag{156.2}$$

When the constants $c_0, c_1, \ldots, c_n$ are determined such that M is a minimum, $f(x)$ is said to be approximated in the mean by the linear aggregate $\sum_{\nu=0}^{n} c_\nu\varphi_\nu(x)$. The coefficients which make M a minimum can readily be shown to have the values

$$c_n = a_n = \int_a^b f(x)\varphi_n(x)dx$$

or if $\varphi_n(x)$ is complex (156.3)

$$c_n = a_n = \int_a^b f(x)\varphi_n^*(x)dx$$

The coefficients $a_0, a_1, \ldots, a_n$ are called the Fourier coefficients of $f(x)$ with respect to the orthonormal system $\varphi_0, \varphi_1, \ldots, \varphi_n$. That M will be a minimum when $c_\nu = a_\nu$ may be shown by performing the integration in the integral (156.2) where possible. Thus

$$\begin{aligned} M &= \int_a^b \left| f(x) - \sum_{\nu=0}^{n} c_\nu\varphi_\nu(x) \right|^2 dx \\ &= \int_a^b \left[f(x) - \sum_{\nu=0}^{n} c_\nu\varphi_\nu(x) \right] \left[f^*(x) - \sum_{\nu=0}^{n} c_\nu^*\varphi_\nu^*(x) \right] dx \\ &= \int_a^b |f(x)|^2 dx - \sum_{\nu=0}^{n} (c_\nu^* a_\nu + c_\nu a_\nu^*) + \sum_{\nu=0}^{n} |c_\nu|^2 \\ &= \int_a^b |f(x)|^2 dx + \sum_{\nu=0}^{n} |c_\nu - a_\nu|^2 - \sum_{\nu=0}^{n} |a_\nu|^2 \end{aligned} \tag{156.4}$$

The right side is readily seen to be least when $c_\nu = a_\nu$. When $c_\nu = a_\nu$, it follows that

$$\sum_{\nu=0}^{n} |a_\nu|^2 \leqq \int_a^b |f(x)|^2 dx \tag{156.5}$$

This is called Bessel's inequality. Since it holds for all n, it establishes the convergence of the series

$$\sum_{\nu=0}^{\infty} \left| a_\nu \right|^2$$

If it is possible to make the quantity

$$\int_a^b \left| f(x) - \sum_{\nu=0}^{n} a_\nu \varphi_\nu \right|^2 dx$$

less than an arbitrarily small positive number ϵ by a suitable choice of n for a function $f(x)$ for which $\int |f(x)|^2 dx$ exists, the orthonormal set $\varphi_1, \varphi_2, \ldots, \varphi_n, \ldots$ is said to form a complete set. Stated in a different way, the φ's constitute a complete orthonormal set if it is possible to approximate a quadratically integrable function $f(x)$ in the mean as exactly as desired by retaining a sufficient number of terms in the aggregate $\sum_{\nu=0}^{n} a_\nu \varphi_\nu$. For a complete set, it therefore follows that

$$\lim_{n\to\infty} \int_a^b \left| f(x) - \sum_{\nu=0}^{n} a_\nu \varphi_\nu \right|^2 dx = 0 \tag{156.6}$$

and from eq. (156.4) this yields

$$\int_a^b |f(x)|^2 dx = \sum_{\nu=0}^{\infty} \left| a_\nu \right|^2 \tag{156.7}$$

The existence of this equality also establishes the completeness of the orthonormal set $\varphi_1, \varphi_2, \ldots, \varphi_n, \ldots$ when

$$a_\nu = \int_a^b f(x)\varphi_\nu^* \, dx$$

A most important result of this analysis is that, given a complete orthonormal set $\varphi_1, \varphi_2, \ldots, \varphi_n$, defined over an interval ab, any function for which $\int_a^b \left| f(x) \right|^2 dx$ exists can be approximated in the mean arbitrarily closely by the sum $\sum_{\nu=0}^{n} a_\nu \varphi_\nu$, where $a_\nu = \int_a^b f(x)\varphi_\nu^* \, dx$. The series $\sum_{\nu=0}^{\infty} a_\nu \varphi_\nu$ is said to converge in the mean to $f(x)$.

It should be carefully observed that although $\sum_{\nu=0}^{\infty} a_\nu \varphi_\nu$ converges to $f(x)$ in the mean, it by no means follows that

$$f(x) = \sum_{\nu=0}^{\infty} a_\nu \varphi_\nu(x)$$

for this would require that it be possible to find a value of n such that

$$\left| f(x) - \sum_{\nu=0}^{n} a_\nu \varphi_\nu(x) \right| < \epsilon$$

where ϵ is an arbitrarily small positive number. This is very different from the requirement

$$\int_a^b \left| f(x) - \sum_{\nu=0}^{n} a_\nu \varphi_\nu(x) \right|^2 dx < \epsilon$$

The latter actually would allow $\sum_{\nu=0}^{n} a_\nu \varphi_\nu(x)$ to differ widely from $f(x)$ at various points.

The least restrictive conditions which must be placed on a function $f(x)$ in order that $\sum_{\nu=0}^{\infty} a_\nu \varphi_\nu(x)$ will converge uniformly to $f(x)$ have not been determined. For a given complete orthonormal set only sufficient conditions can be given. This question has received the most attention for the orthonormal set composed of the trigonometric functions. The matter will be referred to again in connection with particular orthonormal sets. It would seem that in many physical applications only convergence in the mean is required.

It may be mentioned that a piecewise continuous function $f(x)$ is uniquely represented by the coefficients $a_\nu = \int_a^b f(x)\varphi_\nu^*(x)dx$ computed with respect to a complete orthonormal set $\varphi_1, \varphi_2, \ldots, \varphi_n, \ldots$ in the sense that two piecewise continuous functions are identical if they have the same coefficients a_ν with respect to one and the same orthonormal set.

157. Orthogonalization of linearly independent functions. The functions $\varphi_1, \varphi_2, \ldots, \varphi_n, \ldots$ of an orthonormal set have the property of being linearly independent, which means there exists no relation of the form

$$c_1\varphi_1 + c_2\varphi_2 + \ldots + c_n\varphi_n = 0$$

except that for which $c_1 = c_2 = \ldots = c_n = 0$. This follows by direct computation of the constants c_ν. Any constant c_ν can be found by multiplying the above expression by φ_ν^* and integrating over the domain of definition of the orthonormal set. Thus

$$c_\nu \int_a^b \varphi_\nu^* \varphi_\nu \, dx = 0$$

and since

$$\int_a^b \varphi_\nu^* \varphi_\nu \, dx = 1 \qquad c_\nu = 0$$

Given a set of linearly independent functions $f_1(x), f_2(x), \ldots, f_n(x), \ldots$, for which $\int_a^b f_\nu(x)^2 dx$ exists, it would be possible to approximate

a function $f(x)$ in the mean by a linear aggregate $\sum_{\nu=1}^{n} c_\nu f_\nu(x)$. However, the constants $c_1, c_2, \ldots$ cannot be found as directly and as conveniently as for the case of an orthonormal set. An example of how this may be done in the case of the linearly independent functions $1, x, x^2, \ldots, x^n, \ldots$ will be given in the next section.

One way to proceed is to replace the linearly independent functions $f_1(x), f_2(x), \ldots, f_n(x), \ldots$ by the same number of new functions $\varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x), \ldots$, each of which is a linear combination of the $f_1(x), f_2(x), \ldots, f_n(x), \ldots$ such that the new functions are orthonormal. The problem is exactly the same as that already discussed in Sec. 135 where n linearly independent vectors were replaced by an orthogonal system of unit vectors. The procedure for carrying out the orthogonalization is the same as given for vectors in Sec. 135, where the scalar product of two vectors is replaced by the corresponding operation $\int_a^b f(x)g(x)dx$ in function space. Thus one starts by writing

$$\varphi_1(x) = \frac{f_1(x)}{\left(\int_a^b \left| f(x) \right|^2 dx\right)^{1/2}} \tag{157.1}$$

then writing $\varphi_2(x) = c_1\varphi_1(x) + c_2 f_2(x)$. It is required that

$$\begin{aligned} \int_a^b \varphi_1(x)\varphi_2(x) &= 0 \\ &= \int_a^b [c_1\varphi_1^2(x) + c_2 f_2(x)\varphi_1(x)]dx \\ &= c_1 + c_2 \int_a^b f_2(x)\varphi_1(x)dx \end{aligned} \tag{157.2}$$

In order to keep the functions normalized we write

$$\int_a^b [c_1\varphi_1(x) + c_2 f_2(x)]^2 dx = 1$$

or

$$c_1^2 + 2c_1c_2\int_a^b \varphi_1(x)f_2(x)dx + c_2^2\int_a^b f_2^2(x)dx = 1$$

Making use of eq. (157.2), this becomes

$$c_2^2\left[\int_a^b \varphi_1(x)f_2(x)dx\right]^2 - 2c_2^2\left[\int_a^b \varphi_1(x)f_2(x)dx\right]^2 + c_2^2\int_a^b f_2^2(x)dx = 1$$

or

$$c_2 = \frac{1}{\left\{\int_a^b f_2^2(x)dx - \left[\int_a^b \varphi_1(x)f_2(x)dx\right]^2\right\}^{1/2}}$$

and

$$c_1 = \frac{-\int_a^b f_2(x)\varphi_1(x)dx}{\left\{\int_a^b f_2^2(x)dx - \left[\int_a^b \varphi_1(x)f_2(x)dx\right]^2\right\}^{1/2}}$$

This process can be carried on until all the φ's are determined.

Several methods can be used to establish the linear independence of the functions to be orthogonalized. One method has already been given in Sec. 18. Another useful method makes use of the scalar product in function space defined by (155.4). The criterion for the linear independence of a set of functions $f_1(x), f_2(x), \ldots, f_n(x)$ is that the Gramian determinant

$$\Gamma(f_1,f_2,\ldots,f_n) = \begin{vmatrix} (f_1 \cdot f_1) & (f_1 \cdot f_2) & \ldots & (f_1 \cdot f_n) \\ (f_2 \cdot f_1) & (f_2 \cdot f_2) & \ldots & (f_2 \cdot f_n) \\ \ldots\ldots & \ldots\ldots & \ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots & \ldots\ldots \\ (f_n \cdot f_1) & (f_n \cdot f_2) & \ldots & (f_n \cdot f_n) \end{vmatrix} \tag{157.3}$$

should not vanish where $(f_p \cdot f_q)$ is defined by (155.4) and is to be evaluated over the common domain of definition of the functions under consideration.

As an example of the use of the Gramian determinant, let us examine the linear independence of the functions 1, x, x^2, x^3 throughout the interval $0 \leq x \leq 1$. The determinant (157.3) is readily found to be

$$\begin{vmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{vmatrix}$$

This determinant is not zero; hence the functions 1, x, x^2, x^3 are linearly independent. In fact, the set of functions 1, x, x^2, $\ldots$, x^n, $\ldots$ defined over any finite interval $a \leq x \leq b$ are linearly independent.

158. Examples of orthonormal sets. Aside from the most common orthonormal set of functions mentioned in Sec. 155, namely, the set of trigonometric functions, a variety of others may be obtained by orthogonalizing the linearly independent set of functions 1, x, x^2, $\ldots$, x^n, $\ldots$ over various domains of definition. As an example, let us orthogonalize this set on the interval $-1 \leq x \leq 1$. From eq. (157.1), the first of the orthonormal set of functions is found by writing

$$P_0 = \frac{1}{\left(\int_{-1}^{+1} dx\right)^{1/2}} = \frac{1}{\sqrt{2}}$$

Then the next function orthonormal to P_0 will be found by the method outlined in the previous section. First

$$P_1 = c_0 \frac{1}{\sqrt{2}} + c_1 x$$

The c's are now determined so as to make P_1 and P_0 orthonormal. Hence

$$\int_{-1}^{+1}\left(c_0\frac{1}{\sqrt{2}}+c_1x\right)\cdot 1\,dx = 0 = \sqrt{2}c_0$$

and

$$\int_{-1}^{+1} c_1^2x^2\,dx = 1 = c_1^2\frac{2}{3}$$

Consequently $c_0 = 0$ and $c_1 = \sqrt{\frac{3}{2}}$ and $P_1 = \sqrt{\frac{3}{2}}\,x$. In order to obtain P_2 we write

$$P_2 = c_0P_0 + c_1P_1 + c_2x^2$$

and again require that

$$\int_{-1}^{+1}\left(c_0\frac{1}{\sqrt{2}}+c_1\sqrt{\frac{3}{2}}x+c_2x^2\right)dx = 0$$

$$\int_{-1}^{+1}\left(c_0\frac{1}{\sqrt{2}}+c_1\sqrt{\frac{3}{2}}x+c_2x^2\right)\sqrt{\frac{3}{2}}x\,dx = 0$$

$$\int_{-1}^{+1}\left(c_0\frac{1}{\sqrt{2}}+c_1\sqrt{\frac{3}{2}}x+c_2x^2\right)^2dx = 1$$

From these relations we find that

$$c_0 = -\sqrt{\frac{5}{2}},\quad c_1 = 0,\quad c_2 = \frac{3}{2}\sqrt{\frac{5}{2}}$$

$$P_2 = \sqrt{\frac{5}{2}}\left(\frac{3}{2}x^2-1\right)$$

Proceeding thus, a set of polynomials

$$P_0 = \frac{1}{\sqrt{2}},\quad P_1 = \sqrt{\frac{3}{2}}x,\quad P_2 = \sqrt{\frac{5}{2}}\left(\frac{3}{2}x^2-\frac{1}{2}\right),\quad \cdots$$

would be obtained, each of which differ from the well-known Legendre polynomials by a numerical factor. The Legendre polynomials are $1,\ x,\ (\frac{3}{2}x^2-\frac{1}{2}),\ \ldots$. The difference comes only in the normalization. The Legendre polynomials are normalized, so that $P_n(1) = 1$ instead of $\int_{-1}^{+1}P_n^2(x)dx = 1$. Consequently we may regard the Legendre polynomials as an orthogonal set of functions obtained by orthogonalizing the functions $1,\ x,\ x^2,\ \ldots,\ x^n,\ \ldots$ with respect to the interval $-1 \le x \le 1$. It can also be shown that they form a complete set.

A variety of complete sets of orthonormal functions defined over various intervals can be obtained with the help of an auxiliary function $\rho(x)$. To illustrate the procedure, we shall consider subsequent functions to be defined over the interval $a < x < b$ (b may be infinite) and require that

$$\int_a^b \rho(x)x^n dx$$

exist for all positive integer n's. The functions of the system,

$$\sqrt{\rho(x)},\ \sqrt{\rho(x)}x,\ \sqrt{\rho(x)}x^2,\ \ldots,\ \sqrt{\rho(x)}x^n,\ \ldots$$

are linearly independent. Using this function system as a basis, the corresponding orthonormal set may be determined on the interval $a < x < b$ as illustrated above. As a result of this process we will obtain the orthonormal system $\varphi_0, \varphi_1, \varphi_2, \ldots, \varphi_n, \ldots$

where $$\varphi_n(x) = \sqrt{\rho(x)}\ \pi_n(x)$$

and $\pi_n(x)$ is a polynomial of degree n. Here $\pi_n(x)$ is, except for sign, uniquely determined by means of the function $\rho(x)$ and the interval of definition.

There are a number of important special cases, which lead to much used orthonormal systems. These are listed according to the auxiliary function $\rho(x)$ and the interval.

$\rho(x)$	Interval	Name of Polynomial
1	$-1 \leq x \leq 1$	$\pi_n(x) = \sqrt{\frac{2n+1}{2}} \cdot P_n(x)$. ($P_n$ = Legendre polynomial.)
$x^\alpha(1-x)^\beta$ $\alpha, \beta > -1$	$0 \leq x \leq 1$	$\pi_n(x)$ = Jacoby or hypergeometric polynomial (except for constant factor).
e^{-x}	$0 \leq x < \infty$	$\pi_n(x) = L_n(x)$, the Laguerre polynomial.
e^{-x^2}	$-\infty < x < \infty$	$\pi_n(x) = H_n(x)$, the Hermitian polynomial.

The properties of some of these orthonormal systems as well as others, such as Bessel's functions, which are not formed in the above manner will be discussed in later sections of this chapter.

159. Fourier series. Since the complex exponential functions $1, e^{\pm ix}, \ldots, e^{\pm inx}, \ldots$ form a complete orthogonal set we know (see Sec. 156) that if $f(x)$ is defined in the interval $-\pi \leq x \leq \pi$ and $\int_{-\pi}^{\pi} |f(x)|^2 dx$ exists, $f(x)$ can be approximated in the mean as exactly as we wish by the sum

$$S_n(x) = \sum_{\nu=-n}^{+n} a_\nu e^{i\nu x} \tag{159.1}$$

if n is sufficiently large. In accordance with (156.3), the coefficients are given by

$$a_\nu = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-i\nu x} dx \tag{159.2}$$

In addition to this, it is often of interest to know under what conditions the function $f(x)$ can be uniformly approximated by $S_n(x)$, i.e., that

$$\lim_{n\to\infty} |f(x) - S_n(x)| = 0$$

The determination of the conditions which must be satisfied by $f(x)$ in order that $f(x)$ can be uniformly approximated by $S_n(x)$ constitutes the mathematical theory of Fourier series.

In the application of Fourier series, the following sufficient conditions on $f(x)$, which secure uniform convergence, have been found useful.

1. Any function $f(x)$ which is continuous, together with its first derivative in the interval $-\pi \leq x \leq \pi$ with period 2π, can be uniformly approximated by $S_n(x)$.

2. Let $f(x)$ be defined arbitrarily in the interval $-\pi \leq x \leq \pi$ and defined for all other real values of x by the equation

$$f(x+2\pi) = f(x)$$

so that $f(x)$ has a period 2π. Also let $f(x)$ be such that $\int_{-\pi}^{\pi} f(x)dx$ exists, and if this be an improper integral, let it be absolutely convergent. Then if t be any interior point of any interval $a < t < b$ in which $f(t)$ is of bounded variation (see Sec. 53), the series $\lim_{n\to\infty} S_n(t)$ converges, and the sum is

$$\tfrac{1}{2}[f(t+0) + f(t-0)]$$

where

$$f(t+0) = \lim_{h\to 0} f(t+h) \quad \text{and} \quad f(t-0) = \lim_{h\to 0} f(t-h)$$

If $f(x)$ is continuous at $x = t$, the sum is $f(t)$.

This theorem allows $f(x)$ to be discontinuous at one or more points in the interval. If such occurs at a point $x = t$, then $\lim_{n\to\infty} S_n(x)$ at $x = t$ yields the arithmetic mean of the values of the function at the jump point, or $\frac{1}{2}[f(t-0) + f(t+0)]$. Also a function defined over the interval $0 \leq x < \pi$ can, for the purposes of expansion in Fourier series, be continued in a suitable way in the interval $-\pi < x \leq 0$, and to secure periodicity it may be defined for values of x outside this interval by the relation $f(x+2\pi n) = f(x)$, where n is a positive or negative integer. This procedure may lead to discontinuities in the resulting function at $x = \pi$ or $x = -\pi$, and consequently at $x = \pm\pi n$. In accordance with conditions (2) above, this means only that the Fourier series will converge to the arithmetic mean of the function at these jump points.

The above situation is encountered if we wish to represent the function $f(x) = x$ over the interval $0 \leq x < \pi$. In this case the actual function we would represent is shown in Fig. 159.1. Using eq. (159.2), the Fourier coefficients for this function are given by

$$a_\nu = \frac{1}{2\pi}\int_{-\pi}^{\pi} xe^{-i\nu x}dx$$

$$a_0 = 0$$

$$a_\nu = \frac{(-1)^\nu}{-i\nu} \qquad \nu \neq 0$$

Consequently the Fourier representation of $f(x) = x$ in the interval $-\pi < x < \pi$ is

$$x = \sum_{\nu=-\infty}^{\infty}{}' \frac{(-1)^\nu}{-i\nu} e^{-i\nu x} = 2\sum_{\nu=1}^{\infty} (-1)^\nu \frac{\sin \nu x}{\nu}$$

The stroke to the right of the summation sign signifies that the term $\nu = 0$ is omitted. When $x = \pi$, the value of the Fourier series is zero. This furnishes an instance of the fact that the series yields the arithmetic mean of the function at the point where the function undergoes a jump.

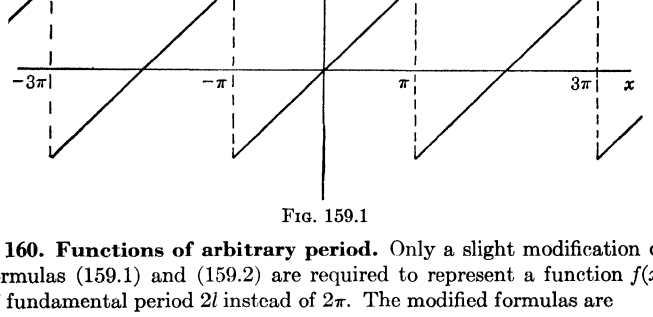


FIG. 159.1

160. Functions of arbitrary period. Only a slight modification of formulas (159.1) and (159.2) are required to represent a function $f(x)$ of fundamental period $2l$ instead of 2π. The modified formulas are

$$S_n(x) = \sum_{\nu=-n}^{n} a_\nu e^{\pi i\nu x/l} \tag{160.1}$$

$$a_\nu = \frac{1}{2l}\int_{-l}^{+l} f(x)e^{-\pi i\nu x/l}dx \tag{160.2}$$

The conditions on $f(x)$ in order that $f(x)$ can be approximated in the mean, or that the Fourier series converge uniformly to $f(x)$, are the same as given in Sec. 159 except that the interval of definition must be $2l$ instead of 2π.

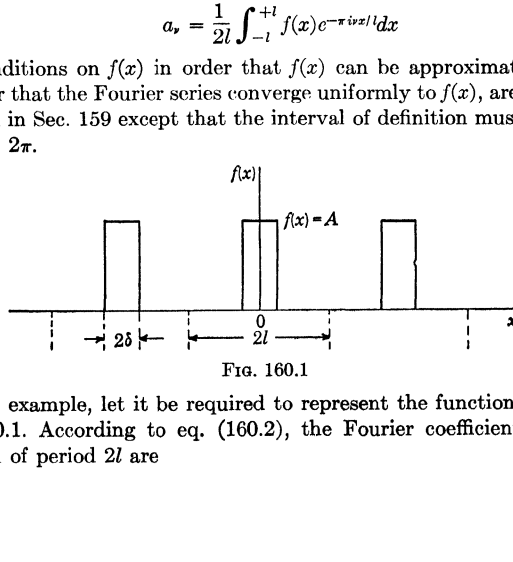


FIG. 160.1

As an example, let it be required to represent the function shown in Fig. 160.1. According to eq. (160.2), the Fourier coefficients for the function of period $2l$ are

$$a_\nu = \frac{1}{2l}\int_{-l}^{+l} f(x)e^{-i\pi\nu x/l}dx$$

$$= \frac{1}{2l}\int_{-\delta}^{+\delta} Ae^{-i\pi\nu x/l}dx = \frac{A}{\pi}\frac{e^{i\pi\nu\delta/l} - e^{-i\pi\nu\delta/l}}{2i\nu}$$

$$= \frac{A}{\pi\nu}\sin\frac{\pi\nu\delta}{l} \tag{160.3}$$

Consequently

$$f(x) = \frac{A}{\pi}\sum_{\nu=-\infty}^{\infty}\frac{\sin(\pi\nu\delta/l)e^{i\pi\nu x/l}}{\nu} \tag{160.4}$$

$$= \frac{2A}{\pi}\sum_{\nu=0}^{\infty}\frac{[\sin(\pi\nu\delta/l)][\cos(\pi\nu\delta/l)]}{\nu} \tag{160.5}$$

This result will be discussed in further detail in the next section in connection with the number of harmonic constituents required to obtain a given degree of approximation of the function $f(x)$.

161. Properties of functions and their Fourier coefficients. In practice it is of great value to have some idea of the way in which general properties of the function to be represented affect the Fourier coefficients. For instance, it is desirable to know what property of a function leads to rapidly decreasing Fourier coefficients, or given a function different from zero over a certain interval, how the number of coefficients required for a certain degree of approximation is related to the magnitude of this interval.

It can readily be seen that the smoothness of variation of a function has a definite effect on the rate at which the Fourier coefficients approach zero. A measure of the smoothness is the number of continuous derivatives the function possesses. Let the function $f(x)$ be defined over the interval $2l$ and have $r - 1$ continuous derivatives and a piecewise continuous derivative of order r in the interval. Then

$$f(x) = \sum_{\nu=-\infty}^{+\infty} a_\nu e^{i\nu\pi x/l} \tag{161.1}$$

$$a_\nu = \frac{1}{2l}\int_{-l}^{+l} f(x)e^{-i\nu\pi x/l}dx \tag{161.2}$$

If one performs r partial integrations,

$$a_\nu = \frac{1}{2l}\left(\frac{l}{\pi}\right)^r\frac{1}{(i\nu)^r}\int_{-l}^{+l} f^{(r)}(x)e^{-i\nu\pi x/l}dx$$

or since

$$|a_\nu| = \frac{1}{2l}\left(\frac{l}{\pi}\right)^r\frac{1}{\nu^r}\left|\int_{-l}^{+l} f^{(r)}(x)e^{-i\nu\pi x/l}dx\right|$$

$$\leqq \frac{1}{2l}\left(\frac{l}{\pi}\right)^r\frac{1}{\nu^r}\int_{-l}^{+l}|f^{(r)}(x)dx| = \frac{K}{\nu^r} \tag{161.3}$$

where K is a constant independent of ν. Thus we see that the smoother a function is, or the greater the number of continuous derivatives it has, the more rapidly the coefficients approach zero.

We may use the Fourier series for the function shown in Fig. 160.1 to infer a relation of a general character between the fraction of the period $2l$ over which the function is different from zero and the number of terms in the series required to obtain a certain degree of approximation to $f(x)$ in the mean. The fraction of the period for which $f(x)$ is different from zero is δ/l. When $\delta/l \ll 1$, the values of $|a_\nu|^2$ found from (160.3) are plotted against ν in Fig. 161.1. It will be convenient to consider the range of ν over which $|a_\nu|^2$ is appreciably different from

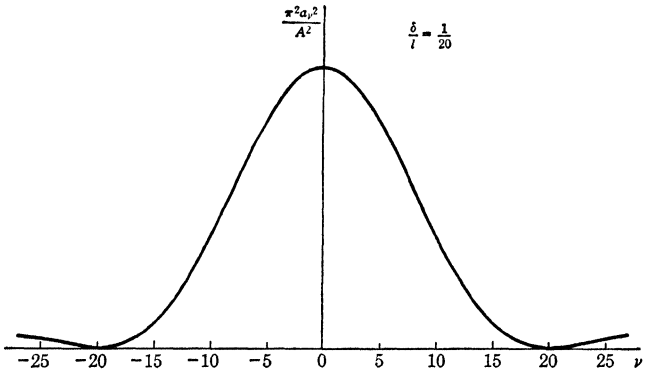


FIG. 161.1

zero to be $-l/\delta < \nu < l/\delta$, i.e., the values of ν lying between the zeros of $[\sin^2(\pi\nu\delta/l)]/\nu^2$ closest to the origin $\nu = 0$. Thus the number of terms n of the series for which $|a_\nu|^2$ is appreciable is $|n| \leq l/\delta$, which is the reciprocal of the fraction of the period over which $f(x)$ is different from zero. Thus the smaller the interval over which $f(x)$ is appreciably different from zero, the greater the value of n for which $|a_\nu|^2$ is appreciably different from zero.

The degree to which $n = l/\delta$ terms of the series approximates the function $f(x)$ in the mean can be found by applying eq. (156.4), which in this case leads to

$$M = \int_{-l}^{+l} \left| f(x) - \sum_{\nu=-n}^{+n} a_\nu e^{-i\pi\nu x/l} \right|^2 dx$$

$$= \int_{-l}^{+l} |f(x)|^2 dx - \sum_{\nu=-n}^{+n} |a_\nu|^2$$

$$= \frac{4A^2}{\pi^2}\left[\sum_{\nu=0}^{\infty}\frac{\sin^2(\pi\nu\delta/l)}{\nu^2} - \sum_{\nu=0}^{n}\frac{\sin^2(\pi\nu\delta/l)}{\nu^2}\right]$$

$$= \frac{4A^2}{\pi^2}\sum_{\nu=n+1}^{\infty}\frac{\sin^2(\pi\nu\delta/l)}{\nu^2}$$

$$\leq \frac{4A^2}{\pi^2}\sum_{\nu=n+1}^{\infty}\frac{1}{\nu^2} \approx \frac{4A^2}{\pi^2}\frac{1}{(n+1)}$$

$$\leq \frac{4A^2}{\pi^2}\frac{\delta}{l}$$

Consequently the mean square deviation from $f(x)$ is inversely proportional to the number of terms over which $|a_\nu|^2$ is appreciably different from zero.

162. Mean square spectral extent of a function. The relation between the "width" of the impulse function and the number of Fourier coefficients for which $|a_\nu|^2$ is appreciably different from zero as exhibited in the preceding section can be established in a more general way for a certain class of functions.

Consider a function $f(x)$ of period $2l$ such that

$$\int_{-l}^{+l}|f(x)|^2dx \quad \text{and} \quad \int_{-l}^{+l}|f'(x)|^2dx$$

exist. We may define the mean square width $\overline{\Delta x^2}$ of the function $f(x)$ as the integral

$$\overline{\Delta x^2} = \int_{-l}^{+l}(x - x_0)^2|f(x)|^2dx \tag{162.1}$$

where x_0 is some reference point, such as the point where $|f(x)|$ is a maximum, or defined such that

$$\int_{-l}^{+l}(x - x_0)|f(x)|dx = 0$$

Let $f(x)$ be represented by the Fourier series

$$f(x) = \sum_{\nu=-\infty}^{+\infty}a_\nu e^{i\nu\pi x/l} \tag{162.2}$$

We may define the mean spectral extent $\overline{\Delta\nu^2}$ of $f(x)$ by the series

$$\overline{\Delta\nu^2} = \sum_{\nu=-\infty}^{+\infty}\nu^2|a_\nu|^2 \tag{162.3}$$

provided this series converges.

To establish a limiting relationship between $\overline{\Delta x^2}$ and $\overline{\Delta\nu^2}$ we first note that

$$\int_{-l}^{+l} |f'|^2 dx = \frac{\pi^2}{l^2} \sum_{\nu=-\infty}^{+\infty} \nu^2 |a_\nu|^2 = \frac{\pi^2}{l^2} \overline{\Delta\nu} \tag{162.4}$$

and then make use of the inequality*

$$\left| \int_{-l}^{+l} g_1 h_1 dx + \int_{-l}^{+l} g_2 h_2 dx \right|^2$$
$$\leqq \left(\int_{-l}^{+l} g_1 g_1^* dx + \int_{-l}^{+l} g_2 g_2^* dx \right) \left(\int_{-l}^{+l} h_1 h_1^* dx + \int_{-l}^{+l} h_2 h_2^* dx \right)$$

in which we let $g_1 = (x - x_0)f$, $h_1 = f^{*\prime}$; and $g_2 = (x - x_0)f^*$, $h_2 = f'$. The inequality then becomes

$$\left| \int_{-l}^{+l} (x - x_0)(ff^{*\prime} + f^* f') dx \right|^2 \leq 4 \int_{-l}^{+l} (x - x_0)^2 |f|^2 dx \int_{-l}^{+l} |f'|^2 dx$$

or

$$\left| \int_{-l}^{+l} (x - x_0) \frac{d}{dx} |f|^2 dx \right|^2 \leq 4 \int_{-l}^{+l} (x - x_0)^2 |f|^2 dx \int_{-l}^{+l} |f'|^2 dx$$

By integrating the integral on the left side by parts, and replacing the integrals on the right side by the quantities from eqs. (162.1) and (162.4), we have

$$\left| (l - x_0)|f(l)|^2 + (l + x_0)|f(-l)|^2 - \int_{-l}^{+l} |f|^2 dx \right|^2$$
$$\leq 4 \overline{\Delta x^2}\, \frac{\pi^2}{l^2}\, \overline{\Delta\nu^2} \tag{162.5}$$

An important case is that for which $f(l) = f(-l) = 0$, so that

$$\overline{\Delta x^2}\, \overline{\Delta\nu^2} \geq \frac{l^2}{4\pi^2} \left| \int_{-l}^{+l} |f|^2 dx \right|^2 \tag{162.6}$$

This result shows that if the area under the curve of $|f|^2$ is held constant, the smaller the mean square width of $|f|^2$ the larger the mean square spectral extent of $f(x)$ must be.

163. Harmonic analysis of curves. In engineering practice it is often necessary to make a harmonic analysis of data. The function so far as it is known in this case takes the form of a curve plotted from the data. Thus the function is not known in explicit analytical form. The approximate representation of such a curve by a Fourier series can be facilitated by making use of the following procedure.

Let it be desired to make a harmonic analysis of the function $f(x)$

* This inequality can be established from the positive definite character of the quadratic form in λ

$$\int [(\lambda|g_1| + |h_1|)^2 + (\lambda|g_2| + |h_2|)^2] dx \geq 0$$

or

$$\lambda^2 \int (|g_1|^2 + |g_2|^2) dx + 2\lambda \int (|g_1||h_1| + |g_2||h_2|) dx$$
$$+ \int (|h_1|^2 + |h_2|^2) dx \geq 0$$

TABLE 163–1

I	II	III	IV	First	Second	Third	Fourth	
	y_{24}	s_0		d_1+d_{11}				0.0216
y_1	y_{23}	s_1	d_1	d_2+d_{10}	$d_1+d_5-d_7-d_{11}$			0.0417
y_2	y_{22}	s_2	d_2	d_3+d_9		$d_1+d_3+d_9+d_{11}-d_5-d_7$		0.0589
y_3	y_{21}	s_3	d_3	d_4+d_8	$d_2+d_4-d_8-d_{10}$		$d_1+d_2+d_4+d_8-d_4-d_5-d_{10}-d_{11}$	0.0722
y_4	y_{20}	s_4	d_4	d_5+d_7				0.0805
y_5	y_{19}	s_5	d_5	d_6	d_3-d_9	$d_2+d_{10}-d_6$		0.0833
y_6	y_{18}	s_6	d_6	$b_1 =$	$b_2 =$	$b_3 =$	$b_4 =$	
y_7	y_{17}	s_7	d_7	s_5-s_7				0.0216
y_8	y_{16}	s_8	d_8	s_4-s_8	$s_2+s_{10}-s_4-s_8$		$s_1+s_5+s_7+s_{11}-s_2-s_4-s_8-s_{10}$	0.0417
y_9	y_{15}	s_9	d_9	s_3-s_9		$s_1+s_7+s_9-s_3-s_5-s_{11}$		0.0589
y_{10}	y_{14}	s_{10}	d_{10}	s_2-s_{10}	$s_1+s_{11}-s_5-s_7$			0.0722
y_{11}	y_{13}	s_{11}	d_{11}	s_1-s_{11}				0.0805
y_{12}		s_{12}		s_0-s_{12}	$s_0+s_{12}-s_6$	$s_0+s_8-s_4-s_{12}$	$s_0+s_6+s_{12}-s_3-s_9$	0.0833
				$a_1 =$	$a_2 =$	$a_3 =$	$a_4 =$	

$$\frac{a_0}{2} = \frac{\sum_{\nu=0}^{12} s_\nu}{24}$$

$$f(x) = \frac{a_0}{2} + \sum_{\nu=1} (a_\nu \cos \nu x + b_\nu \sin \nu x)$$

shown in the curve Fig. 163.1. For convenience we shall suppose the region of definition is $0 \leq x \leq 2\pi$. If the function does not vary too rapidly, it may be sufficient to divide the interval into 24 equal parts, numbering the division points 1, 2, ... , 24 as shown. The values of $f(x)$ at these division points are denoted by $y_1, y_2, \ldots, y_{24}$ and tabulated as shown in columns I and II in Table 163-1. The sums $s_0 = y_{24}$'

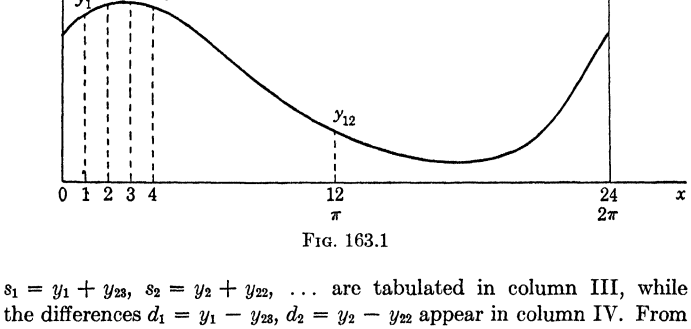


FIG. 163.1

$s_1 = y_1 + y_{23}$, $s_2 = y_2 + y_{22}$, ... are tabulated in column III, while the differences $d_1 = y_1 - y_{23}$, $d_2 = y_2 - y_{22}$ appear in column IV. From the entries in the table, the Fourier coefficients in the series of real functions

$$f(x) = \frac{a_0}{2} + \sum_{\nu=1}^{\infty} (a_\nu \cos \nu x + b_\nu \sin \nu x)$$

can be computed conveniently. First

$$\frac{a_0}{2} = \frac{1}{2\pi} \int_0^{2\pi} f(x)dx \approx \frac{2\pi}{2\pi 24} (y_1 + y_2 + \ldots + y_{24})$$

$$\approx \frac{s_0 + s_1 + \ldots + s_{12}}{24}$$

Therefore $a_{0/2}$ is found by adding the entries in column III and dividing the sum by 24. From the formula

$$a_1 = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos x \, dx$$

the coefficient a_1 can be computed approximately in terms of $f(x)$ and $\cos x$ at the points of subdivision. Thus

$$a_1 \approx \frac{1}{\pi} \left(y_1 \cos \frac{\pi}{12} + y_2 \cos \frac{2\pi}{12} + \ldots + y_{24} \cos 2\pi \right) \frac{2\pi}{24}$$

$$\approx \frac{1}{12} \left[(y_1 - y_{11} - y_{13} + y_{23}) \cos \frac{\pi}{12} + (y_2 - y_{10} - y_{14} + y_{22}) \cdot \right.$$

$$\cos\frac{2\pi}{12} + \ldots + y_{24}\cos 2\pi\Big]$$

$$\approx \frac{1}{12}\Big[(s_1 - s_{11})\cos\frac{\pi}{12} + (s_2 - s_{10})\cdot$$

$$\cos\frac{2\pi}{12} + (s_3 - s_9)\cos\frac{3\pi}{12}$$

$$+ (s_4 - s_8)\cos\frac{4\pi}{12} + (s_5 - s_7)\cos\frac{5\pi}{12} + s_0\Big]$$

The entries in the column labeled First in the lower part of the table are the numbers $(s_5 - s_7)$, etc. as indicated, multiplied by the numbers 0.0216 etc. in the column at the extreme right, which are just the values of $\frac{1}{12}\cos\frac{5}{12}\pi$, etc. appearing in the last equation. The value of the coefficient a_1 is then found by adding the numbers of the First column in the lower part of the table. The procedure for computing a_2, a_3, etc., and b_1, b_2, etc. is sufficiently clear from an inspection of the table.

The accuracy of this procedure becomes poor for more than 4 or 5 coefficients because the number of subdivisions in the interval is not great enough to obtain a very good approximate value of the pertinent integrals.

164. Hermitian polynomials and orthogonal Hermitian functions. In Sec. 158, it was mentioned that an orthonormal set of functions built on the interval $-\infty < x < \infty$ by means of an orthogonalization procedure based on the auxiliary function e^{-x^2} led to the Hermitian polynomials except for a constant factor. The Hermitian polynomials $H_n(x)$ may also be obtained with the help of a generating function $g(x,t)$, from which several very useful relations can be derived.

A generating function $g(x,t)$ is one from which the whole set of functions $H_n(x)$ in question appear as the coefficients of successive powers of t when $g(x,t)$ is expanded as a power series in t. That is,

$$g(x,t) = \sum_{n=0}^{\infty} \frac{H_n(x)}{n!}\, t^n \tag{164.1}$$

For the Hermitian polynomials

$$g(x,t) = e^{-t^2+2tx} = e^{x^2}e^{-(t-x)^2} = \sum_{n=0}^{\infty} \frac{H_n(x)}{n!}\, t^n \tag{164.2}$$

It is evident that $H_n(x)$ may be obtained by differentiating the above equation n times with respect to t, and thereafter setting t equal to zero. Thus

$$H_n(x) = \frac{\partial^n}{\partial t^n}\, g(x,t)\Big]_{t=0} \tag{164.3}$$

Since $$\frac{\partial}{\partial t} e^{-(t-x)^2} = -\frac{\partial}{\partial x} e^{-(t-x)^2}$$

we have

$$H_n(x) = \frac{\partial^n}{\partial t^n} e^{x^2} e^{-(t-x)^2}\Big]_{t=0} = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \tag{164.4}$$

which allows one to determine $H_n(x)$ by merely differentiating n times the function e^{-x^2}.

Differentiating the equality

$$g(x,t) = e^{-t^2+2tx}$$

with respect to x gives

$$\frac{\partial g(x,t)}{\partial x} = 2tg(x,t)$$

or $$\sum_{n=0}^{\infty} \frac{H_n'(x)}{n!} t^n = 2t \sum_{n=0}^{\infty} \frac{H_n(x)}{n!} t^n$$

Consequently $$H_n'(x) = 2nH_{n-1}(x) \tag{164.5}$$

This provides a means of expressing derivatives of $H_n(x)$ in terms of polynomials of lower degree.

A recurrence formula for the H's can be obtained by differentiating the equality

$$g(x,t) = e^{x^2} e^{-(t-x)^2}$$

with respect to t. Then

$$\frac{\partial g}{\partial t} = -2(t - x)g(x,t)$$

or $$\sum_{n=1}^{\infty} \frac{H_n}{(n-1)!} t^{n-1} + 2(t - x) \sum_{n=0}^{\infty} \frac{H_n(x)t^n}{n!} = 0$$

Consequently

$$\sum_{n=0}^{\infty} \left[\frac{H_{n+1}}{n!} - 2x \frac{H_n(x)}{n!} + \frac{2H_{n-1}(x)}{(n-1)!}\right] t^n = 0$$

or $$H_{n+1}(x) - 2xH_n(x) + 2nH_{n-1}(x) = 0 \tag{164.6}$$

Once two Hermitian polynomials are known, all others can be computed in terms of them by means of the recurrence formula.

The differential equation satisfied by $H_n(x)$ can readily be obtained by using the relation (164.5) in eq. (164.6) in such a way that only polynomials of the same degree remain. It is

$$H_n''(x) - 2xH_n'(x) + 2nH_n(x) = 0 \tag{164.7}$$

The first six Hermitian polynomials are

$$H_0(x) = 1$$
$$H_1(x) = 2x$$
$$H_2(x) = 4x^2 - 2$$

$$H_3(x) = 8x^3 - 12x$$
$$H_4(x) = 16x^4 - 48x^2 + 12$$
$$H_5(x) = 32x^5 - 160x^3 + 96x + 24$$

The generating function can often be used to great advantage in evaluating definite integrals involving $H_n(x)$. As an illustration we shall use it to discuss the properties of the integral

$$I_{mn} = \int_{-\infty}^{+\infty} H_m(x)H_n(x)e^{-x^2}dx \tag{164.8}$$

Instead of working with this integral directly, it is convenient to consider the integral

$$\int_{-\infty}^{+\infty} g(x,t)H_n(x)e^{-x^2}dx = \int_{-\infty}^{+\infty} \left(\sum_{m=0}^{\infty} \frac{H_m(x)t^m}{m!}\right) H_n(x)e^{-x^2}dx$$

The series representation (164.2) has replaced $g(x,t)$. The order of summation and integration may be inverted so that

$$\int_{-\infty}^{+\infty} g(x,t)H_n(x)e^{-x^2}dx = \sum_{m=0}^{\infty} \frac{t^m}{m!} \int_{-\infty}^{+\infty} H_m(x)H_n(x)e^{-x^2}dx$$
$$= \sum_{m=0}^{\infty} I_{mn}\frac{t^m}{m!}$$

We have thus obtained a kind of generating function for all integrals of the form (164.8), so that

$$I_{mn} = \left[\frac{\partial^m}{\partial t^m} \int_{-\infty}^{+\infty} g(x,t)H_n(x)e^{-x^2}dx\right]_{t=0} \tag{164.9}$$

This integral may now be evaluated in an especially simple manner by replacing $g(x,t)$ by its explicit form from (164.2) and replacing $H_n(x)$ by an expression of the form (164.3). Then

$$\int_{-\infty}^{+\infty} g(x,t)H_n(x)e^{-x^2}dx = \int_{-\infty}^{+\infty} e^{-(t-x)^2}\left[\frac{\partial^n}{\partial s^n} e^{x^2-(s-x)^2}\right]_{s=0} dx$$
$$= \left[\frac{\partial^n}{\partial s^n} \int_{-\infty}^{+\infty} e^{-(x-s-t)^2+2st}dx\right]_{s=0}$$
$$= \left[\frac{\partial^n}{\partial s^n} e^{2st} \int_{-\infty}^{+\infty} e^{-y^2}dy\right]_{s=0}$$
$$= \sqrt{\pi}\left[\frac{\partial^n}{\partial s^n} e^{2st}\right]_{s=0}$$

From 164.9 we finally have

$$I_{mn} = \sqrt{\pi}\left[\frac{\partial^m}{\partial t^m}\frac{\partial^n}{\partial s^n} e^{2st}\right]_{\substack{s=0\\t=0}}$$

We see that I_{mn} is merely the coefficient of the term $t^m s^n$ in the power series expansion of $\sqrt{\pi}\, e^{2st}$. On the other hand, the power series expan-

sion of e^{2st} contains only terms having the same power of s as of t, i.e., only terms in which $m = n$. Thus we have the result

$$\begin{aligned} I_{mn} &= 0 \qquad m \neq n \\ I_{nn} &= \sqrt{\pi}\,\frac{\partial^n}{\partial s^n}\frac{\partial^n}{\partial t^n} e^{2st}\Big]_{\substack{s=0\\t=0}} \\ &= \sqrt{\pi}\, 2^n(n!) \end{aligned} \tag{164.10}$$

This result brings out the orthogonality of the set of functions $H_0(x)e^{-x^2/2}$, $H_1(x)e^{-x^2/2}$, etc., which was already mentioned in Sec. 158. From the value of I_{nn} we may convert this set of orthogonal functions to an orthonormal set $\varphi_0(x)$, $\varphi_1(x)$, $\varphi_2(x)$, ... , $\varphi_n(x)$ ... by writing

$$\varphi_n(x) = \frac{e^{-x^2/2}}{\sqrt{\sqrt{\pi}2^n n!}} H_n(x) \quad n = 0, 1, 2, \ldots \tag{164.11}$$

The differential equation satisfied by $\varphi_n(x)$ may readily be found by solving the above equation for $H_n(x)$ and substituting the result in the differential eq. (164.7) for $H_n(x)$. The equation is

$$\varphi_n''(x) + (2n + 1 - x^2)\varphi_n(x) = 0 \tag{164.12}$$

The functions $\varphi_n(x)$ form a complete orthonormal set, so that functions defined over the interval $-\infty < x < \infty$ may be expanded in terms of them in the sense of Sec. 156.

Solutions of the above differential equation when n is not an integer, but an arbitrary complex number ν can be found by the following extension of the definition of (164.4). First e^{-x^2} is represented by means of a contour integral (see Sec. 86); thus

$$e^{-x^2} = \frac{1}{2\pi i}\int_C \frac{e^{-z^2}dz}{(z - x)}$$

where the contour encloses the point $z = x$. From (164.4) and (87.2),

$$\begin{aligned} H_n(x) &= \frac{(-1)^n}{2\pi i} e^{x^2}\frac{d^n}{dx^n}\int_C \frac{e^{-z^2}dz}{(z - x)} \\ &= \frac{\Gamma(n+1)}{2\pi i} e^{x^2}\int_{C'} \frac{e^{-z^2}dz}{(z + x)^{n+1}} \end{aligned} \tag{164.13}$$

The contour C' encloses the point $z = -x$. Since the functions on the right are defined for any complex number $n = \nu$ except negative integers,

$$\varphi_\nu(x) = C\,\frac{\Gamma(\nu+1)}{2\pi i} e^{x^2/2}\int_{C'} \frac{e^{-z^2}dz}{(z + x)^{\nu+1}} \tag{164.14}$$

where C is any constant.

165. Laguerre polynomials and the Laguerre orthonormal functions. The set of orthogonal functions built on the interval $0 \leq x < \infty$ by means of the procedure outlined in Sec. 158 with the

help of the auxiliary functions e^{-x} leads to the Laguerre polynomial $L_n(x)$. The generating function (see Sec. 164) for these polynomials is

$$g(x,t) = \frac{e^{-xt/(1-t)}}{1-t} = \sum_{n=0}^{\infty} \frac{L_n(x)}{n!} t^n \tag{165.1}$$

so that

$$L_n(x) = \frac{\partial^n g}{\partial t^n}\Big]_{t=0} = e^x \frac{d^n}{dx^n}(x^n e^{-x}) \tag{165.2}$$

$$= (-1)^n \left[x^n - \frac{n^2}{1!} x^{n-1} + \frac{n^2(n-1)^2}{2!} x^{n-2} - \cdots + (-1)^n n! \right] \tag{165.3}$$

The equivalence of the nth derivative of g with respect to t at $t = 0$, and the nth derivative with respect to x as expressed in eq. (165.2) can be demonstrated in an instructive way by using a contour integral representation of $g(x,t)$, which incidentally leads to certain useful integral representations of $L_n(x)$. To do this $g(x,t)$ can be expressed in terms of the following Cauchy integral (see Sec. 86).

$$g(x,t) = \frac{1}{2\pi i} \int_C \frac{e^{-xz/(1-z)}}{(1-z)(z-t)}\, dz$$

where the contour C encloses the point $z = t$, and for our purpose $t < 1$. The contour C must not enclose the point $z = 1$. Expressing $L_n(x)$ in terms of the above integral,

$$L_n(x) = \frac{\partial^n g(x,t)}{\partial t^n}\Big]_{t=0} = \frac{\Gamma(n+1)}{2\pi i} \int_{C'} \frac{e^{-xz/(1-z)}}{(1-z)z^{n+1}}\, dz \tag{165.4}$$

where C' encloses the point $z = 0$. The above equation furnishes one integral representation of $L_n(x)$, but this can be made somewhat more tractable by a change of complex variable from z to s according to the transformation

$$\frac{xz}{1-z} = s - x \quad \text{or} \quad z = \frac{s-x}{s}$$

In terms of the new variable

$$L_n(x) = \frac{\Gamma(n+1)}{2\pi i} e^x \int_{C''} \frac{e^{-s} s^n}{(s-x)^{n+1}}\, ds \tag{165.5}$$

wherein C'' encloses the point $s = x$ in the s plane. From eq. (87.2) the right side of the above equation is

$$L_n(x) = e^x \frac{d^n}{dx^n}(x^n e^{-x})$$

Equation (165.5) provides a useful integral representation of $L_n(x)$ and serves to define $L_n(x)$ in those cases where n is not an integer. If n is

not an integer, $s = 0$ and $s = x$ are branch points, and the contour must enclose both points in order to remain on the same branch of the integrand.

When n is an integer, $L_n(x)$ can also be expressed as a definite integral, which is often useful in evaluating integrals involving $L_n(x)$. This integral representation is

$$L_n(x) = \frac{e^x}{2\Gamma(n+1)} \int_0^\infty \left(\frac{u^2}{4}\right)^n e^{-u^2/4} J_0(u\sqrt{x})u \, du \qquad (165.6)$$

where $J_0(y)$ is the Bessel function of order zero (see Sec. 169).

The first few Laguerre polynomials are listed below.

$$L_0(x) = 1 \qquad L_1(x) = -x + 1$$
$$L_2(x) = x^2 - 4x + 2 \qquad L_3(x) = -x^3 + 9x^2 - 18x + 6$$
$$L_4(x) = x^4 - 16x^3 + 72x^2 - 96x + 24$$

With the help of the generating function (165.1), several important recurrence relations between the Laguerre polynomials can be derived. For instance, by differentiating $g(x,t)$ partially with respect to t we can establish the relation

$$(1 - t)^2 \frac{\partial g(x,t)}{\partial t} = (1 - t - x)g(x,t)$$

On taking the nth partial derivative of this equation with respect to t, the following relation is obtained.

$$(1 - t)^2 \frac{\partial^{n+1} g}{\partial t^{n+1}} - 2n(1 - t)\frac{\partial^n g}{\partial t^n} + \frac{2n(n-1)}{2!} \frac{\partial^{n-1} g}{\partial t^{n-1}}$$
$$= (1 - t - x)\frac{\partial^n g}{\partial t} - n\frac{\partial^{n-1} g}{\partial t^{n-1}}$$

Placing $t = 0$ and using the relation (165.2), the first important recurrence formula results, namely,

$$L_{n+1}(x) - (2n + 1 - x)L_n(x) + n^2 L_{n-1}(x) = 0 \qquad (165.7)$$

This recurrence relation allows one to compute one Laguerre polynomial in terms of two others.

By carrying out similar operations on the relation

$$(1 - t)\frac{\partial g}{\partial x} = -tg$$

one obtains another relation between the L's, namely,

$$L_n'(x) - nL_{n-1}'(x) = -nL_{n-1}(x) \qquad (165.8)$$

which is a relation involving first derivatives of the L's and only two Laguerre polynomials. Using this relation to eliminate $L_{n+1}(x)$ in eq. (165.7), a somewhat more useful relation results, i.e.,

$$xL_n'(x) - nL_n(x) + n^2 L_{n-1}(x) = 0 \qquad (165.9)$$

The differential equation satisfied by $L_n(x)$ can readily be obtained by differentiating eq. (165.7) once and using relations (165.8) and (165.9) to obtain a relation involving only $L_n(x)$ and its derivatives. This operation gives

$$xL_n''(x) + (1 - x)L_n'(x) + nL_n(x) = 0 \tag{165.10}$$

The above recurrence formulas hold for n noninteger provided that $L_n(x)$ is appropriately defined, as for example by the integral (165.5).

As pointed out at the beginning of the chapter, the $L_n(x)$'s in connection with the auxiliary function e^{-x} form an orthogonal set of functions on the interval $0 \leq x < \infty$, and by methods similar to those used in establishing the result given in (164.10), it can be shown that

$$\int_0^\infty e^{-x}L_m(x)L_n(x)dx = \begin{cases} 0 & \text{for } m \neq n \\ (n!)^2 & \text{for } m = n \end{cases} \tag{165.11}$$

On the basis of this property the orthonormal set of functions built on the Laguerre polynomials is

$$\varphi_n(x) = \frac{e^{-x/2}L_n(x)}{n!} \qquad n = 0, 1, 2, \ldots$$

This set of orthonormal functions is also a complete set.

As an example of formal manipulation with this set of orthonormal functions, let it be required to represent the function

$$f(x) = \begin{matrix} 1 & 0 \leq x \leq a \\ 0 & a < x \end{matrix}$$

in terms of the above set of functions, i.e.,

$$f(x) = \sum_{n=0}^{\infty} c_n\varphi_n(x) \tag{165.12}$$

where in accordance with (156.3),

$$\begin{aligned} c_n &= \int_0^a \varphi_n(x)dx \\ &= \frac{1}{n!}\int_0^a e^{-x/2}L_n(x)dx \end{aligned} \tag{165.13}$$

Two procedures for evaluating these coefficients will be illustrated. In a practical case one may prove more convenient than the other, depending on the use to be made of the series (165.12).

The first procedure will be to obtain a recurrence relation in terms of which all the c_n's can be computed once one of them is known. To do this, the above integral is integrated once by parts, which leads to

$$\int_0^a e^{-x/2}L_n(x)dx = -2e^{-a/2}L_n(a) + 2L_n(0) + 2\int_0^a e^{-x/2}L_n'(x)dx$$

This is an opportune time to use the recurrence formula (165.8), in which case the last integral becomes

$$2n\int_0^a e^{-x/2}[L'_{n-1}(x) - L_{n-1}(x)]dx$$
$$= 2n[e^{-a/2}L_{n-1}(a) - L_{n-1}(0)] - n\int_0^a e^{-x/2}L_{n-1}(x)dx$$

wherein another integration by parts has been performed. Combining the above two equations and noting from (165.3) that $L_n(0) = n!$, one obtains the following recurrence formula for c_n.

$$c_n = \frac{2}{n!}e^{-a/2}[nL_{n-1}(a) - L_n(a)] - c_{n-1} \tag{165.14}$$

which holds for $n \geq 1$. Here c_0 can be readily computed directly from eq. (165.13), since $L_0 = 1$.

As an alternate procedure the integral (165.13) can be evaluated. To do this it is convenient to use an integral representation of $L_n(x)$ and invert the order of integration. However, on looking at the integral representation (165.5), the variable x appears in the denominator, which makes it difficult to integrate the function $e^{+x/2}/(s-x)^{n+1}$. Consequently it is best to modify the integral representation (165.5) by making a transformation of complex variable from s to z by means of the relation

$$s - x = x\left(z - \frac{1}{2}\right)$$

whereupon the integral in (165.5) becomes

$$L_n(x) = \frac{\Gamma(n+1)}{2\pi i}e^{x/2}\int_C e^{-xz}\left(z+\frac{1}{2}\right)^n\left(z-\frac{1}{2}\right)^{-n-1}dz \tag{165.15}$$

The path of integration C must enclose the point $z = \frac{1}{2}$ if n is a positive integer or zero. If n is not an integer, the path must enclose both branch points $z = \frac{1}{2}$ and $z = -\frac{1}{2}$ in order to remain on the same branch of the integrand. The presence of x only in the exponential term makes this representation extremely useful in practice.

With the above representation the formula for c_n, eq. (165.13) becomes

$$c_n = \frac{1}{2\pi i}\int_0^a dx\int_C e^{-xz}\left(z+\frac{1}{2}\right)^n\left(z-\frac{1}{2}\right)^{-n-1}dz$$
$$= \frac{1}{2\pi i}\int_C\left(z+\frac{1}{2}\right)^n\left(z-\frac{1}{2}\right)^{-n-1}(1-e^{-az})\frac{dz}{z}$$

where the order of integration has been reversed and the integration with respect to x has been performed. The resulting contour integral can be evaluated by the method of residues [see Sec. 93 and in particular formula (94.2)], whereupon

$$c_n = \lim_{z \to 1/2} \frac{1}{n!} \frac{d^n}{dz^n} \frac{\left(z + \frac{1}{2}\right)^n (1 - e^{-az})}{z}$$

It is evident that this will lead to polynomials in a for the coefficients c_n.

166. Associated Laguerre polynomials. Another set of polynomials much used in quantum theory are closely related to the Laguerre polynomials discussed in the last section. These also can be used as a basis for an orthonormal set on the interval $0 \leq x < \infty$. These are known as the Associated Laguerre polynomials $L_{n+k}^k(x)$ defined in terms of $L_n(x)$ (see Sec. 165) according to the relation

$$L_{n+k}^k(x) = \frac{d^k}{dx^k} L_{n+k}(x) \tag{166.1}$$

The generating function for $L_{n+k}^k(x)$ can be found from that for $L_n(x)$ by writing the relation (165.1) in the form

$$\begin{aligned} \frac{e^{-xt/(1-t)}}{(1-t)} &= \sum_{m=0}^{k-1} \frac{L_m(x)t^m}{m!} + \sum_{m=k}^{\infty} \frac{L_m(x)t^m}{m!} \\ &= \sum_{n=0}^{k-1} \frac{L_n(x)t^n}{n!} + \sum_{n=0}^{\infty} \frac{L_{n+k}(x)}{(n+k)!} t^{n+k} \end{aligned}$$

This result may be differentiated k times with respect to x, in which case the first term on the right is reduced to zero, since only polynomials of degree $k - 1$ in x appear. Consequently

$$\frac{(-t)^k}{(1-t)^{k+1}} e^{-xt/(1-t)} = \sum_{n=0}^{\infty} \frac{L_{n+k}^k(x)}{(n+k)!} t^{n+k}$$

and the generating function is

$$g(x,t) = \frac{(-1)^k}{(1-t)^{k+1}} e^{-xt/(1-t)} = \sum_{n=0}^{\infty} \frac{L_{n+k}^k(x)}{(n+k)!} t^n \tag{166.2}$$

and

$$L_{n+k}^k(x) = \frac{(n+k)!}{n!} \frac{\partial^n}{\partial t^n} g(x,t) \Big]_{t=0} \tag{166.3}$$

Recurrence relations between these polynomials can be found in the same manner as that used for $L_n(x)$. For instance, the recurrence relation corresponding to (165.7) is

$$L_{n+k+1}^k(x) - (2n + 1 + k - x)L_{n+k}^k(x) + n(n+k)L_{n+k-1}^k = 0 \tag{166.4}$$

This relation is particularly convenient when it is desired to compute polynomials with a different lower index while the upper index remains the same.

The relation corresponding to (165.8) is

$$\frac{d}{dx} L_{n+k}^k(x) - n \frac{d}{dx} L_{n+k-1}^k(x) + nL_{n+k-1}^k = 0 \tag{166.5}$$

Combining the formulas (166.4) and (166.5), the recurrence relation corresponding to (165.9) is

$$x\frac{d}{dx}L_{n+k}^{k}(x) - nL_{n+k}^{k}(x) + n(n+k)L_{n+k-1}^{k}(x) = 0 \qquad (166.6)$$

The differential equation for $L_{n+k}^{k}(x)$ is

$$x\frac{d^2}{dx^2}L_{n+k}^{k}(x) + (k+1-x)\frac{d}{dx}L_{n+k}^{k}(x) + nL_{n+k}^{k}(x) = 0 \qquad (166.7)$$

and it can be obtained by using methods similar to those used in obtaining the differential equation for $L_n(x)$ in the last section.

The associated Laguerre polynomials form an orthonormal set when used with the auxiliary function $x^k e^{-x}$. Since the orthogonality condition*

$$\int_0^\infty x^k e^{-x} L_{n+k}^{k}(x) L_{m+k}^{k}(x)dx = \begin{cases} 0 & m \neq n \\ \dfrac{[(n+k)!]^3}{n!} & m = n \end{cases} \qquad (166.8)$$

is satisfied we can form the orthonormal set of functions

$$\varphi_{nk}(x) = \sqrt{\frac{n!}{[(n+k)!]^3}}\, x^{k/2} e^{-x/2} L_{n+k}^{k}(x) \qquad (166.9)$$

As an example of the use of this set of functions let it be required to represent the function $x^{k/2}e^{-(\alpha+i\beta)x}$ in terms of the above orthonormal set, where α and β are real. Thus

$$x^{k/2}e^{-(\alpha+i\beta)x} = \sum_{n=0}^{\infty} c_{nk}\varphi_{nk}(x) \qquad (166.10)$$

and

$$c_n = \int_0^\infty x^{k/2}e^{-(\alpha+i\beta)x}\varphi_{nk}(x)dx \qquad (166.11)$$

$$= \sqrt{\frac{n!}{[(n+k)!]^3}}\int_0^\infty x^k e^{-(\alpha+i\beta)x} e^{-x/2} L_{n+k}^{k}(x)dx$$

In evaluating this integral it is convenient to deduce an integral representation of $x^k e^{-x/2}L_{n+k}^{k}(x)$, then invert the order of integration and evaluate the resulting integrals. In order to determine the above-mentioned integral representation it is convenient to start with the integral representation of $L_n(x)$ given by (165.5) and replace n by $n+k$, so that

$$L_{n+k}(x) = \frac{\Gamma(n+k+1)}{2\pi i} e^x \int_C e^{-s} s^{n+k}(s-x)^{-n-k-1}ds$$

Since according to (166.1), this has to be differentiated k times with respect to x to obtain $L_{n+k}^{k}(x)$, and it is in turn awkward to carry out

* For the value of more general integrals of this type see E. Schrodinger, *Ann. Physik*, **80**, 486 (1926).

this differentiation with this integral in the above form, we make the transformation of variable $\xi = s - x$, so that

$$L_{n+k}(x) = \frac{\Gamma(n+k+1)}{2\pi i}\int_C e^{-\xi}(\xi + x)^{n+k}\xi^{-n-k-1}d\xi$$

where the path of integration encloses the point $\xi = 0$. If $n + k$ is not an integer, the path must enclose both branch points $\xi = 0$ and $\xi = -x$. The k-fold differentiation mentioned above can now be readily carried through, and the result is

$$L^k_{n+k}(x) = \frac{[\Gamma(n+k+1)]^2}{2\pi i\Gamma(n+1)}\int_C e^{-\xi}(\xi + x)^n\xi^{-n-k-1}d\xi \qquad (166.12)$$

This provides one possible integral representation of $L^k_{n+k}(x)$, which may serve also to define an associate Laguerre function when n or k or both are not integers, as is sometimes the case in quantum theory applications. For use in connection with the integral (166.11) it is more convenient to have the x appear in the exponential. This can be achieved by the further transformation of variable $\xi = x(z - \frac{1}{2})$, after which we have

$$x^k e^{-x/2}L^k_{n+k}(x) =$$
$$\frac{[\Gamma(n+k+1)]^2}{2\pi i\Gamma(n+1)}\int_C e^{-xz}\left(z + \frac{1}{2}\right)^n\left(z - \frac{1}{2}\right)^{-n-k-1}dz \qquad (166.13)$$

the path C enclosing the points $z = \frac{1}{2}$ and $z = -\frac{1}{2}$.

Inserting this integral representation in (166.11) and inverting the order of integration,

$$c_n = \sqrt{\frac{(n+k)!}{n!}}\frac{1}{2\pi i}\int_C\left(z + \frac{1}{2}\right)^n\left(z - \frac{1}{2}\right)^{-n-k-1}dz\int_0 e^{-(z+\alpha+i\beta)x}dx$$
$$= \sqrt{\frac{(n+k)!}{n!}}\frac{1}{2\pi i}\int_C\frac{\left(z + \frac{1}{2}\right)^n\left(z - \frac{1}{2}\right)^{-n-k-1}}{z + \alpha + i\beta}dz$$

The integration over x has introduced a new singular point, and it must be determined whether it lies inside the contour C or outside it. To do this we note first that C need not enclose the point $z = -\frac{1}{2}$, as shown by the dashed curve in Fig. 166.1. The value of the integral over x has sense only if (real part of $z + \alpha$) > 0; consequently all points on the path of integration must lie to the right of $-\alpha$, so that we can conclude that the point $z = -(\alpha + i\beta)$ lies outside the contour C. Since the value of the above integral is not changed by deforming the path of integration so long as the contour excludes the point $z = -(\alpha + i\beta)$. Therefore it is convenient to deform the path C to coincide with the solid curve shown in Fig. 166.1. It can readily be shown that the integral over the circle of radius ρ is proportional to ρ^{-k-1} and hence vanishes

when $\rho \to \infty$. Then since the integrals over paths a and b cancel, we are left with just the negative of the residue at the point $z = -\alpha - i\beta$. Taking the residue (see Sec. 94),

$$c_n = -\sqrt{\frac{(n+k)!}{n!}}\left(-\alpha - i\beta + \frac{1}{2}\right)^n \left(-\alpha - i\beta - \frac{1}{2}\right)^{-n-k-1} \tag{166.14}$$

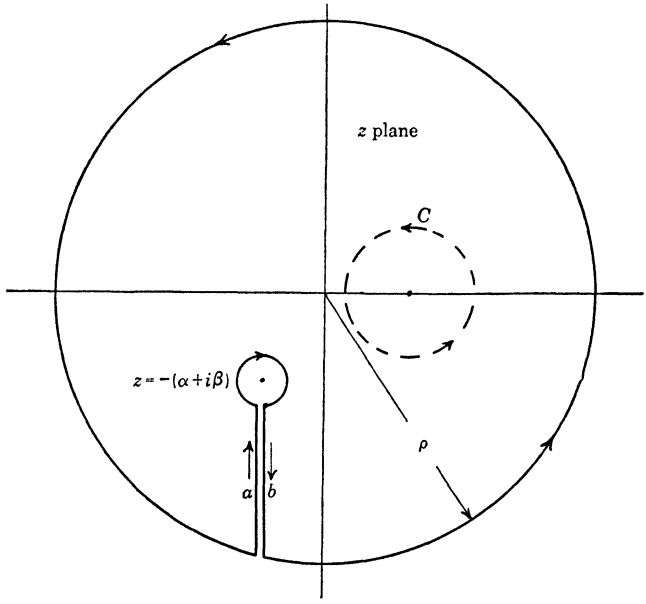


FIG. 166.1

167. Legendre polynomials. These polynomials next to the trigonometric functions are the most often used set of orthonormal functions in theoretical physics. They arose from the representation of the quantity $(r^2 + R^2 - 2rR\cos\theta)^{-1/2}$ as a power series in $r/R < 1$, where r and R are the magnitudes of vectors making an angle θ with each other, and the above square root is the distance between their extremities. The polynomials are the coefficients of the various powers of r/R. Consequently the generating function (see Sec. 164) for the Legendre polynomials is

$$g(x,t) = (1 - 2xt + t^2)^{-1/2} = \sum_{n=0}^{\infty} P_n(x)t^n \tag{167.1}$$

The polynomials $P_n(x)$ are defined over the interval $-1 \leq x \leq 1$, as we would infer from the fact that in writing the generating function x has replaced $\cos\theta$. From (167.1),

$$P_n(x) = \frac{1}{n!}\left[\frac{\partial^n g(x,t)}{\partial t^n}\right]_{t=0} \tag{167.2}$$

Also the generating function can readily be expanded as a power series in t to obtain the first few polynomials. This gives

$$\begin{aligned} g(x,t) &= (1 - 2xt + t^2)^{-1/2} \\ &= 1 + xt + \tfrac{1}{2}(3x^2 - 1)t^2 + \tfrac{1}{2}(5x^3 - 3x)t^3 + \ldots \end{aligned} \tag{167.3}$$

whereupon

$$\begin{aligned} P_0(x) &= 1 & P_0(\cos\theta) &= 1 \\ P_1(x) &= x & P_1(\cos\theta) &= \cos\theta \\ P_2(x) &= \tfrac{1}{2}(3x^2 - 1) & P_2(\cos\theta) &= \tfrac{1}{4}(3\cos 2\theta + 1) \\ P_3(x) &= \tfrac{1}{2}(5x^3 - 3x) & P_3(\cos\theta) &= \tfrac{1}{8}(5\cos 3\theta + 3\cos\theta) \end{aligned}$$

Just as the polynomials in the previous section can be expressed in terms of derivatives of a function of x alone, so can the Legendre polynomials. To do this it is convenient to obtain an integral representation of $P_n(x)$ with the help of the generating function. First the generating function (167.1) can be expressed in terms of a contour integral by the Cauchy method (see Sec. 86). This gives

$$g(x,t) = \frac{1}{2\pi i}\int_C \frac{(1 - 2xz + z^2)^{-1/2}}{z - t}\,dz$$

The contour C must enclose the point $z = t$, but on account of the fact that the numerator of the integrand is a multiple-valued function (see Sec. 92), the branch points $z = x \pm i\sqrt{1 - x^2}$ for $x < 1$ should also be included in the contour in order to remain on the same branch, in this case the positive branch, of the two-valued function. The appropriate path is shown in Fig. 167.1. From (167.2) and (87.2),

$$P_n(x) = \frac{1}{n!}\frac{\partial^n g(x,t)}{\partial t^n}\Big]_{t=0} = \frac{1}{2\pi i}\int_C \frac{(1 - 2xz + z^2)^{-1/2}}{z^{n+1}}\,dz \tag{167.4}$$

It must now be borne in mind that C must include $z = 0$, since the special value of $t = 0$ has been used. The above integral can be transformed so as to represent the nth derivative of a function of x. To do this, the transformation from the z plane to an s plane according to the relation

$$1 - 2xz + z^2 = (1 - zs)^2$$

is effected. Under these circumstances, the path C should include the points $s = x \pm i\sqrt{1 - x^2}$ and $s = x$. The transformed integral is

$$P_n(x) = \frac{1}{2\pi i}\frac{1}{2^n}\int_C \frac{(s^2-1)^n ds}{(s-x)^{n+1}} \tag{167.5}$$

This provides a useful integral representation of $P_n(x)$ for n integer. From (87.2), it is clear that

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n \tag{167.6}$$

This is known as Rodrigues' formula for $P_n(x)$.

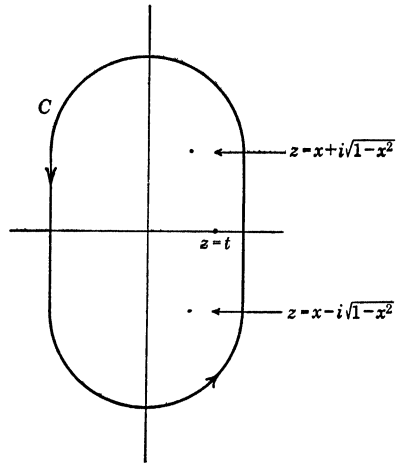


FIG. 167.1

The Legendre polynomials have the property that

$$\int_{-1}^{+1} P_n(x)P_m(x)dx = \begin{cases} 0 & m \neq n \\ \dfrac{2}{2n+1} & m = n \end{cases} \tag{167.7}$$

Since this result can be obtained by a useful method of using the generating function for $P_n(x)$, the method will be illustrated. First multiply two generating functions of the form (167.1) together with suitably changed sum indexes and dummy variable.

$$(1-2xt+t^2)^{-1/2}(1-xs+s^2)^{-1/2} = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} P_n(x)P_m(x)s^n t^m$$

This is next integrated with respect to x over the interval $-1 \leq x \leq 1$, and on the right side the order of summation and integration is interchanged. Thus

$$\int_{-1}^{+1} (1 - 2xt + t^2)^{-1/2}(1 - 2xs + s^2)^{-1/2}dx$$
$$= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\left[\int_{-1}^{+1} P_n(x)P_m(x)dx\right] s^n t^m$$
$$= \frac{1}{\sqrt{st}} \log \frac{1 + \sqrt{st}}{1 - \sqrt{st}} \tag{167.8}$$

the integral having been evaluated by means similar to those used in Sec. 28. On expanding the logarithmic term in powers of st, we have the identity

$$\sum_{n=0}^{\infty} \frac{2}{2n+1} s^n t^n = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\left[\int_{-1}^{+1} P_n(x)P_m(x)dx\right] s^n t^m$$

Since these power series in the two independent variables s and t can be identical only if the coefficients of the various powers $s^n t^m$ are equal, we see at once that all coefficients on the right for which m is not equal to n vanish, so that

$$\int_{-1}^{+1} P_n(x)P_m(x)dx = \begin{cases} 0 & m \neq n \\ \dfrac{2}{2n+1} & m = n \end{cases}$$

Consequently, the orthonormal set of functions based on the Legendre polynomials is

$$\varphi_n = \sqrt{\frac{2n+1}{2}}\, P_n(x) \qquad n = 0, 1, 2, \ldots \tag{167.9}$$

Recurrence Relations. Recurrence relations can be obtained for the Legendre polynomials in the same way as for the Hermitian or Laguerre polynomials in Secs. 164 and 165, from the relations

$$(1 - 2xt + t^2)\frac{\partial g}{\partial t} = (x - t)g$$
$$t\frac{\partial g}{\partial t} = (x - t)\frac{\partial g}{\partial x}$$

The most useful recurrence relations are written below.

$$(n+1)P_{n+1}(x) - (2n+1)xP_n(x) + nP_{n-1}(x) = 0 \tag{167.10}$$
$$xP_n'(x) - P_{n-1}'(x) - nP_n(x) = 0 \tag{167.11}$$
$$P_{n+1}'(x) - xP_n'(x) - (n+1)P_n(x) = 0 \tag{167.12}$$
$$P_{n+1}'(x) - P_{n-1}'(x) - (2n+1)P_n(x) = 0 \tag{167.13}$$
$$(x^2 - 1)P_n'(x) - nxP_n(x) + nP_{n-1}(x) = 0 \tag{167.14}$$

By means of these formulas the various order polynomials and derivatives can be computed in terms of two others. These relations are especially useful in evaluating integrals involving Legendre polynomials. An example of their use in this connection will be given later in this section.

Also by differentiating and combining some of the above relations one finds that $P_n(x)$ is a solution of the differential equation

$$(1 - x^2)P_n''(x) - 2xP_n'(x) + n(n + 1)P_n(x) = 0 \qquad (167.15)$$

The other linearly independent solution of this differential equation, often denoted by $Q_n(x)$ has a singularity at $x = 1$. Since in most applications this would mean a singularity at an angle of $\theta = 0$, these Legendre functions do not often appear in physical problems.

As an example of an expansion in terms of the orthonormal set of functions (167.9), we may consider the representation of the function $P_l(\cos\theta)\sin^2\theta$ in terms of the φ_n's, i.e.,

$$P_l(\cos\theta)\sin^2\theta = \sum_{n=0}^{\infty} c_n\varphi_n(\cos\theta)$$

or changing over to the variable $x = \cos\theta$,

$$P_l(x)(1 - x^2) = \sum_{n=0}^{\infty} c_n\varphi_n(x) = \sum_{n=0}^{\infty} c_n\sqrt{\frac{2n+1}{2}}\,P_n(x)$$

From the orthonormal property of the φ_n's

$$c_n = \int_{-1}^{+1} P(x)(1 - x^2)\varphi_n(x)dx = \sqrt{\frac{2n+1}{2}}\int_{-1}^{+1}(1 - x^2)P_l(x)P_n(x)dx$$

where c_n may be evaluated readily in terms of the integral formula (167.7) if the recurrence formula (167.10) is applied twice to express $x^2P_l(x)$ in terms of P's alone. When this is done it follows that

$$\sqrt{\frac{2n+1}{2}}\,c_n =$$

$$\frac{2n+1}{2}\left\{\left[1 - \frac{(l+1)^2}{(2l+1)(2l+3)} - \frac{l^2}{(2l+1)(2l-1)}\right]\int_{-1}^{+1} P_l(x)P_n(x)dx\right.$$

$$- \frac{l(l-1)}{(2l+1)(2l-1)}\int_{-1}^{+1} P_{l-2}(x)P_n(x)dx$$

$$\left. - \frac{(l+1)(l+2)}{(2l+1)(2l+3)}\int_{-1}^{+1} P_{l+2}(x)P_n(x)dx\right\}$$

From the orthogonality of the P's it is seen that c_n is zero for all n not equal to $l - 2$, l, or $l + 2$. From the formula (167.7),

$$\sqrt{\frac{2l-3}{2}}\,c_{l-2} = \frac{-l(l-1)}{(2l-1)(2l+1)}$$

$$\sqrt{\frac{2+1}{2}}\,c_l = 1 - \frac{4l^3 + 6l^2 - 1}{(2l-1)(2l+1)(2l+3)}$$

$$\sqrt{\frac{2+5}{2}}\,c_{l+2} = -\frac{(l+1)(l+2)}{(2l+1)(2l+3)}$$

Since the solution of Laplace's equation for potential distributions which depend only on r and θ; i.e., spherical distributions not depending on the azimuth angle φ can always be expressed in the form

$$V = \sum_{n=0}^{\infty} c_n \left(\frac{r}{a}\right)^n P_n(\cos\theta)$$

or

$$V = \sum_{n=0}^{\infty} c_n' \left(\frac{a}{r}\right)^{n+1} P_n(\cos\theta)$$

where a is the radius of a sphere on which the potential is a given function of the angle θ, i.e., at $r = a$, $V = f(\theta)$. The coefficients c_n, c_n' are then determined with the help of (167.7).

168. Associated Legendre functions and spherical harmonics. Just as the associated Laguerre polynomials (see Sec. 166) were defined in terms of the Laguerre polynomials by a process of differentiation, so also is the associated Legendre function defined. Two definitions of the associated Legendre function are in common use. For real values of x in the interval $-1 \leq x \leq 1$, Ferrer's definition is customarily used, namely,

$$P_n^m(x) = (1 - x^2)^{m/2} \frac{d^m}{dx^m} P_n(x) \qquad m \leq n \tag{168.1}$$

For rare applications where x is replaced by the complex variable z, it is convenient to use Hobson's definition,

$$P_n^m(z) = (z^2 - 1)^{m/2} \frac{d^m}{dz^m} P_n(z)$$

The properties of the associated Legendre function according to Ferrer will be given in the following.

The generating function for $P_n^m(x)$ can be obtained from (167.1) and (168.1) in a manner similar to that used to obtain the associated Laguerre polynomial (see Sec. 166). The relevant generating function is

$$g(x,t) = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2m + 1)(1 - x^2)^{m/2}(1 - 2xt + t^2)^{-(2m+1)/2}$$

$$= \sum_{\nu=0}^{\infty} P_{\nu+m}^m(x) t^\nu \tag{168.2}$$

Thus

$$P_{\nu+m}^m(x) = \nu! \left[\frac{\partial^\nu g}{\partial t^\nu}\right]_{t=0} \tag{168.3}$$

Some associated Legendre functions $P_n^m(x)$ of low degree n and order m are shown in the following table.

TABLE FOR $P_n^m(x)$

n	$m = 1$	$m = 2$	$m = 3$
1	$(1 - x^2)^{1/2}$		
2	$3(1 - x^2)^{1/2}x$	$3(1 - x)^2$	
3	$\frac{3}{2}(1 - x^2)^{1/2}(5x^2 - 1)$	$15(1 - x^2)x$	$15(1 - x^2)^{3/2}$
4	$\frac{5}{2}(1 - x^2)^{1/2}(7x^3 - 3x)$	$\frac{15}{2}(1 - x^2)(7x^2 - 1)$	$105(1 - x^2)^{3/2}x$

By differentiating the integral representation of $P_n(x)$ in (167.5) m times with respect to x, the following integral representation for $P_n^m(x)$ can be obtained.

$$P_n^m(x) = \frac{(n+m)!}{2^n n!} \frac{1}{2\pi i} \int_C \frac{(1-x^2)^{m/2}(s^2-1)^n}{(2-x)^{n+m+1}} ds \qquad (168.4)$$

where the path of integration must include the point $s = x$.

Recurrence Formulas. Recurrence formulas for the associated Legendre functions can be established by methods used in the preceding sections or by means of the integral representation for $P_n^m(x)$. A useful relation in which the order m remains constant is

$$(n-m+1)P_{n+1}^m(x) - (2n+1)xP_n^m(x) + (n+m)P_{n-1}^m(x) = 0 \qquad (168.5)$$

A similar formula which holds when the degree n remains constant is

$$P_n^{m+2}(x) - 2(m+1)\frac{x}{(1-x^2)^{1/2}}P_n^{m+1}(x) + (n-m)(n+m+1)P_n^m = 0 \qquad (168.6)$$

When $x = \cos\theta$, this relation has the familiar form

$$P_n^{m+2}(\cos\theta) - 2(m+1)(\cot\theta)P_n^{m+1}(\cos\theta) + (n-m)(n+m+1)P_n^m(\cos\theta) = 0 \qquad (168.7)$$

A very useful composite recurrence relation between the associated Legendre functions is

$$(1-x^2)^{1/2}P_n^m(x) = \frac{1}{2n+1}[P_{n+1}^{m+1}(x) - P_{n-1}^{m+1}] \qquad (168.8)$$

or

$$(\sin\theta)P_n^m(\cos\theta) = \frac{1}{2n+1}[P_{n+1}^{m+1}(\cos\theta) - P_{n-1}^{m+1}(\cos\theta)] \qquad (168.9)$$

The differential equation satisfied by $P_n^m(x)$ is

$$(1-x^2)\frac{d^2}{dx^2}P_n^m(x) - 2x\frac{d}{dx}P_n^m(x) + \left[n(n+1) - \frac{m^2}{1-x^2}\right]P_n^m(x) = 0 \qquad (168.10)$$

or in the more familiar form in terms of the angle θ, it is

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left[(\sin\theta)\frac{d}{d\theta}P_n^m(\cos\theta)\right] + \left[n(n+1) - \frac{m^2}{\sin^2\theta}\right]P_n^m(\cos\theta) = 0 \qquad (168.11)$$

The set of functions $P_n^m(x)$ satisfy the orthogonality relation

$$\int_{-1}^{+1} P_l^m(x)P_n^m(x)dx = \begin{cases} 0 & l \neq n \\ \dfrac{2}{2n+1}\dfrac{(n+m)!}{(n-m)!} & l = n \end{cases} \qquad (168.12)$$

Consequently the general member of the relevant orthonormal set is

$$\varphi_n^m(x) = \sqrt{\frac{2n+1}{2}\frac{(n-m)!}{(n+m)!}}\,P_n^m(x) \tag{168.13}$$

The most general representation of a given function of x in terms of this orthonormal set would be

$$f(x) = \sum_{n=0}^{\infty}\sum_{m=0}^{n} c_{nm}\varphi_n^m(x) \tag{168.14}$$

More complicated representations than the one above are obtained as solutions of Laplace's equation in polar coordinates r, θ, φ. Two important solutions are

$$V = \sum_{n=0}^{\infty}\sum_{m=0}^{n} (c_{nm}\cos m\varphi + d_{nm}\sin m\varphi)r^n P_n^m(\cos\theta) \tag{168.15}$$

$$V = \sum_{n=0}^{\infty}\sum_{m=0}^{n} (c_{nm}\cos m\varphi + d_{nm}\sin m\varphi)r^{-n-1} P_n^m(\cos\theta) \tag{168.16}$$

The function of θ and φ given by the sum over m for any given n, i.e.,

$$Y_n(\theta,\varphi) = \sum_{m=0}^{n} (c_{nm}\cos m\varphi + d_{nm}\sin m\varphi)P_n^m(\cos\theta) \tag{168.17}$$

is called a surface spherical harmonic of degree n. An arbitrary function of θ and φ defined over the surface of a sphere can be represented as a sum of surface spherical harmonics.

The functions $r^n Y_n(\theta,\varphi)$ and $r^{-n-1}Y_n(\theta,\varphi)$ are sometimes referred to as solid spherical harmonics of the nth degree.

As an example involving manipulations with spherical harmonics, let it be required to represent the function $(\cos 2\varphi)(\sin^2\theta)$ in terms of spherical harmonics, i.e.,

$$\begin{aligned}(\cos 2\varphi)(\sin^2\theta) &= \sum_{n=0}^{\infty} a_n Y_n(\theta,\varphi)\\ &= \sum_{n=0}^{\infty}\sum_{m=0}^{n} (c_{nm}\cos m\varphi + d_{nm}\sin m\varphi)P_n^m(\cos\theta)\end{aligned}$$

Since the functions $\cos m\varphi$ and $\sin m\varphi$ are members of an orthogonal set (see Sec. 155), it follows that the only coefficient c_{nm} and d_{nm} on the right which is different from zero is c_{n2}. Consequently,

$$\sin^2\theta = \sum_{n=2}^{\infty} c_{n2}P_n^2(\cos\theta)$$

or

$$(1-x^2) = \sum_{n=2}^{\infty} c_{n2}P_n^2(x)$$

Using the property (168.12),

$$\frac{2}{2n+1}\frac{(n+2)!}{(n-2)!}c_{n2} = \int_{-1}^{+1}(1-x^2)P_n^2(x)dx$$

From the preceding table for the function $P_n^m(x)$ we may write

$$c_{n2} = \frac{2n+1}{2}\frac{(n-2)!}{(n+2)!}\int_{-1}^{+1}\frac{1}{3}P_2^2(x)P_n^2(x)dx = \begin{cases} 0 & n \neq 2 \\ \frac{1}{3} & n = 2 \end{cases}$$

Consequently

$$(\cos 2\varphi)(\sin^2\theta) = \frac{1}{3}(\cos 2\varphi)P_2^2(\cos\theta)$$

which of course could have been arrived at by inspection of the table for $P_n^m(x)$.

Another example which at first sight looks like an even simpler expansion problem, however, turns out to be somewhat more difficult. Let it be required to determine the potential at any point inside a spherical shell of radius a upon which the potential is known to be $V_0 \cos 2\varphi$. The relevant solution of Laplace's equation is given by (168.15). That is

$$V = \sum_{n=0}^{\infty}\sum_{m=0}^{n}(c_{nm}\cos m\varphi + d_{nm}\sin m\varphi)\left(\frac{r}{a}\right)^n P_n^m(\cos\theta)$$

at $r = a$, the potential is known, so that

$$V_0 \cos 2\varphi = \sum_{n=0}^{\infty}\sum_{m=0}^{n}(c_{nm}\cos m\varphi + d_{nm}\sin m\varphi)P_n^m(\cos\theta)$$

For the same reasons as given in the example above, this expansion reduces to

$$V_0 \cos 2\varphi = \cos 2\varphi \sum_{n=2}^{\infty} c_{n2}P_n^2(\cos\theta)$$

Consequently we have the problem of representing a constant V_0 in terms of the orthogonal set of functions $P_n^2(\cos\theta)$, i.e.,

$$V_0 = \sum_{n=2}^{\infty} c_{n2}P_n^2(x)$$

Multiplying through by $P_n^2(x)$ and using eq. (168.12),

$$\frac{2}{2n+1}\frac{(n+2)!}{(n-2)!}c_{n2} = V_0\int_{-1}^{+1}P_n^2(x)dx$$

The integral on the right can be evaluated by writing it

$$\int_{-1}^{+1}P_n^2(x)dx = \int_{-1}^{+1}(1-x^2)\frac{d^2}{dx^2}P_n(x)dx$$

and performing 2 integrations by parts, which yields

$$\int_{-1}^{+1} P_n^2(x)dx = 2[P_n(1) + P_n(-1)] - 2\int_{-1}^{+1} P_n(x)dx$$

The integral on the right is zero, since $n = 2$, and the integral is just a special case of the orthogonality relation (167.7). The values of $P_n(1)$ and $P_n(-1)$ can be conveniently found from the generating function (167.1), since

$$(1 - 2t + t^2)^{-1/2} = \frac{1}{1-t} = \sum_{n=0}^{\infty} P_n(1)t^n = \sum_{n=0}^{\infty} t^n$$

$$(1 + 2t + t^2)^{-1/2} = \frac{1}{1+t} = \sum_{n=0}^{\infty} P_n(-1)t^n = \sum_{n=0}^{\infty} (-1)^n t^n$$

Thus $P_n(1) = 1$ for all n and $P_n(-1) = (-1)^n$. Consequently

$$c_{n2} = V_0 \frac{2n+1}{2} \frac{(n-2)!}{(n+2)!} \cdot 2[(1) - (-1)^n]$$

and the final expression for the potential inside the spherical shell is

$$V = V_0 \cos 2\varphi \sum_{n=2}^{\infty} (2n+1) \frac{(n-2)!}{(n+2)!} [1 - (-1)^n] \left(\frac{r}{a}\right)^n P_n^2(\cos\theta)$$

The Addition Formula for Legendre Polynomials. The addition formula for Legendre polynomials allows one to express the Legendre polynomial of z, where

$$z = xy + (1 - x^2)^{1/2}(1 - y^2)^{1/2} \cos\omega \qquad (168.18)$$

in terms of associated Legendre functions of x and y alone, and trigonometric functions of ω. As a later example will indicate, this possibility is of great importance in connection with expansions of the inverse distance between two points. In its customary form the formula in question is

$$P_n(z) = P_n(x)P_n(y) + 2\sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} P_n^m(x)P_n^m(y)\cos m\omega \qquad (168.19)$$

Another very convenient form can be obtained from this by writing $\cos m\omega$ in terms of exponentials. In this case the formula can be rearranged to yield

$$P_n(z) = \sum_{m=-n}^{+n} \frac{(n - |m|)!}{(n + |m|)!} P_n^{|m|}(x)P_n^{|m|}(y)e^{im\omega} \qquad (168.20)$$

As an example of the practical use of this formula, let it be required to expand the function

$$\frac{\cos\varphi_1}{\sqrt{r_1^2 - 2r_1r_2\cos\psi + r_2^2}} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} f_{nm} P_n^{|m|}(\cos\theta_1)e^{im\varphi_1} \qquad (168.21)$$

as indicated, in which the denominator is the distance between two points whose polar coordinates are r_1, θ_1, φ_1 and r_2, θ_2, φ_2, respectively, in terms of the associated Legendre function $P_n^m(\cos\theta_1)$. The angle between the two radius vectors r_1 and r_2 is then the angle ψ, the cosine of which in terms of θ_1, φ_1, θ_2, φ_2 is well known to be

$$\cos\psi = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos(\varphi_1 - \varphi_2) \qquad (168.22)$$

By applying the orthogonality relations (155.10) and (168.12) for the functions $e^{im\varphi_1}$ and $P_n^m(\cos\theta)$, we find for the expansion coefficients the result

$$2\pi\frac{2}{2n+1}\frac{(n+|m|)!}{(n-|m|)!}f_{nm}$$
$$= \int_0^\pi d\theta_1\int_0^{2\pi}\frac{\cos\theta_1 P_n^{|m|}(\cos\theta_1)e^{-im\varphi_1}\sin\theta_1\,d\varphi_1}{\sqrt{r_1^2 - 2r_1r_2\cos\psi + r_2^2}}$$

In order to effect the indicated integrations, it will be assumed that $r_2 < r_1$, so that

$$\sqrt{r_1^2 - 2r_1r_2\cos\psi + r_2^2} = r_1\sqrt{1 - 2t\cos\psi + t^2}$$

where $t = r_2/r_1 < 1$. From the generating function for $P_n(\cos\omega)$, eq. (167.1), the above square root can be expanded as follows.

$$\frac{1}{\sqrt{r_1^2 - 2r_1r_2\cos\psi + r_2^2}} = \frac{1}{r_1}\sum_{\nu=0}^{\infty}P_\nu(\cos\psi)t^\nu$$

so that

$$f_{nm} = \frac{1}{2\pi}\frac{2n+1}{2}\frac{(n-m)!}{(n+m)!}\frac{1}{r_1}$$
$$\cdot\int_0^\pi d\theta_1\int_0^{2\pi}\sum_{\nu=0}^{\infty}P_\nu(\cos\psi)t^\nu P_n^{|m|}(\cos\theta_1)e^{-im\varphi_1}\cos\theta_1\sin\theta_1\,d\varphi_1 \qquad (168.23)$$

The integrations still cannot be carried out directly because P_ν is a complicated function of θ_1, θ_2, φ_1, and φ_2. The application of the addition formula (168.19) allows us to integrate over separate functions of θ_1, θ_2, φ_1, and φ_2 alone. To do this it should be noted that if in (168.18) the following substitutions are made, i.e., $z = \cos\omega$, $x = \cos\theta_1$, $y = \cos\theta_2$, and $\cos\omega = \cos(\varphi_1 - \varphi_2)$ just the needed expression (168.22) is obtained, whence the addition formula (168.20) becomes

$$P_\nu(\cos\psi) = \sum_{\mu=-\nu}^{+\nu}\frac{(\nu-|\mu|)!}{(\nu+|\mu|)!}P_\nu^{|\mu|}(\cos\theta_1)P_\nu^{|\mu|}(\cos\theta_2)e^{i\mu(\varphi_1-\varphi_2)}$$

Replacing $P_\nu(\cos\psi)$ by the right-hand side we note that if the order of summation and integration are interchanged in (168.23), we have the following integrals to deal with.

$$\int_0^{2\pi} e^{i\mu\varphi_1 - im\varphi_1} d\varphi_1 = \begin{matrix} 0 & \mu \neq m \\ 2\pi & \mu = m \end{matrix}$$

$$\int_0^{\pi} P_n^{|m|}(\cos\theta_1) P_\nu^{|m|}(\cos\theta_1) \cos\theta_1 \sin\theta_1 \, d\theta_1$$

$$= \int_{-1}^{+1} P_n^{|m|}(x) P_\nu^{|m|}(x) x \, dx$$

$$= \begin{cases} \dfrac{2}{2n+1} \dfrac{1}{2n-1} \dfrac{(n+|m|)!}{(n-|m|)!} & \nu = n-1 \\ \dfrac{2}{2n+1} \dfrac{1}{2n+3} \dfrac{(n+|m|+1)!}{(n-|m|)!} & \nu = n+1 \end{cases}$$

In evaluating the last integral, the recurrence relation (168.5) has been used, together with (168.12). Consequently

$$f_{nm} = \frac{1}{r_1} \frac{(n-|m|)!}{(n+|m|)!} \left[\frac{n+|m|}{2n-1} \left(\frac{r_2}{r_1} \right)^{n-1} P_{n-1}^{|m|}(\cos\theta_2) + \frac{(n+1-|m|)!}{2n+3} \left(\frac{r_2}{r_1} \right)^{n+1} P_{n+1}^{|m|}(\cos\theta_2) \right] e^{-im\varphi_1}$$

and the expansion coefficients for the proposed problem have been successfully calculated with the help of the addition formula for the Legendre polynomials.

169. Bessel functions and Hankel functions. Widely used functions in physics and engineering are the Bessel functions, which often arise in problems involving axial symmetry. They also can be used as the elements of a complete orthonormal set of functions, though their use as such occurs in a somewhat different manner than the orthonormal functions considered previously in the chapter, as will be pointed out later in this section.

The generating function $g(x,t)$ for the Bessel functions $J_n(x)$ is

$$g(x,t) = e^{(x/2)(t-1/t)} = \sum_{n=-\infty}^{+\infty} J_n(x) t^n \tag{169.1}$$

where the region of definition of $J_n(x)$ is $0 \leq x < \infty$. Since this is a Laurent series (see Sec. 80), one cannot obtain the coefficients $J_n(x)$ by a simple differentiation procedure, and they are consequently not simple polynomials. They can be obtained by the usual means of determining coefficients in a Laurent series (see Sec. 91) or by actually expanding $g(x,t)$ in the following way.

$$e^{(x/2)(t-1/t)} = e^{xt/2} \cdot e^{-x/2t} = \sum_{r=0}^{\infty} \left(\frac{x}{2} \right)^r \frac{t^r}{r!} \cdot \sum_{s=0}^{\infty} \left(\frac{-x}{2} \right)^s \frac{t^{-s}}{s!} \tag{169.2}$$

For any given value of s the term in the above product yielding t^n is

$$\left(\frac{x}{2} \right)^{n+s} \frac{1}{(n+s)!} \left(\frac{-x}{2} \right)^s \frac{t^{n+s-s}}{s!}$$

Consequently the coefficient of t^n in (169.2) will be found by summing the above terms over all s. Thus

$$J_n(x) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s!(n+s)!}\left(\frac{x}{2}\right)^{n+2s} \tag{169.3}$$

If in a similar way we had selected the coefficient of t^{-n} in (169.2), we would have found

$$J_{-n}(x) = \sum_{s=0}^{\infty} \frac{(-1)^{n+s}}{s!(n+s)!}\left(\frac{x}{2}\right)^{n+2s} \tag{169.4}$$

so that

$$J_{-n}(x) = (-1)^n J_n(x) \tag{169.5}$$

Graphs of the Bessel functions of the first few orders are shown in Fig. 169.1.

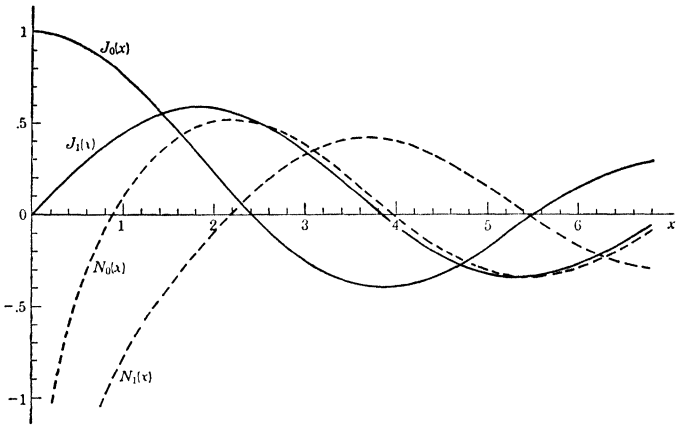


FIG. 169.1

Recurrence Formulas. Very useful recurrence formulas between Bessel functions of the above type can be obtained from the generating function (169.1) by differentiating $g(x,t)$ with respect to t. From (169.1),

$$\frac{\partial}{\partial t} g(x,t) = \frac{1}{2} x \left(1 + \frac{1}{t^2}\right) g(x,t) = \sum_{n=-\infty}^{+\infty} n J_n(x) t^{n-1}$$

$$\frac{1}{2} x \left(1 + \frac{1}{t^2}\right) \sum_{n=-\infty}^{+\infty} J_n(\nu) t^n = \sum_{n=-\infty}^{+\infty} n J_n(x) t^{n-1}$$

Comparing like coefficients of t on the right and left sides, it is evident that

$$J_{n-1}(x) + J_{n+1}(x) = \frac{2n}{x} J_n(x) \tag{169.6}$$

Proceeding as above, except first carrying out a differentiation with respect to x, the next formula

$$J_{n-1}(x) - J_{n+1}(x) = 2J_n'(x) \tag{169.7}$$

can be established. For $n = 0$ this reduces to the very simple result

$$J_0'(x) = -J_1(x) \tag{169.8}$$

If (169.6) and (169.7) be added and subtracted, two other useful formulas are obtained. They can be written in the form

$$\frac{d}{dx}[x^n J_n(x)] = x^n J_{n-1}(x) \tag{169.9}$$

$$\frac{d}{dx}[x^{-n} J_n(x)] = -x^{-n} J_{n+1}(x) \tag{169.10}$$

By differentiation and combination of the above relations it may be shown that $J_n(x)$ satisfies the differential equation

$$\frac{d^2 J_n}{dx^2} + \frac{1}{x}\frac{dJ_n}{dx} + \left(1 - \frac{n^2}{x^2}\right)J_n = 0 \tag{169.11}$$

Useful Integral Representations of $J_n(x)$. An integral representation of $J_n(x)$ can be obtained simply by determining the coefficient of the Laurent expansion of the generating function $g(x,t)$. In accordance with the procedure of Sec. 91, the coefficient $J_n(x)$ in (169.1) is

$$J_n(x) = \frac{1}{2\pi i}\int_C \frac{g(x,t)}{t^{n+1}}\,dt = \frac{1}{2\pi i}\int_C t^{-n-1} e^{(x/2)(t-1/t)} dt \tag{169.12}$$

where t is regarded as a complex variable, and the contour C encloses the point $t = 0$. A well-known definite integral representation of $J_n(x)$ can be obtained by taking the contour C to be a circle in the t plane with unit radius. On this circle $t = e^{i\theta}$, so that

$$J_n(x) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} e^{-in\theta + ix\sin\theta} d\theta \tag{169.13}$$

This can readily be written

$$J_n(x) = \frac{1}{\pi}\int_0^{\pi} \cos(n\theta - x\sin\theta) d\theta \tag{169.14}$$

Another useful integral representation can be obtained by a transformation of (169.12). On making a transformation of complex variable defined by $t = 2z/x$, eq. (169.12) becomes

$$J_n(x) = \frac{1}{2\pi i}\left(\frac{x}{2}\right)^n \int_{C'} z^{-n-1} e^{z - x^2/4z} dz \tag{169.15}$$

If n is an integer, the path C must enclose the point $z = 0$. This integral however, is often used as the definition of the Bessel function when n is noninteger. In this case the path C must be chosen somewhat differ-

ently, since z^{-n} does not return to its initial value on a path once around the point $z = 0$. However, if the path C is deformed into C' shown in Fig. 169.2, the integrand will have the same value at the beginning and end of the path after going around the point $z = 0$, and (169.15) will constitute a definition of $J_n(x)$ satisfying the differential eq. (169.11) when n is not an integer.

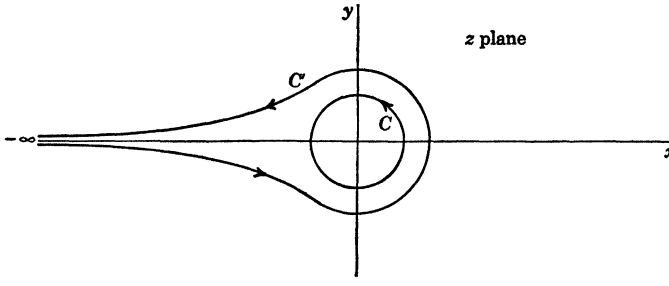


FIG. 169.2

Bessel Functions of Half Odd Integer Order. In certain applications one encounters Bessel functions of the form $J_{n+1/2}(x)$. Such functions can always be expressed in the form

$$J_{n+1/2}(x) = P_n \sin x + Q_n \cos x \tag{169.16}$$

where P_n and Q_n are polynomials in $x^{-1/2}$. Formulas for $J_{n+1/2}(x)$ can be built up by first computing $J_{1/2}(x)$ by means of (169.15). Then

$$J_{1/2}(x) = \frac{1}{2\pi i}\left(\frac{x}{2}\right)^{1/2}\int_{-\infty}^{(0+)} z^{-1/2-1}e^{z}\cdot e^{-x^2/4z}dz$$

On expanding $e^{-x^2/4z}$ and integrating term by term it is found that

$$J_{1/2}(x) = \left(\frac{x}{2}\right)^{1/2}\left[I_0 - \frac{1}{4}I_1x^2 + \frac{1}{4^2 2!}I_2x^4 - \ldots + \frac{(-1)^k}{4^k k!}I_kx^{2k} + \ldots\right]$$

where

$$I_k = \frac{1}{2\pi i}\int_{-\infty}^{(0+)} z^{-k-1/2-1}e^{z}dz$$

If the complex variable z is replaced by $-z$ in the integral, the integral is substantially the gamma function. From formula (102.3),

$$I_k = -\frac{\sin \pi(k+\frac{1}{2})}{\pi}\,\Gamma\left(-k-\frac{1}{2}\right)$$

Using the relations (109.6a) and (109.9) between gamma functions, I_k may be expressed as

$$I_k = \frac{1}{(k+\frac{1}{2})\Gamma(k+\frac{1}{2})} = \frac{2^{2k-1}\Gamma(k)}{\sqrt{\pi}(k+\frac{1}{2})\Gamma(2k)}$$

Remembering that $z\Gamma(z) = \Gamma(z+1)$, and replacing I_k in the expression for $J_{1/2}(x)$ by the above formula, it can be readily verified that

$$J_{1/2}(x) = \left(\frac{2}{\pi x}\right)^{1/2} \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!} = \left(\frac{2}{\pi x}\right)^{1/2} \sin x \qquad (169.17)$$

In order to express $J_{n+1/2}(x)$ in terms of $J_{1/2}(x)$, one may start with the recurrence relation (169.10), which may conveniently be written

$$x^{-n-1}J_{n+1}(x) = \frac{-1}{x}\frac{d}{dx}x^{-n}J_n(x)$$

Operating on this equation with the operator $\frac{1}{x}\frac{d}{dx}$ and applying the above relation again, and repeating this r times, gives

$$x^{-n-r}J_{n+r}(x) = (-1)^r\left(\frac{1}{x}\frac{d}{dx}\right)^r[x^{-n}J_n(x)]$$

Since the recurrence formulas hold for all n, we may replace n by $\frac{1}{2}$ and obtain

$$\begin{aligned} J_{r+1/2}(x) &= (-1)^r x^{+r+1/2}\left(\frac{1}{x}\frac{d}{dx}\right)^r[x^{-1/2}J_{1/2}(x)] \\ &= (-1)^r x^{r+1/2}\left(\frac{1}{x}\frac{d}{dx}\right)^r\left(\frac{\sin x}{x}\right) \end{aligned} \qquad (169.18)$$

wherein $\left(\frac{1}{x}\frac{d}{dx}\right)^r$ means that this operation is to be carried out r successive times. This will clearly reduce to the form (169.16).

Other Solutions of Bessel's Differential Equation. Several other solutions of Bessel's eq. (169.11) are often necessary in the solution of physical problems. A solution usually denoted by $N_n(x)$ which, together with $J_n(x)$, form a fundamental set of solutions in the sense that

$$Z_n(x) = c_1J_n(x) + c_2N_n(x) \qquad (169.19)$$

is a general solution of Bessel's equation. Here $N_n(x)$ is often called a Bessel function of the second kind, and is defined as

$$N_n(x) = \frac{J_n(x)\cos n\pi - J_{-n}(x)}{\sin n\pi} \qquad (169.20)$$

if n is noninteger and is defined as the limit of the expression on the right as n approaches the integer n in question for integer values of n. When n is an integer, $N_n(x)$ may be written

$$\begin{aligned} \pi N_n(x) = 2\left(\gamma + \log\frac{x}{2}\right)J_n(x) - \left(\frac{x}{2}\right)^{-n}\sum_{m=0}^{n}\frac{(n-m-1)!}{m!}\left(\frac{x}{2}\right)^{2m} \\ - \sum_{m=0}^{\infty}\frac{(-1)^m(x/2)^{n+2m}}{m!(n+m)!}\left(\frac{1}{1}+\frac{1}{2}+\ldots+\frac{1}{m}+\frac{1}{1}+\frac{1}{2}+\ldots+\frac{1}{n+m}\right) \end{aligned} \qquad (169.21)$$

where γ is Euler's constant [see eq. (109.3a)].

All Bessel functions of the second kind become infinite at $x = 0$ as shown in Fig. 169.1, so that in physical problems involving Bessel functions at $x = 0$ the functions of the second kind are ruled out by the boundary conditions or initial conditions at $x = 0$. However, in heat-flow problems and potential problems and the like which hold outside a cylinder of finite radius, the functions $N_n(x)$ are required as well as $J_n(x)$. Examples of this are given later in this section.

The two linearly independent solutions of Bessel's equation considered above are analogous to, and play the same roll as, the trigonometric functions $\sin nx$ and $\cos nx$, which form a fundamental set of solutions of the differential equation

$$\frac{d^2u}{dx^2} + n^2u = 0$$

However, there is another important fundamental set of solutions of this equation, namely, e^{inx} and e^{-inx}. These solutions are necessary in problems involving running waves. In an analogous way there are two other linearly independent solutions of Bessel's equation which become very convenient in handling certain radiation problems. These solutions are called Hankel functions, and are denoted by $H_n^{(1)}(x)$ and $H_n^{(2)}(x)$. Thus a general solution of Bessel's equation can also be written

$$c_1H_n^{(1)}(x) + c_2H_n^{(2)}(x)$$

The analogy between $J_n(x)$ and $N_n(x)$ and the trigonometric functions $\cos nx$ and $\sin nx$ and between $H_n^{(1)}(x)$ and $H_n^{(2)}(x)$ and the exponentials e^{inx} and e^{-inx} is all the more meaningful when the asymptotic behavior of the various types of Bessel functions is noted (see Sec. 110). For large value of x,

$$\left.\begin{aligned} J_n(x) &\approx \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{n\pi}{2} - \frac{\pi}{4}\right) \\ N_n(x) &\approx \sqrt{\frac{2}{\pi x}} \sin\left(x - \frac{n\pi}{2} - \frac{\pi}{4}\right) \\ H_n^{(1)}(x) &\approx \sqrt{\frac{2}{\pi x}}\, e^{i\left(x - \frac{n\pi}{2} - \frac{\pi}{4}\right)} \\ H_n^{(2)}(x) &\approx \sqrt{\frac{2}{\pi x}}\, e^{-i\left(x - \frac{n\pi}{2} - \frac{\pi}{4}\right)} \end{aligned}\right\} \quad (169.22)$$

It is often this asymptotic behavior which determines the choice of the type of Bessel function to be used in a given physical problem.

The above type of Bessel functions are connected by the following relations.

$$J_n(x) = \frac{H_n^{(1)}(x) + H_n^{(2)}(x)}{2} = \frac{N_{-n}(x) - N_n(x)\cos n\pi}{\sin n\pi} \quad (169.23)$$

$$N_n(x) = \frac{J_n(x)\cos n\pi - J_{-n}(x)}{\sin n\pi} = \frac{H_n^{(1)}(x) - H_n^{(2)}(x)}{2i} \tag{169.24}$$

$$H_n^{(1)}(x) = \frac{J_{-n}(x) - e^{-n\pi i}J_n(x)}{i\sin n\pi} = \frac{N_{-n}(x) - e^{-n\pi i}N_n(x)}{\sin n\pi} \tag{169.25}$$

$$H_n^{(2)}(x) = \frac{e^{n\pi i}J_n(x) - J_{-n}(x)}{i\sin n\pi} = \frac{N_{-n}(x) - e^{+n\pi i}N_n(x)}{\sin n\pi} \tag{169.26}$$

When n is an integer it must be understood that the limit of these expressions is meant as n approaches the integer in question. All of the Bessel functions satisfy the recurrence relations given under (169.6), (169.7), (169.9), and (169.10), and of course all satisfy eq. (169.11).

Orthogonal Character of the Bessel Functions. An orthonormal set of functions over the interval $0 \leq x \leq 1$ can be formed from the Bessel function $J_n(x)$. However, in this case the set is not enumerated by changing the index n as has been the case in the preceding examples. To form the orthonormal set in this case let $\alpha_{n1}, \alpha_{n2}, \ldots, \alpha_{nk}, \ldots$ be the roots of $J_n(x)$ arranged in order of increasing magnitude, i.e., $\alpha_{n1} < \alpha_{n2} < \ldots < \alpha_{nk}, \ldots$, and consider the set of functions

$$\sqrt{x}J_n(\alpha_{n1}x),\ \sqrt{x}J_n(\alpha_{n2}x),\ \ldots,\ \sqrt{x}J_n(\alpha_{nk}x),\ \ldots$$

These functions form a complete orthogonal set over the interval $0 < x < 1$. The complete orthonormal set is

$$\frac{\sqrt{2x}J_n(\alpha_{n1}x)}{J_{n+1}(\alpha_n)},\ \frac{\sqrt{2x}J_n(\alpha_{n2}x)}{J_{n+1}(\alpha_{n2})},\ \ldots,\ \frac{\sqrt{2x}J_n(\alpha_{nk}x)}{J_{n+1}(\alpha_{nk})},\ \ldots$$

It can readily be shown that

$$\int_0^1 \varphi_{nk}(x)\varphi_{nl}(x)dx = \begin{matrix} 0 & k \neq l \\ 1 & k = l \end{matrix} \tag{169.27}$$

where

$$\varphi_{nk} = \frac{\sqrt{2x}J_n(\alpha_{nk}x)}{J_{n+1}(\alpha_{nk})} \tag{169.28}$$

It is thus seen that an orthonormal set can be formed from Bessel functions of all orders. The one most often used in practice is formed from the Bessel's function $J_0(x)$ of zero order.

Example of the Use of the Orthonormal Properties of the Bessel's Functions. The general solution of the wave equation

$$a^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = \frac{\partial^2 u}{\partial t^2} \tag{169.29}$$

applied to the case where u is the displacement of an elastic circular membrane is

$$u = \sum_{k=1}^{\infty}\sum_{n=0}^{\infty}[(A_{nk}\cos\mu_{nk}at + B_{nk}\sin\mu_{nk}at)\cos n\theta + (C_{nk}\cos\mu_{nk}at + D_{nk}\sin\mu_{nk}at)\sin n\theta]J_n(\mu_{nk}r) \tag{169.30}$$

The constants μ_{nk} are as yet undetermined constants as are also A_{nk}, B_{nk}, C_{nk}, and D_{nk}, while r, θ, t are the polar coordinates which locate a point on the circular membrane, and the time, respectively. The constants must be determined from the boundary conditions and initial conditions imposed on the membrane. To see how they are determined, let us suppose that on a circle $r = R$ the membrane is fastened so that on this circle $u = 0$ for all t. This condition can be met if $\mu_{nk}R = \alpha_{nk}$ where α_{nk} is the kth root of the Bessel function $J_n(x)$. Hence $\mu_{nk} = \alpha_{nk/R}$ is determined. Let us further suppose that at $t = 0$, the displacement of the membrane is given, and indeed is specified by

$$u_{t=0} = f(r,\theta) \tag{169.31}$$

membrane is zero, i.e.,

$$\left.\frac{\partial u}{\partial t}\right]_{t=0} = 0 \tag{169.32}$$

On differentiating (169.30) with respect to t, it is obvious that the condition (169.32) requires that $B_{nk} = D_{nk} = 0$. Consequently at $t = 0$ we require that

$$f(r,\theta) = \sum_{k=1}^{\infty}\sum_{n=0}^{\infty}(A_{nk}\cos n\theta + C_{nk}\sin n\theta)J_n\left(\alpha_{nk}\frac{r}{R}\right) \tag{169.33}$$

To determine the constants A_{nk}, we may first use the orthonormal property of the trigonometric functions as used in Sec. 163 and multiply the above equation through by $\cos n\theta$, and integrate from 0 to 2π. This gives

$$\int_0^{2\pi} f(r,\theta)\cos n\theta\, d\theta = \pi\sum_{k=1}^{\infty} A_{nk}J_n\left(\alpha_{nk}\frac{r}{R}\right)$$

Next the orthogonal property of the Bessel functions as given in eq. (169.27) can be used by multiplying the above equation through by $rJ_n[\alpha_{nk}(r/R)]$, and integrating over r from 0 to R. Since according to (169.27) and (169.28),

$$\int_0^R rJ_n\left(\alpha_{nk}\frac{r}{R}\right)J_n\left(\alpha_{nl}\frac{r}{R}\right)dr = R^2\int_0^1 xJ_n(\alpha_{nk}x)J_n(\alpha_{nl}x)dx$$

$$= \begin{cases} 0 & k \neq l \\ \dfrac{R^2}{2}[J_{n+1}(\alpha_{nk})]^2 & k = l \end{cases}$$

we finally have

$$A_{nk} = \frac{2}{\pi R^2[J_{n+1}(\alpha_{nk})]^2}\int_0^R\int_0^{2\pi} rf(r,\theta)J_n\left(\alpha_{nk}\frac{r}{R}\right)\cos n\theta\, d\theta\, dr \tag{169.34}$$

The coefficients C_{nk} can be determined in a similar way by first multiplying (169.33) through by $\sin n\theta$ and proceeding as just outlined.

$$T_0 = \sum_{n=1}^{\infty} C_n H_0^{(1)}\left(i\frac{n\pi}{l}a\right)\sin\left(\frac{n\pi}{l}z\right) \tag{169.39}$$

Using the orthogonality properties of the trigonometric functions as in the preceding example, we readily find that

$$C_n = \frac{2}{lH_0^{(1)}[i(n\pi/l)a]}\int_0^l T_0 \sin\left(\frac{n\pi}{l}z\right)dz \tag{169.40}$$

By differentiating (169.38) with respect to r, it can be verified that heat flows in the direction of increasing r as it should, and never in the opposite direction.

Addition Formulas for the Bessel Function. Just as in the case of the Legendre polynomials [see eq. (168.19)] there are useful addition formulas for Bessel functions. In particular

$$J_n(x + y) = \sum_{k=-\infty}^{\infty} J_k(x)J_{n-k}(y) \tag{169.41}$$

This allows one to express a Bessel function of the combination variable $x+y$ in terms of Bessel functions of x alone and y alone.

Another useful addition formula which applies to J_0 is

$$J_0(R) = J_0(x)J_0(y) + 2\sum_{n=1}^{\infty} J_n(x)J_n(y)\cos n\psi \tag{169.42}$$

where $R = \sqrt{x^2 + y^2 - 2xy\cos\psi}$. This allows one to express the Bessel function whose argument is the distance between the two points whose polar coordinates are r_1, θ_1, φ_1, and r_2, θ_2, φ_2, respectively, in terms of higher order Bessel's functions of the radii r_1 and r_2 alone and the angle ψ between the radii to the two points in question.

A useful related formula is

$$\frac{\sin Rs}{Rs} = \pi\sum_{n=0}^{\infty}\left(n + \frac{1}{2}\right)\frac{J_{n+1/2}(sx)}{\sqrt{sx}}\frac{J_{n+1/2}(sy)}{\sqrt{sy}}P_n(\cos\psi) \tag{169.43}$$

where R, x, y, $\cos\psi$ have the same significance as above. By integrating the above equation with respect to s from 0 to ∞, a useful separated expansion for $1/R$ is obtained, namely,

$$\frac{1}{R} = \sum_{n=0}^{\infty}(2n + 1)P_n(\cos\psi)\int_0^{\infty}\frac{J_{n+1/2}(sx)}{\sqrt{sx}}\frac{J_{n+1/2}(sy)}{\sqrt{sy}}ds \tag{169.44}$$

This formula can be used to advantage in evaluating integrals of products of functions of x and y and $1/R$, since $1/R$ is then expressed as products of functions of x and y alone.

SUGGESTED REFERENCES

R. Courant and D. Hilbert, *Methoden der mathematischen Physik*, I., Berlin, Julius Springer, 1931.

M. H. Stone, *Linear Transformations in Hilbert Space*, New York, American Mathematical Society, 1932.

E. T. Whittaker and G. N. Watson, *Modern Analysis*, London, Cambridge University Press, 1920.

G. N. Watson, *Theory of Bessel Functions*, London, Cambridge University Press, 1922.

W. E. Byerly, *Fourier Series and Spherical, Cylindrical, and Ellipsoidal Harmonics*, Boston, Ginn & Company, 1893.

E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics*, London, Cambridge University Press, 1931.

A. Gray and G. B. Mathews, *A Treatise on Bessel Functions*, New York, The Macmillan Company, 1895.

PROBLEMS

157.1. Form a normal orthogonal set from the functions $1, x, x^2, x^3, \ldots$ on the interval 0 to 1.

157.2. Form a normal orthogonal set from the functions $e^{-x}, e^{-2x}, e^{-3x}, e^{-4x}$ on the interval 0 to ∞.

Expand the following functions in series of sines and cosines.

159.1. $f(x) = -1, \quad 0 < x < \pi. \qquad f(x) = +1, \quad \pi < x < 2\pi.$

159.2. $f(x) = x, \qquad 0 \leq x \leq \pi/2.$

$f(x) = \pi - x, \qquad \pi/2 \leq x \leq 3\pi/2.$

$f(x) = x - 2\pi, \qquad 3\pi/2 \leq x \leq 2\pi.$

Expand the following functions in Fourier series.

159.3. $f(x) = |\sin x|.$

159.4. $f(x) = \sin x, \quad 0 \leq x \leq \pi. \qquad f(x) = 0, \quad -\pi \leq x \leq 0.$

159.5. $f(x) = 1 - a(x + 1), \; -\pi < x < -\pi/2, \; 0 < a \ll 1/\pi.$

$f(x) = \cos x$ or $1 - a(x + 1)$, whichever is larger for $-\pi/2 < x < 0.$

where $f(x)$ repeats in interval $0 < x < \pi$. This is a crude representation of the output of a full-wave rectifier and capacitance filter.

159.6. In an amplifier which is being overdriven the wave form is given by $f(x) = \sin x$ or by $f(x) = a$, for $0 < a < 1$, whichever is less. Find the amplitudes of the second and third harmonics relative to that of the first.

159.7. Expand $r \sin x/(1 - 2r \cos x + r^2)$ in a Fourier series.

160.1. Expand the function e^{-x} in Fourier series in the interval 0 to 21.

160.2. Expand the function $f(x) = (x^2 - 1)^2$ in Fourier series on the interval -1 to 1.

161.1. Estimate the number of terms required in order that the Fourier representation of the function

$$f(x) = \cos^{2n} x \text{ for } -\pi/2 \leq x \leq \pi/2.$$

$$f(x) = 0 \text{ for } -\pi \leq x \leq -\pi/2 \text{ and } \pi/2 \leq x \leq \pi.$$

will have a mean square deviation from $f(x)$ less than ϵ, where n is a positive integer.

163.1. The torque in arbitrary units on the crankshaft of a compressor is given in the following table, for every 15 degrees of rotation of the shaft. Determine numerically the first four harmonics in a harmonic analysis of the data.

θ in degrees	Torque	θ in degrees	Torque
0	11,940	195	19,715
15	17,365	210	18,000
30	23,810	225	16,740
45	22,460	240	14,210
60	20,440	255	11,470
75	18,250	270	11,940
90	18,240	285	17,715
105	19,365	300	22,450
120	19,360	315	21,920
135	17,280	330	18,660
150	15,990	345	14,030
165	15,690	360	11,940
180	18,240		

164.1. Prove from the recursion formula and the first two Hermitian polynomials

$$H_n(x) = \sum_{l=0}^{n} (-2)^l (2x)^{n-2l} \binom{n}{2l} (1 \cdot 3 \cdot 5 \cdot \ldots \cdot 2l - 1)$$

$$= (2x)^n - 2\frac{n!}{(n-2)!2!}(2x)^{n-2} + 4\frac{n!}{(n-4)!4!}(2x)^{n-4}(1 \cdot 3) \ldots .$$

164.2. Prove by expansion in a series of Hermitian polynomials that

$$H_m(x + a) = \sum_{l=0}^{m} \binom{m}{l} (2a)^{m-l} H_l(x)$$

164.3. Evaluate the following integrals, using the generating function for $H_n(x)$.

$$\int_{-\infty}^{\infty} e^{-x^2} H_n(x) H_m(x) x \, dx, \quad \int_{-\infty}^{\infty} e^{-x^2} H_n(x) H_m(x) x^2 dx.$$

165.1. By the use of the generating function, derive the orthogonality relation for Laguerre polynomials.

165.2. Evaluate $\int_0^{\infty} e^{-x} L_n(x) L_m(x) x \, dx$.

165.3. Evaluate $\int_0^{\infty} e^{-\lambda x} L_n(x) dx$.

165.4. Express $L_n(x + a)$ in terms of $L_n(x)$.

166.1. Derive the orthogonality relation for associated Laguerre polynomials.

166.2. Express $L_{n+k}^k(x)$ in terms of $L_m(x)$.

166.3. Evaluate $\int_0^{\infty} x^l e^{-x} L_n^{\alpha}(x) L_m^{\beta}(x) dx$ where $l \geq \alpha$, $l \geq \beta$.

166.4. Express $L_{n+k}^k(ax)$ in terms of $L_{m+l}^l(x)$.

167.1. By means of the recurrence relations prove that

$$P_n(x) = \frac{(2n)!}{2^n n! n!} \left[x^n - \frac{n(n-1)}{2(2n-1)} x^{n-2} + \frac{n(n-1)(n-2)(n-3)}{2 \cdot 4 \cdot (2n-1)(2n-3)} x^{n-4} - \ldots \right].$$

167.2. Show that $P_n(1) = 1$, $P_n(-1) = (-1)^n$,

$$P_{2n+1}(0) = 0, \qquad P_{2n}(0) = (-1)^n \frac{(2n)!}{2^{2n} n! n!}.$$

167.3. Show that $\int_{-1}^{1} x^k P_n(x)dx = 0$, $k < n$.

167.4. Expand in a series of Legendre polynomials the function

$$f(x) = \begin{cases} -1, & -1 < x < 0. \\ +1, & 0 < x < 1. \end{cases}$$

167.5. The charge density on the surface of a sphere of radius R is given by $K \cos\theta$, where θ is the polar angle. Find the potential at points inside and outside the sphere.

168.1. By using the particular solution of Poisson's equation

$$\varphi(r) = \frac{1}{4\pi}\int_V \frac{\rho(r')}{r - r'}dV'$$

find the potential due to the charge distribution

$$\rho = \frac{1}{64\pi a^5} r^2 e^{-r/a} \sin^2\theta$$

in spherical coordinates r, θ, φ.

168.2. By using $$A(r) = \frac{1}{4\pi c}\int_V \frac{J(r')}{|r - r'|}dV'$$

find the vector potential due to the current distribution

$$J = J_\varphi = \frac{\hbar}{m}\frac{1}{64\pi a^5} e^{-r/a} r \sin\theta.$$

The two problems above represent a hydrogen atom in a $2p$ state.

169.1. Separate the wave equation

$$\nabla^2 u - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0$$

in spherical coordinates and find the solutions appropriate to the exterior of a sphere of radius a, the values of u being given on the surface of the sphere. Separate outgoing and incoming waves.

169.2. Show that $e^{ik\cdot z}$ is a solution of the wave equation $\nabla^2 u + k^2 u = 0$ in rectangular coordinates. By finding the solution in spherical coordinates which has the same behavior at the origin and at infinity, show that

$$e^{ikr\cos\theta} = \sum_{n=0}^{\infty} (2n + 1) i^n \left(\frac{\pi}{2kr}\right)^{1/2} J_{n+1/2}(kr).$$

169.3. If V is the velocity potential for a sound wave, V satisfies the wave equation. By combining the expression above with the expressions for outgoing waves, find the velocity potential for an incident plane wave which is scattered by a hard sphere of radius a.

169.4. The velocity potential of a perfect fluid satisfies $\nabla^2\Phi = 0$. Find the velocity potential for flow around a sphere of radius a when the velocity at infinity is uniform and in the z direction.

169.5. Separate the wave equation

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} p = 0$$

in cylindrical coordinates and find the solution which represents an outgoing wave and which takes the value $p_0 e^{-i\omega t}\cos\theta$ at the surface of a cylinder of radius a. If p is the sound pressure, show that this represents the radiation of sound from a vibrating wire. Find the corresponding velocity potential.

169.6. Show that $J_n(z) = \sum_{r=0}^{\infty} (-1)^r \frac{(\frac{1}{2}z)^{n+2r}}{r!(n + r)!}$.

Chapter 13

ORTHONORMAL FUNCTIONS WITH A CONTINUOUS SPECTRUM

170. The expansion problem in the continuous spectrum. In Sec. 149 the representation of a function in terms of an orthonormal set of functions in the form $f(x) = \sum_{\nu=0}^{\infty} a_\nu \varphi_\nu(x)$ was discussed, where $\varphi_\nu(x)$, $\nu = 0, 1, 2, \ldots$ was in most cases a countably infinite set of functions such as the trigonometric functions which form the basis for the usual Fourier series. In order to represent a certain class of functions, however, it is necessary to call upon sets of functions which are no longer countably infinite in number, but form a dense set in the sense that the designating parameter ν takes on all values over a continuous range of real numbers. One of the simplest examples of such a set of functions is e^{iux}, where u takes on all real numbers in the interval $-\infty < u < +\infty$ in contrast to the orthonormal set of functions $e^{\nu x}/\sqrt{2\pi}$ where $\nu = 0, \pm 1, \pm 2, \ldots$. Both functions are solutions of the same differential equation

$$\frac{d^2\varphi}{dx^2} + w^2\varphi = 0$$

This equation, for instance, applies to the lateral vibrations of an elastic string. The function φ gives the displacement of the string as a function of the distance along it. If the string is held fixed at two points, these boundary conditions can be satisfied only by certain discrete values of w. These solutions form an orthonormal set of the kind typified by $e^{ivx}/\sqrt{2\pi}$. If, on the other hand, the string is infinite in length, no such restriction can be placed on w, i.e., all real values of w would be allowed, or as we say, we have a set of solutions e^{iux} over a continuous spectrum of frequencies u. Under these circumstances it would be natural to replace the representation

$$f(x) = \sum_{\nu=-\infty}^{+\infty} a_\nu \frac{e^{-ivx}}{\sqrt{2\pi}}$$

by the representation

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} A(u) e^{-iux} du \tag{170.1}$$

This is called the Fourier integral representation of a function $f(x)$. In general if we may write

$$f(x) = \int F(u) \varphi(u,x) du \tag{170.2}$$

where the integration extends over an appropriate region of u we shall speak of the representation of $f(x)$ in terms of the continuous spectrum functions $\varphi(u,x)$. In the next sections we shall discuss some important examples of such functions and discuss the sense in which they have orthonormal properties.

171. Fourier integrals and transforms. Probably the most widely used set of functions possessing a continuous spectrum are the exponential functions e^{-iux}, $-\infty < u < \infty$. If $\int_{-\infty}^{+\infty} |f(x)|^2 dx$ exists, $f(x)$ possesses a Fourier integral representation and

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} F(u) e^{-iux} du \tag{171.1}$$

The function $F(u)$ is called the Fourier transform of $f(x)$. Knowing $F(u)$ is equivalent to knowing $f(x)$. Here $F(u)$ is a measure of the spectral distribution necessary to represent the function $f(x)$. Thus in electrical practice, if a potential difference which had a time dependence given by $f(t)$ is impressed on a frequency analyzing network, the frequency analysis would yield $F(u)$. For electric network theory, a knowledge of Fourier transforms is very important.

Drawing upon the similarity of the Fourier integral representation and the representation in terms of an orthonormal set, we would expect to be able to calculate $F(u)$ by means of the formula

$$F(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) e^{iux} dx \tag{171.2}$$

This is indeed the correct Fourier transform of $f(x)$, provided $f(x)$ is piecewise continuous and $\int_{-\infty}^{+\infty} |f(x)|^2 dx$ exists. This would indicate that there must be some property of the functions e^{iux}, $-\infty < u < \infty$ akin to orthogonality over the interval $-\infty < x < \infty$. That it cannot be the same as that possessed by the functions $e^{i\nu x}$, $\nu = 0, \pm 1, \pm 2, \ldots$ for $-\pi < x < \pi$ (see Sec. 155) is readily seen from the fact that

$$\int_{-\infty}^{+\infty} e^{iux} \cdot e^{-iu'x} dx \tag{171.3}$$

does not exist.

The formal concept of orthogonality with regard to functions with a

continuous spectrum of parameters is so important from the standpoint of modern quantum theory that it will be developed in some detail in the case of the functions e^{iux} in the next section.

172. The orthogonality concept for functions with a continuous spectrum. The difficulty encountered in the convergence of the usual orthogonality integral (171.3) can be sidestepped by a formal artifice. This artifice consists in defining a function $\varphi(u,x)$ as follows.

$$\varphi_\epsilon(u,x) = \frac{1}{\sqrt{2\pi}2\epsilon}\int_{u-\epsilon}^{u+\epsilon} e^{-i\xi x}d\xi \tag{172.1}$$

where ϵ is arbitrary though usually regarded as a small quantity, i.e., $\epsilon/u \ll 1$. H. Weyl* has shown that functions so defined satisfy the same differential equation as e^{iux} and are orthogonal in the sense illustrated below. The orthogonality condition now becomes

$$2\epsilon\int_{-\infty}^{+\infty} \varphi_\epsilon^*(u'x)\varphi_\epsilon(u,x)dx = \begin{cases} 0 & \text{when } \left|\dfrac{u'-u}{\epsilon}\right| > 1 \\ 1 & \text{when } \left|\dfrac{u'-u}{\epsilon}\right| < 1 \end{cases} \tag{172.2}$$

This means that $\varphi_\epsilon^*(u'x)$ is orthogonal to $\varphi_\epsilon(u,x)$ if u lies outside the interval $(u'-\epsilon, u'+\epsilon)$. If u is such that $u'-\epsilon < u < u'+\epsilon$, the integral can be made to have the value unity when the functions are properly normalized. All operations with the functions φ are to be performed with the ultimate idea of making ϵ vanishingly small. Such limits are taken at the termination of all manipulations.

In order to facilitate working with this type of orthonormal function, the above orthogonality integral will be worked out in detail. Using the definition (172.1), the above integral becomes

$$\begin{aligned} &\frac{1}{4\epsilon\pi}\int_{-\infty}^{+\infty} dx\int_{u'-\epsilon}^{u'+\epsilon} e^{i\eta x}d\eta\int_{u-\epsilon}^{u+\epsilon} e^{-i\xi x}d\xi \\ &= \frac{2}{4\epsilon\pi}\int_{-\infty}^{+\infty} dx\int_{u'-\epsilon}^{u'+\epsilon} e^{i\eta x}d\eta\,\frac{e^{-iux}\sin\epsilon x}{x} \end{aligned} \tag{172.3}$$

where the indicated integration over ξ has been carried out. The order of integration over η and x may now be inverted, since the resulting integral over x converges. This integral can be evaluated as follows.

$$\begin{aligned} \int_{-\infty}^{+\infty} e^{i(\eta-u)x}\frac{\sin\epsilon x}{x}dx &= \int_{-\infty}^{+\infty} e^{i(\eta-u)s/\epsilon}\frac{\sin s}{s}ds \\ &= 2\int_0^\infty \cos\left(\frac{\eta-u}{\epsilon}\right)s\frac{\sin s}{s}ds \\ &= \begin{cases} 0 & \text{when } \left|\dfrac{\eta-u}{\epsilon}\right| > 1 \\ \pi & \text{when } \left|\dfrac{\eta-u}{\epsilon}\right| < 1 \end{cases} \end{aligned}$$

* H. Weyl, *Math. Am.* **68**, 220 (1910).

For the case when the above integral gives π, we are left with the final integral

$$\frac{1}{2\epsilon\pi}\int_{u'-\epsilon}^{u'+\epsilon}\pi d\eta = 1$$

Consequently the validity of eq. (172.2) is established. It is agreeable that the normalizing integral is independent of the length of the interval 2ϵ.

Using this concept of orthogonality of the complex exponentials with a continuous spectrum, it is possible to obtain formally the Fourier formulas (171.1) and (171.2). To do this we seek a representation of a function $f(x)$ in terms of the functions $\varphi_\epsilon(u,x)$ defined by (172.1). In accordance with (170.2), this representation is to be written

$$f(x) = \int_{-\infty}^{+\infty} F(u)\varphi_\epsilon(u,x)dx \tag{172.4}$$

In order to determine $F(u)$ we proceed essentially in the same way as that in which the unknown coefficients were determined in the case of an orthonormal set of functions with a discrete spectrum (see Sec. 156). Thus we multiply the above equation by $\varphi_\epsilon^*(u',x)$, integrate over the entire range of x, and multiply the result by 2ϵ, so that

$$\begin{aligned}2\epsilon\int_{-\infty}^{+\infty}\varphi^*(u',x)f(x)dx &= 2\epsilon\int_{-\infty}^{+\infty}\varphi^*(u',x)dx\int_{-\infty}^{+\infty}F(u)\varphi(u,x)du\\ &= \int_{-\infty}^{+\infty}F(u)du\,2\epsilon\int_{-\infty}^{+\infty}\varphi^*(u',x)\varphi(u,x)dx\end{aligned}$$

On making use of the orthonormal character (172.2) of the φ's, the above equation becomes

$$2\epsilon\int_{-\infty}^{+\infty}\varphi^*(u',x)f(x)dx = \int_{u'-\epsilon}^{u'+\epsilon}F(u)du \approx F(u')2\epsilon$$

In the limit $\epsilon \to 0$, this gives

$$F(u') = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}f(x)e^{iu'x}dx$$

From the above discussion, the use of orthogonal functions in the sense just described is convenient in determining the spectral distribution required to represent a given function.

173. Relation between the mean extent of a Fourier integral and the mean spectral extent of its Fourier transform. An important relation exists between the interval Δx over which the function $|f(x)|^2$ is appreciably different from zero, and the interval Δu over which the Fourier transform $|F(u)|^2$ is appreciably different from zero. As one would conclude from the case of Fourier series (see Sec. 162), it is a necessary consequence of the properties of $f(x)$ and its Fourier transform that when the mean extent of $|f(x)|^2$ is small, $|F(u)|^2$ is

widely extended. This can be shown with the help of the orthonormal functions (172.1).

First it can be shown readily that the area under the curve of $|f(x)|^2$ is equal to the area under the curve $|F(u)|^2$. Writing $f(x)$ in the form (172.4) we have

$$f(x)f^*(x) = \int_{-\infty}^{+\infty} F(u)\varphi_\epsilon(u,x)du \cdot \int_{-\infty}^{+\infty} F^*(u')\varphi_\epsilon^*(u',x)du'$$

Integrating both sides with respect to x and inverting the order of integration of x and u and u' on the right, we have

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx =$$

$$\frac{1}{2\epsilon}\int_{-\infty}^{+\infty} F(u)du \cdot \int_{-\infty}^{+\infty} F^*(u')du' \cdot 2\epsilon \int_{-\infty}^{+\infty} \varphi_\epsilon(u,x)\varphi_\epsilon^*(u',x)dx$$

Using the orthonormal property (172.2), this becomes

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx = \frac{1}{2\epsilon}\int_{-\infty}^{+\infty} F(u)du \int_{u-\epsilon}^{u+\epsilon} F^*(u')du'$$

In the limit $\epsilon \to 0$, we have the result already mentioned above,

$$\int_{-\infty}^{+\infty} |f|^2 dx = \int_{-\infty}^{+\infty} |F(u)|^2 du \tag{173.1}$$

In order to obtain the relation between the mean spectral extent of $F(u)$ mentioned at the beginning of the section, we first define these terms to be, respectively

$$\overline{\Delta x^2} = \int_{-\infty}^{+\infty} x^2 |f|^2 dx \tag{173.2}$$

and

$$\overline{\Delta u^2} = \int_{-\infty}^{+\infty} u^2 |F(u)|^2 du \tag{173.3}$$

where the origin of x and u are chosen so that

$$\int_{-\infty}^{+\infty} x|f|^2 dx = \int_{-\infty}^{+\infty} u|F(u)|^2 du = 0$$

It is easy to show that

$$\overline{\Delta x^2} = \int_{-\infty}^{+\infty} |F'|^2 du \tag{173.4}$$

by writing

$$\begin{aligned}
\overline{\Delta x^2} &= \int_{-\infty}^{+\infty} x^2 f(x) f^*(x) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 f^*(x) dx \int_{-\infty}^{+\infty} F(u) e^{-iux} du \\
&= \frac{-1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} F(u) \frac{d^2}{du^2} \left(\int_{-\infty}^{+\infty} f^*(x) e^{-iux} dx \right) du \\
&= -\int_{-\infty}^{+\infty} F(u) \frac{d^2}{du^2} F^*(u) du = \int_{-\infty}^{+\infty} |F'(u)|^2 du
\end{aligned}$$

In the inequality

$$\left| \int_{-\infty}^{+\infty} g_1(u)h_1(u)du + \int_{-\infty}^{+\infty} g_2(u)h_2(u)du \right|^2$$
$$\leq \left(\int_{-\infty}^{+\infty} g_1 g_1^* du + \int_{-\infty}^{+\infty} g_2 g_2^* du \right) \left(\int_{-\infty}^{+\infty} h_1 h_1^* du + \int_{-\infty}^{+\infty} h_2 h_2^* du \right)$$

used in Sec. 162, we now make the substitutions

$$g_1 = uF(u) \qquad h_1 = F^{*\prime}$$
$$g_2 = uF^*(u) \qquad h_2 = F'$$

This gives

$$\left| \int_{-\infty}^{+\infty} u(FF^{*\prime} + F^*F')du \right|^2 \leq 4 \int_{-\infty}^{+\infty} u^2 \mid F \mid^2 du \int_{-\infty}^{+\infty} \mid F' \mid^2 du$$

The integral on the left can be integrated by parts,

$$\int_{-\infty}^{+\infty} u \frac{d}{du} \mid F \mid^2 du = -\int_{-\infty}^{+\infty} \mid F \mid^2 du$$

since, if the integrals on the right exist, it can be shown that the integrated-out part uFF^* vanishes at the limits of integration. The result, together with (173.2), (173.3), and (173.4), give

$$\overline{\Delta u^2}\,\overline{\Delta x^2} \geq \frac{1}{4} \left| \int_{-\infty}^{+\infty} \mid F \mid^2 du \right|^2 = \frac{1}{4} \left| \int_{-\infty}^{+\infty} \mid f(x) \mid^2 dx \right|^2 \quad (173.5)$$

This result, together with (173.1), shows that if the function $\mid f(x) \mid^2$ is narrow and high, the function $\mid F(u) \mid^2$ must be small and widely spread out, and vice versa. This result is identical with the famous indetermination principle of quantum mechanics.

174. Some examples of Fourier integrals and transforms. In electric circuit theory and the theory of probability, etc., the representation of various functions by Fourier integrals is of considerable importance. In this section a few examples of the representation of functions by means of Fourier integrals will be worked out. The problem in each case is that of determining the Fourier transform of the function to be represented.

Example 1. Let it be required to represent the Gaussian distribution function $f(x) = e^{-x^2/a^2}$ as a Fourier integral. According to eqs. (171.2) and (171.1) the Fourier transform is

$$F(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/a^2} e^{ixu} dx$$
$$= \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\xi^2} e^{iau\xi} d\xi = \frac{a}{\sqrt{2}} e^{-a^2u^2/4}$$

The integral above was evaluated in Sec. 96. The function

$$f(x) = e^{-x^2/a^2} \quad (174.1)$$

and its Fourier transform

$$F(u) = \frac{a}{\sqrt{2}} e^{-a^2u^2/4} \tag{174.2}$$

are especially illustrative of the relation (173.5) connecting the width of the function $f(x)$ and the spectral extent of $f(x)$, i.e., the width of the Fourier transform $F(u)$. If a is large, the width of $f(x)$ is large, while the width of the transform is small. For this case the inequality (173.5) becomes

$$\overline{\Delta x^2}\,\overline{\Delta u^2} \geq \frac{\pi a^2}{2}$$

Example 2. A particularly interesting and useful example is that of the discontinuous function shown in Fig. 174.1a, i.e.,

$$\begin{aligned} f(x) &= 0 & x &< 0 \\ f(x) &= \tfrac{1}{2} & x &= 0 \\ f(x) &= 1 & 0 &< x < d \\ f(x) &= \tfrac{1}{2} & x &= d \\ f(x) &= 0 & d &< x \end{aligned}$$

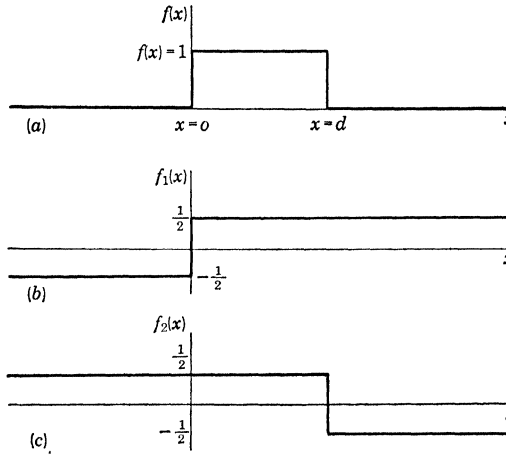


FIG. 174.1

In this case the Fourier transform (171.2) becomes

$$F(u) = \frac{1}{\sqrt{2\pi}} \int_0^d e^{-iux} dx = \frac{e^{iud} - 1}{\sqrt{2\pi} iu} \tag{174.3}$$

and consequently

$$f(x) = \frac{1}{2\pi i} \int_{-\infty}^{+\infty} \frac{e^{iud} - 1}{u} e^{-iux} du \tag{174.4}$$

It is instructive to express the above integral in terms of the sum $f(x) = f_1(x) + f_2(x)$ as follows.

$$f(x) = \frac{1}{2\pi i}\int_{-\infty}^{+\infty} \frac{e^{iux}}{u}\,du + \frac{1}{2\pi i}\int_{-\infty}^{+\infty} \frac{e^{i(d-x)u}}{u}\,du \qquad (174.5)$$

The principal value of the integrals representing $f_1(x)$ and $f_2(x)$ [see eq. (100.3)] has already been found. The first integral has the values

$$f_1(x) = \begin{cases} -\frac{1}{2} & \text{for } x < 0 \\ 0 & \text{for } x = 0 \\ +\frac{1}{2} & \text{for } x > 0 \end{cases}$$

This function $f_1(x)$ is sketched in Fig. 174.1b. The second integral has the values

$$f_2(x) = \begin{cases} -\frac{1}{2} & \text{for } d - x < 0 \\ 0 & \text{for } d - x = 0 \\ +\frac{1}{2} & \text{for } d - x > 0 \end{cases}$$

The function $f_2(x)$ is sketched in Fig. 174.1c. It is very clear from an inspection of Figs. 174.1b and 174.1c how these add up to give the original function $f(x)$ of Fig. 174.1a. Furthermore one can, in a sense, call the Fourier transform of the impulse function $f_1(x)$, much used in electric circuit theory, the function $1/\sqrt{2\pi}iu$.

175. Representations of functions in terms of a general class of functions with discrete and continuous spectra. From the theory of differential equations it is known that the Sturm-Liouville differential equation

$$[p(x)\psi']' - q(x)\psi + \rho(x)E\psi = 0 \qquad (175.1)$$

has solutions useful in physical problems for discrete and continuous values of the parameter E for $0 \leq x < \infty$. The solution $\psi_k(x)$ for discrete values of E, say E_k, occur for $E < 0$, and the solutions $\psi(E,x)$ for the continuous spectrum occur for all $E > 0$.

The most general representation of a function in terms of solutions of this differential equation must be written as a combination of a sum over the orthogonal functions $\psi_k(x)$ and an integral involving $\psi(E,x)$ in the following manner.

$$f(x) = \sum_{k=0}^{\infty} c_k\psi_k(x) + \int_0^{\infty} g(E)\psi(E,x)dE \qquad (175.2)$$

The coefficients c_k and the spectral distribution function $g(E)$ can be determined from the orthogonal properties of ψ_k and the artifice defined in Sec. 172 applied to $\psi(E,x)$. If we define $\varphi_\delta(E,x)$ to be

$$\varphi_\delta(E,x) = \frac{1}{\delta}\int_E^{E+\delta} \psi(\epsilon,x)d\epsilon \qquad (175.3)$$

it can be shown that ψ_k and $\varphi_\delta(E,x)$ satisfy the following orthogonality relations.

$$\int_0^\infty \rho(x)\psi_k^*\psi_l\,dx = \begin{matrix} 0, & k \neq l \\ 1, & k = l \end{matrix} \tag{175.4}$$

$$\int_0^\infty \rho(x)\psi_k^*(x)\varphi_\delta(E,x)dx = 0, \quad \text{for all } k \text{ and } E \tag{175.5}$$

$$\delta\int_0^\infty \rho(x)\varphi_\delta^*(E',x)\varphi_\delta(E,x)dx = \begin{matrix} 0, & E \text{ not in interval } (E', E'+\delta) \\ 1, & E' < E < E'+\delta \end{matrix} \tag{175.6}$$

It has of course been assumed that solutions $\psi_k(x)$ and $\psi(E,x)$ have been chosen so as to achieve the indicated normalization.

These properties can be used in a formal manner, the same as that used in Sec. 156 for determining coefficients in representations in terms of orthonormal sets. To do this we formally write

$$f(x) = \sum_{k=0}^{\infty} c_k\psi_k(x) + \int_0^\infty g(E)\varphi_\delta(E,x)dE \tag{175.7}$$

realizing that in the end this is to hold in the limit $\delta \to 0$. One multiplies through first by $\rho(x)\psi_{k'}^*(x)$ and integrates over the interval $0 \leq x < \infty$ in which case

$$c_{k'} = \int_0^\infty \rho(x)f(x)\psi_{k'}^*(x)dx$$

from the orthogonality relations (175.4) and (175.5). To find $g(E)$ it is necessary only to multiply through eq. (175.7) by $\varphi_\delta^*(E',x)\rho(x)$ and integrate with respect to x, and thereafter multiply the whole equation by δ, thus

$$\begin{aligned}\delta\int_0^\infty \rho(x)\varphi_\delta^*(E',x)f(x)dx &= \delta\int_0^\infty \rho(x)\varphi_\delta^*(E',x)dx\int_0^\infty g(E)\varphi_\delta(E,x)dE \\ &= \int_E^{E'+\delta} g(E)dE \approx \delta g(E') \qquad \text{for } \delta \text{ small}\end{aligned}$$

after using the orthogonality relation (175.6). In the limit $\delta \to 0$,

$$g(E') = \int_0^\infty \rho(x)f(x)\psi(E',x)dx \tag{175.8}$$

This method of obtaining orthonormal functions in the case of the continuous spectrum is much used in quantum theory. H. Weyl called the differential mean functions defined as in (173.3) eigen-differentials.

SUGGESTED REFERENCES

S. Bochner, *Vorlesungen über Fouriersche Integrale*, Leipzig, Akademische Verlagsgesellschaft, 1932.

Norbert Wiener, *The Fourier Integral and Certain of Its Applications*, London, Cambridge University Press, 1933.

PROBLEMS

Find the Fourier transforms of the following functions.

174.1. $f(x) = \dfrac{1}{(1 + x/\omega)^2}$.

174.2. $f(x) = e^{-x^2/2a^2} \sin bt$.

174.3. $f(x) = \sin \omega x$ for $-2\pi n/\omega \leq x \leq 2\pi n/\omega$.
$f(x) = 0$ elsewhere.

174.4. $f(x) = e^{-(1+i)x}$, if $x > 0$.
$f(x) = -e^{(1+i)x}$, if $x < 0$.

174.5. Let $f(x)$ satisfy the equation

$$\frac{d^2f}{dx^2} + \varphi(x)f = 0,$$

and let $f(x)$ be the Fourier transform of $F(t)$ and $\varphi(x)$ be the Fourier transform of $\chi(t)$. Show that $F(t)$ satisfies the equation

$$t^2F(t) - \int_{-\infty}^{\infty} \chi(\alpha)F(t - \alpha)d\alpha = 0.$$

Chapter 14

INTEGRAL EQUATIONS

176. Integral equations in general. An integral equation is a special form of functional equation in which the unknown function appears in the integrand of an integral. In applications they arise directly from setting up the physical relationship in a physical problem or in many cases by transforming a differential equation to an equivalent integral equation. In certain problems, especially where approximations are used, it is sometimes more convenient to transform a differential equation into an integral equation. However, integral equations are in general difficult to solve and indeed it sometimes happens that no unique solutions exist. The material in this chapter will largely be devoted to reasonably practical methods of solving certain types of linear integral equations which commonly occur in practice.

Linear integral equations are divided into two main classes, those having the form

$$f(x) = \int_a^b K(x,t)u(t)dt \tag{176.1}$$

which is called an integral equation of the first kind with the unknown function $u(t)$ appearing under the integral sign only, and those having the form

$$u(x) = f(x) + \lambda \int_a^b K(x,t)u(t)dt \tag{176.2}$$

which is called an integral equation of the second kind and which has the unknown function outside as well as under the sign of integration.

In these equations, the functions $f(x)$ and $K(x,t)$, as well as the limits a and b, are assumed known. The function $K(x,t)$ is normally called the kernel of the equation. In some cases these equations have the upper limit of integration b replaced by x, but this is really a special case of the above equations in which the kernel $K(x,t)$ vanishes identically for $t > x$, so that it makes no difference whether b is replaced by x.

177. Some integral equations of the first kind. Probably one of the first examples of an integral equation of the first kind arose in con-

nection with the Fourier integral representation of a function $f(x)$. In this case

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u(t) e^{-ixt} dt \tag{177.1}$$

When $u(t)$ [now known as the Fourier transform of $f(x)$, see Sec. 171] is regarded as the unknown function this becomes an integral equation of the first kind. From Sec. 171 we know how to invert this equation and thus obtain the solution of the integral equation whose kernel is $e^{-ixt}/\sqrt{2\pi}$.

A famous integral equation of the first kind arose very early in mechanics and was studied by Abel, who succeeded in obtaining its solution. This equation arises from the following problem. A particle starts from rest at P (see Fig. 177.1) and slides on a smooth frictionless curve lying in a vertical plane. The velocity of the particle of mass m at the point O is independent of the shape of the curve down which it slides. The time taken to traverse the curve from P to O is, however, dependent on the shape of the curve. The time taken to traverse a given distance along the curve can be easily found as follows: Let the coordinates of the point P be (x,y), and those at an intermediate point be (ξ,η), where the particle has arrived after a time t. The velocity of the particle at the intermediate point is found from the energy equation

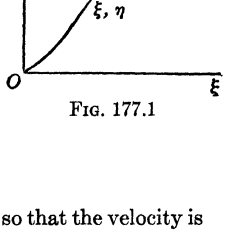


FIG. 177.1

$$\frac{1}{2}m\left(\frac{ds}{dt}\right)^2 = mg(y - \eta)$$

so that the velocity is

$$\frac{ds}{dt} = -\sqrt{2g(y - \eta)}$$

The minus sign occurs because s is measured positively from the origin O. From this we find the time elapsed to be

$$t = -\frac{1}{\sqrt{2g}} \int_{x,y}^{\xi,\eta} \frac{ds}{\sqrt{y - \eta}}$$

When the distance s along the curve is expressed in terms of the coordinate η, then $s = u(\eta)$ and $ds = u'(\eta)d\eta$, whereupon the total time of descent is given by

$$T = \frac{1}{\sqrt{2g}} \int_0^y \frac{u'(\eta)d\eta}{\sqrt{y - \eta}} \tag{177.2}$$

If, of course, $u'(\eta)$ is regarded as known, this just gives the total time required for the particle to descend, but if a time is specified and we

wish to know what the shape of curve must be to give this time, $u'(\eta)$ is unknown and we have to deal with an integral equation of the first kind. The solution of this type of equation will be given in Sec. 180.

178. Integral equations of the second kind; connection with linear differential equations. From a mathematical point of view much more is known about the properties and solutions of integral equations of the second kind, the most general form of which we shall consider is

$$u(x) = f(x) + \lambda \int_a^b K(x,t)u(t)dt \tag{178.1}$$

A commonly encountered special case of this equation is that for which $f(x) \equiv 0$, so that

$$u(x) = \lambda \int_a^b K(x,t)u(t)dt \tag{178.2}$$

This form of equation is called a homogeneous linear integral equation of the second kind, and as we shall see, has some interesting special properties.

Another special form of eq. (178.1) is

$$u(x) = f(x) + \lambda \int_0^x K(x,t)u(t)dt \tag{178.3}$$

This is known as Volterra's equation. It is of especial interest because of its connection with the linear differential equation

$$\frac{d^n y}{dx^n} + P_1(x)\frac{d^{n-1}y}{dx^{n-1}} + \ldots + P_n(x)y = F(x) \tag{178.4}$$

This differential equation can be transformed into an integral equation of the above type by the following procedure. First let

$$\frac{d^n y}{dx^n} = u(x) \tag{178.5}$$

Then it follows that

$$\frac{d^{n-1}y}{dx^{n-1}} = \int_0^x u(t_1)dt_1 + c_1$$

$$\frac{d^{n-2}y}{dx^{n-2}} = \int_0^x \int_0^{t_2} u(t_1)dt_1\,dt_2 + c_1 x + c_2$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$y = \int_0^x \int_0^{t_n} \cdots \int_0^{t_2} u(t_1)dt_1 \ldots dt_n + \frac{c_1 x^{n-1}}{(n-1)!} + \frac{c_2 x^{n-2}}{(n-2)!} + \ldots + c_n$$

From these relations the differential eq. (178.4) may be written in the form

$$u(x) + P_1(x)\int_0^x u(t_1)dt_1 + \ldots + P_n(x)\cdot \int_0^x \int_0^{t_n} \cdots \int_0^{t_2} u(t_1)dt_1 \ldots dt_n = F(x) - \sum_{i=1}^{n} c_i\varphi_i(x) \tag{178.6}$$

where

$$\varphi_i(x) = P_i(x) + P_{i+1}(x)\frac{x}{1!} + \ldots + P_n(x)\frac{x^{n-i}}{(n-i)!}$$

By integration by parts it can be established that

$$\int_0^x \int_0^{t_2} u(t_1)dt_1\, dt_2 = \int_0^x (x-t)u(t)dt$$

and in general that

$$\int_0^x \int_0^{t_n} \cdots \int_0^{t_2} u(t_1)dt_1 \ldots dt_n = \int_0^x \frac{(x-t)^{n-1}}{(n-1)!}u(t)dt$$

Using this last result, eq. (178.6) may be written

$$u(x) + \int_0^x \left[P_1(x) + P_2(x)(x-t) + P_3(x)\frac{(x-t)^2}{2!} + \ldots + P_n(x)\frac{(x-t)^{n-1}}{(n-1)!}\right]u(t)dt = f(x)$$

where

$$f(x) = F(x) - \sum_{i=1}^{n} c_i\varphi_i(x)$$

Equation (178.7) is identical with the Volterra integral eq. (178.3), where the kernel is

$$K(x,t) = P_1(x) + P_2(x)\frac{(x-t)}{1!} + \ldots + P_n(x)\frac{(x-t)^{n-1}}{(n-1)!} \tag{178.8}$$

Thus it is possible to transform a linear differential equation into an integral equation of the second kind. To make the transformation complete it is necessary to know the initial conditions to be used with the differential equations, since the constants $c_1, \ldots, c_n$ are determined from these conditions.

179. Types of solutions of integral equations. Before taking up the methods of obtaining solutions of integral equations in detail it is useful to see something of the form of solution to be obtained by the several methods. In the main there are four ways of obtaining solutions of integral equations. Certain of these may readily form the basis of obtaining a numerical solution.

1. One method, especially useful in the case of equations of the first kind, makes use of certain inversion formulas, of which the Fourier

transform (Sec. 171) furnishes an example. Equations of the form of (177.2) can be inverted by means of an inversion formula (see Sec. 180).

2. In the case of integral equations of the second kind, there is a method of obtaining a solution which is due largely to Neumann, Liouville, and Volterra. This method gives $u(x)$ in the form of a power series in λ, the coefficients being functions of x. This method yields a solution in those cases where the resulting power series in λ converges.

3. The method of Fredholm provides a solution for equations of the second kind, which is a ratio of two power series in λ, each of which has an infinite radius of convergence. This method gives

$$u(x) = \frac{F_0(x) + F_1(x)\lambda + F_2(x)\lambda^2 + \dots}{a_0 + a_1\lambda + a_2\lambda^2 + \dots}$$

where the denominator is independent of x. For the particular values of λ which make the denominator vanish, there is in general no solution. However, the method provides a solution for those cases where a solution exists. In cases where the power series converge slowly, the functions $F_n(x)$ and constants a_n are difficult to determine in practice for large values of n.

4. A method for solving equations of the second kind has been given by Hilbert and Schmidt. This method expresses $u(x)$ in terms of a set of characteristic orthogonal functions $u_1(x)$, $u_2(x)$, etc., which are solutions of the homogeneous equation

$$u(x) = \lambda \int_a^b K(x,t)u(t)dt$$

In general, this form of equation has solutions other than $u(x) \equiv 0$ only for a certain characteristic set of numbers $\lambda_1, \lambda_2, \dots$. The corresponding finite solutions $u_1(x)$, $u_2(x)$, etc., form an orthogonal set of characteristic solutions (see Sec. 185). Writing the above equation in the form

$$\int_a^b K(x,t)u_n(t)dt = \frac{1}{\lambda_n} u_n(x)$$

brings out the fact that $1/\lambda_n$ and $u_n(x)$ are the eigen-values and eigenfunctions of the operator $\int_a^b K(x,t)dt$. When the eigen-values and eigenfunctions of this operator have been determined, the solution of the equation of the second kind is given by

$$u(x) = \sum_{n=1}^{\infty} c_n u_n(x)$$

where the constants c_n depend on $f(x)$.

180. Method of inversion for equations of the first kind. There is no general method for obtaining an inversion of an integral equation

of the first kind. There are, however, special forms of such equations which can be solved by inversion with the help of certain inversion formulas. Two examples of the method will be discussed here.

1. Equations invertable by means of Fourier transforms. Equations of the form

$$f(x) = \int_{-\infty}^{+\infty} k(x - t)u(t)dt \tag{180.1}$$

can be inverted by means of Fourier transforms. It will be apparent in the following that this is a consequence of the fact that the kernel is a function of $(x - t)$ only. In order to invert the above equation we multiply through by $(1/\sqrt{2\pi})e^{izx}$ and integrate with respect to x. Thus

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(x)e^{izx}dx &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} dx\, e^{izx}\int_{-\infty}^{+\infty} k(x - t)u(t)dt \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} u(t)e^{izt}dt\int_{-\infty}^{+\infty} k(x - t)e^{iz(x-t)}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} u(t)e^{izt}dt\int_{-\infty}^{+\infty} k(s)e^{izs}ds\end{aligned}$$

where the order of integration was inverted and the change of variable $s = x - t$ was made. From eq. (171.2) these integrals are formally just the Fourier transforms of $f(x)$, $u(t)$, and $k(s)$, so that the above equation can be written in terms of these transforms, in which case

$$F(z) = \sqrt{2\pi}U(z)K(z)$$

Thus the Fourier transform $U(z)$ of the unknown function is determined. From the theory of Fourier transforms (see Sec. 171) the unknown function $u(x)$ is then given by

$$u(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{F(z)}{K(z)}e^{-izx}dz \tag{180.2}$$

This method of course fails if the pertinent Fourier transforms do not exist. This requires that

$$\int_{-\infty}^{+\infty} |\, k(s)\,|^2 ds, \quad \int_{-\infty}^{+\infty} |\, f(x)\,|^2 dx, \quad \text{etc., exist}$$

2. Inversion of the Abel type of equation of the first kind. A somewhat more general equation than Abel's eq. (177.2) can be solved by an inversion method. The equation to be considered is

$$f(x) = \int_a^x \frac{u(t)}{(x - t)^\alpha}dt \tag{180.3}$$

where $f(a) = 0$ and $f'(x)$ is continuous, and $0 < \alpha < 1$. To invert this equation, multiply through by $1/(z - x)^{1-\alpha}$ and integrate from a to z. This gives

$$\int_a^z \frac{f(x)dx}{(z-x)^{1-\alpha}} = \int_a^z dx \int_a^z \frac{u(t)dt}{(z-x)^{1-\alpha}(x-t)^{\alpha}} \tag{180.4}$$

One may now make use of the Dirichlet formula

$$\begin{aligned} &\int_a^z dt \int_t^z (t-a)^{\lambda-1}(z-x)^{\mu-1}(x-t)^{\nu-1}\psi(x,t)dx \\ &= \int_a^z dx \int_a^x (t-a)^{\lambda-1}(z-x)^{\mu-1}(x-t)^{\nu-1}\psi(x,t)dt \end{aligned} \tag{180.5}$$

for changing the order of integration, where the integrand has the special character shown above and $0 < \lambda \leq 1$, $0 < \mu \leq 1$, $0 < \nu \leq 1$.

Equation (180.4) then becomes

$$\int_a^z \frac{f(x)dx}{(z-x)^{1-\alpha}} = \int_a^z u(t)dt \int_t^z \frac{dx}{(z-x)^{1-\alpha}(x-t)^{\alpha}} \tag{180.6}$$

It is now possible to evaluate the integral

$$I = \int_t^z \frac{dx}{(z-x)^{1-\alpha}(x-t)^{\alpha}}$$

This integral may readily be expressed in terms of the beta function [see eq. (62.2)] by a transformation of variable according to the equation

$$\eta = \frac{x-t}{z-t}$$

whereupon

$$I = \int_0^1 \eta^{-\alpha}(1-\eta)^{\alpha-1}\,d\eta = B(1-\alpha,\alpha) \tag{180.7}$$

Using eq. (62.1) connecting the beta function with the gamma function,

$$I = \Gamma(1-\alpha)\Gamma(\alpha)$$

According to eq. (109.7) this combination of gamma functions yields the simple result that

$$I = \frac{\pi}{\sin \pi\alpha}$$

With this result used in eq. (180.6), we have

$$\int_a^z \frac{f(x)dx}{(z-x)^{1-\alpha}} = \frac{\pi}{\sin \pi\alpha} \int_a^z u(t)dt \tag{180.8}$$

The reason for choosing the exponent of $(z-x)$ to be $1-\alpha$ was simply to achieve the simple result for the integral I.

Since $f'(x)$ is assumed continuous, we may take the derivative of the above equation with respect to z, which completes the inversion of the original equation and furnishes the solution of the integral equation in the form

$$u(z) = \frac{\sin \pi\alpha}{\pi} \frac{d}{dz} \int_a^z \frac{f(x)dx}{(z-x)^{1-\alpha}} \tag{180.9}$$

That this is a solution of the integral eq. (180.3) can be verified by direct substitution in it.

181. Solution of integral equations of the second kind by the method of successive approximations. This method of obtaining a solution of the integral equation

$$u(x) = f(x) + \lambda \int_a^b K(x,t)u(t)dt, \tag{181.1}$$

where $f(x) \not\equiv 0$ and is continuous in the interval $a \leq x \leq b$, and $K(x,t)$ is finite and has at worst only scattered discontinuities in the rectangle $a \leq x \leq b$ and $a \leq t \leq b$, and λ is a constant, yields the solution in the form of a power series in λ in accordance with method 2 of Sec. 179. When the resulting series converges rapidly, this method is particularly useful for obtaining approximate solutions of the integral equation, and because of the connection with linear differential equations so also for approximate solutions of differential equations (see Sec. 178).

The method of successive approximations may be illustrated by applying the method to the solution of a simple algebraic equation, namely,

$$y = 1 + \lambda y \tag{181.2}$$

If we suppose that λ is not too large we may obtain a zero-order approximation by setting $y = 1$; hence

$$y_0 = 1$$

We may expect to obtain a closer approximation to the solution by replacing y on the right side of eq. (181.2) by y_0; hence

$$y_1 = 1 + \lambda y_0 = 1 + \lambda$$

Continuing this process,

$$y_2 = 1 + \lambda y_1 = 1 + \lambda + \lambda^2$$
$$y_3 = 1 + \lambda y_2 = 1 + \lambda + \lambda^2 + \lambda^3$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

If the process is continued indefinitely we obtain the series representation of $y = 1/(1 - \lambda)$, which is the solution of eq. (181.2). The series thus obtained is a solution of the original equation for those values of λ for which it converges, i.e., $|\lambda| < 1$. If λ is small, a sufficiently good approximation to the solution might be obtained by repeating the process only a few times.

In the case of the integral eq. (181.1), one may adopt a similar procedure, whereupon the zero-order approximation can be taken to be $f(x)$ itself, so that

$$u_0 = f(x)$$

$$u_1 = f(x) + \lambda\int_a^b K(x,t)u_0(t)dt = f(x) + \lambda\int_a^b K(x,t)f(t)dt$$

$$u_2 = f(x) + \lambda\int_a^b K(x,t)f(t)dt + \lambda^2\int_a^b K(x,t)dt\int_a^b K(t,t_1)f(t_1)dt_1$$

. . . . . .

$$u_n = s_0(x) + \lambda s_1(x) + \lambda^2 s_2(x) + \ldots + \lambda^n s_n(x) \qquad (181.3)$$

where

$$s_0(x) = f(x)$$

$$s_1(x) = \int_a^b K(x,t)f(t)dt$$

$$s_2(x) = \int_a^b K(x,t)\int_a^b K(t,t_1)f(t_1)dt\,dt_1$$

. . . . . .

$$s_n(x) = \int_a^b K(x,t)\int_a^b K(t,t_1)\int_a^b K(t_1,t_2)\ \ldots$$

$$\int_a^b K(t_{n-2},t_{n-1})f(t_{n-1})dt_{n-1}\ \ldots\ dt_1\,dt$$

It is to be expected that the solution of the original equation would be

$$u(x) = \lim_{n\to\infty} u_n(x) = \lim_{n\to\infty} \sum_{j=0}^{n} \lambda^j s_j(x) \qquad (181.4)$$

if the indicated limit exists or the series on the right converges. In order to obtain some idea of conditions which will insure the convergence of this series, it may be noted that

$$|\lambda^n s_n(x)| \leq |\lambda^n|\,|f_{max}K^n_{max}|\,|b-a|^n$$

where f_{max} denotes the maximum value of the function $f(x)$, and K_{max} is the maximum value of the kernel $K(x,t)$ within the domain of x and t under consideration. By the comparison text (see Sec. 76) the series under consideration will converge if the series formed by replacing each $\lambda^n s_n(x)$ by the term appearing on the right side of the above inequality converges. By the ratio test this converges if

$$|\lambda|\,|K_{max}|\,|b-a| < 1 \qquad (181.5)$$

If this condition is satisfied there is a continuous solution of the original equation given by eq. (181.4). There may be a continuous solution of the integral equation when the above condition is not satisfied, but it would not be found by this method of solution.

A procedure entirely similar to that discussed above can be formulated for the Volterra eq. (178.3) where the upper limit of integration b is replaced by x.

Illustrative Example. The above method will be adapted to the solution of the integral equation

$$u(x) = x + \int_0^x (t - x)u(t)dt \tag{181.6}$$

Following the procedure outlined in this section,

$$s_0(x) = x$$

$$s_1(x) = \int_0^x (t - x)t\,dt = -\frac{x^3}{3!}$$

$$s_2(x) = \int_0^x (t - x)\int_0^t (t_1 - t)t_1\,dt_1\,dt = \frac{x^5}{5!}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$s_n(x) = (-1)^n \frac{x^{2n+1}}{(2n + 1)!}$$

From eq. (181.4) the solution sought is

$$u(x) = \lim_{n\to\infty} \sum_{j=0}^{n} (-1)^j \frac{x^{2j+1}}{(2j + 1)!} = \sin x \tag{181.7}$$

182. The Fredholm method. Here we take up the third method mentioned in Sec. 179. This provides a general solution of the equation

$$u(x) = f(x) + \lambda \int_a^b K(x,t)u(t)dt \tag{182.1}$$

Although it is sometimes rather difficult to carry out the general solution of the equation by the method of Fredholm, the method can be used as a basis for an approximate numerical solution in those cases where it is not practical to carry the method through for a complete solution.

Fredholm's method consists in setting up an analogy between the solution of the above equation and a system of ordinary linear algebraic equations. The analogy is set up by dividing the interval of integration (ab) into n equal parts, designating the points of subdivision by t_0, t_1, t_2, $\ldots$, t_n, where $t_0 = a$ and $t_n = b$. The length of each interval in the subdivision is then

$$h = \frac{b - a}{n} \tag{182.2}$$

and consequently

$$t_0 = a; \quad t_1 = a + h; \quad t_2 = a + 2h; \quad \ldots; \quad t_n = a + nh \tag{182.3}$$

On the basis of this scheme of subdivision, the integral in eq. (182.1) can of course be expressed as the limit of the following sum.

$$\int_a^b K(x,t)u(t)dt = \lim_{n\to\infty} \sum_{i=1}^{n} K(x,t_i)u(t_i)h \tag{182.4}$$

If the magnitude of h is not too large, we can obtain a reasonably good approximation for the above integral without passing to the limit.

Thus we may replace the integral in (182.1) by the above sum, so that the integral eq. (182.1) can be replaced by the approximate equation

$$u(x) - \lambda h[K(x,t_1)u(t_1) + K(x,t_2)u(t_2) + \ldots + K(x,t_n)u(t_n)] = f(x) \tag{182.5}$$

The original integral equation must of course hold for all values of x and consequently in particular for $x = t_1, t_2, \ldots, t_n$. Substituting these values of x in (182.5), we obtain the following n linear algebraic equations

$$\left.\begin{aligned}
u(t_1) - \lambda h[K(t_1,t_1)u(t_1) + K(t_1,t_2)u(t_2) + \ldots \\
+ K(t_1,t_n)u(t_n)] = f(t_1) \\
u(t_2) - \lambda h[K(t_2,t_1)u(t_1) + K(t_2,t_2)u(t_2) + \ldots \\
+ K(t_2t_n)u(t_n)] = f(t_2) \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
u(t_n) - \lambda h[K(t_n,t_1)u(t_1) + K(t_n,t_2)u(t_2) + \ldots \\
+ K(t_n,t_n)u(t_n)] = f(t_n)
\end{aligned}\right\} \tag{182.6}$$

By the procedure adopted above, the original integral equation has been replaced by a system of n approximate nonhomogeneous linear algebraic equations to solve for the n unknowns $u(t_1), u(t_2), \ldots, u(t_n)$. The solution of these equations by means of ordinary algebraic methods (see Sec. 132) will furnish the values of the unknown function at n points. One can then draw a curve through these points by interpolation and thus obtain an approximate solution.

The exact solution of the integral equation can indeed be obtained by proceeding with the solution of the above system of linear equations. Considerable simplification of the ensuing work will be obtained by adopting the following simpler notation. Let

$$\begin{aligned} u(t_i) &= u_i \\ K(t_i,t_j) &= K_{ij} \\ f(t_i) &= f_i \end{aligned}$$

Written in this notation the eqs. (182.6) are

$$\left.\begin{aligned}
u_1 - \lambda h[K_{11}u_1 + K_{12}u_2 + \ldots + K_{1n}u_n] = f_1 \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
u_n - \lambda h[K_{n1}u_1 + K_{n2}u_2 + \ldots + K_{nn}u_n] = f_n
\end{aligned}\right\} \tag{182.7}$$

In accordance with the theory of linear equations (see secs. 131 and 132), the solution of these equations is

$$u_r = \frac{A_r}{A} \tag{182.8}$$

where A is the determinant of the coefficients of the u_ν's, i.e.,

$$A = \begin{vmatrix} (1 - \lambda h K_{11}) & -\lambda h K_{12} & \dots & -\lambda h K_{1n} \\ -\lambda h K_{21} & (1 - \lambda h K_{22}) & \dots & -\lambda h K_{2n} \\ \dots\dots\dots & \dots\dots\dots & \dots & \dots\dots\dots \\ \dots\dots\dots & \dots\dots\dots & \dots & \dots\dots\dots \\ -\lambda h K_{n1} & -\lambda h K_{n2} & \dots & (1 - \lambda h K_{nn}) \end{vmatrix} \tag{182.9}$$

and A_ν is the determinant formed by replacing the νth column in the above determinant by $f_1, f_2, \dots, f_n$.

Following Laplace's development of A_ν in terms of its first minors with respect to column ν, we have (see Sec. 131)

$$A_\nu = f_1 A_{1\nu} + f_2 A_{2\nu} + \dots + f_n A_{n\nu} \tag{182.10}$$

where $A_{\mu\nu}$ is the first minor of A_ν with respect to the μth row and νth column. Combining eqs. (182.10) and (182.7), an approximate solution to the integral equation is given by

$$u_\nu = \frac{\sum_{k=1}^{n} f_k A_{k\nu}}{A} \qquad \nu = 1, 2, \dots, n \tag{182.11}$$

This solution should approach the exact solution as the number of subdivisions in the interval (ab) increases or as $h \to 0$. In order to ascertain the limits approached by $A_{\mu\nu}$ and A as a result of this process, it is most convenient to expand two-rowed, three-rowed, etc., determinants in powers of λ, and deduce the result by induction.

Starting with a two-rowed determinant of the form (182.9), we have

$$\begin{vmatrix} (1 - \lambda a_{11}) & -\lambda a_{12} \\ -\lambda a_{21} & (1 - \lambda a_{22}) \end{vmatrix} = 1 - \lambda(a_{11} + a_{22}) + \lambda^2 \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \tag{182.12}$$

On expanding a similar three-rowed determinant,

$$\begin{aligned} &\begin{vmatrix} (1 - \lambda a_{11}) & -\lambda a_{22} & -\lambda a_{13} \\ -\lambda a_{21} & (1 - \lambda a_{22}) & -\lambda a_{23} \\ -\lambda a_{31} & -\lambda a_{32} & (1 - \lambda a_{33}) \end{vmatrix} \\ &= 1 - (a_{11} + a_{22} + a_{33}) \\ &\quad + \lambda^2 \left[\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \right] \\ &\quad - \lambda^3 \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &= 1 - \lambda \sum_{i=1}^{3} a_{ii} + \frac{\lambda^2}{2} \sum_{i,j=1}^{3} \begin{vmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{vmatrix} \\ &\quad - \frac{\lambda^3}{6} \sum_{i,j,k=1}^{3} \begin{vmatrix} a_{ii} & a_{ij} & a_{ik} \\ a_{ji} & a_{jj} & a_{jk} \\ a_{ki} & a_{kj} & a_{kk} \end{vmatrix} \end{aligned} \tag{182.13}$$

Proceeding by induction we infer that A of (182.9) may be written

$$A = 1 - \lambda \sum_{i=1}^{n} K_{ii}h + \frac{\lambda^2}{2!} \sum_{i,j=1}^{n} h^2 \begin{vmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{vmatrix} - \frac{\lambda^3}{3!} \sum_{i,j,k=1}^{n} h^3 \begin{vmatrix} K_{ii} & K_{ij} & K_{ik} \\ K_{ji} & K_{jj} & K_{jk} \\ K_{ki} & K_{kj} & K_{kk} \end{vmatrix} + \dots \tag{182.14}$$

On a formal basis we are led to expect that as $n \to \infty$ or $h \to 0$, eq. (175.14) goes over into

$$\lim_{n\to\infty} A = D(\lambda) = 1 - \lambda \int_a^b K(t,t)dt + \frac{\lambda^2}{2!} \int_a^b \int_a^b \begin{vmatrix} K(t_1,t_1) & K(t_1,t_2) \\ K(t_2,t_1) & K(t_2,t_2) \end{vmatrix} dt_1\, dt_2 - \frac{\lambda^3}{3!} \int_a^b \int_a^b \int_a^b \begin{vmatrix} K(t_1,t_1) & K(t_1,t_2) & K(t_1,t_3) \\ K(t_2,t_1) & K(t_2,t_2) & K(t_2,t_3) \\ K(t_3,t_1) & K(t_3,t_2) & K(t_3,t_3) \end{vmatrix} dt_1\, dt_2\, dt_3 + \dots \tag{182.15}$$

where $D(\lambda)$ is called the determinant of the kernel K.

In order to infer the limit of $A_{\mu\nu}$, it is convenient first to expand $A_{\nu\nu}$ in powers of λ. Following the same procedure as above it is easy to verify that

$$A_{\nu\nu} = 1 - \lambda \sum_{i=1}^{n}{}' K_{ii}h + \frac{\lambda^2}{2} \sum_{i,j=1}^{n}{}' h^2 \begin{vmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{vmatrix} - \dots$$

where $\sum'$ means that the term $i = \nu$ is omitted. In the limit as $n \to \infty$ the above series would have the same limit as A, so that

$$\lim_{n\to\infty} A_{\nu\nu} = D(\lambda) \tag{182.16}$$

Then expanding the determinant $A_{\mu\nu}$, one obtains

$$A_{\mu\nu} = \lambda h \left(K_{\nu\mu} - \lambda \sum_{i=1}^{n} h \begin{vmatrix} K_{\nu\mu} & K_{\nu i} \\ K_{i\mu} & K_{ii} \end{vmatrix} + \frac{\lambda^2}{2!} \sum_{i,j=1}^{n} h^2 \begin{vmatrix} K_{\nu\mu} & K_{\nu i} & K_{\nu j} \\ K_{i\mu} & K_{ii} & K_{ij} \\ K_{j\mu} & K_{ji} & K_{jj} \end{vmatrix} - \dots \right)$$

On letting $D_{\nu\mu} = A_{\mu\nu}/h$ and $n \to \infty$ in such a way that $t_\nu \to x$ and $t_\mu \to y$, it is to be inferred that

$$\lim_{n\to\infty} D_{\nu\mu} = D(x,y,\lambda) = \lambda K(x,y) - \lambda^2 \int_a^b \begin{vmatrix} K(x,y) & K(x,t) \\ K(t,y) & K(t,t) \end{vmatrix} dt + \frac{\lambda^3}{2!} \int_a^b \int_a^b \begin{vmatrix} K(x,y) & K(x,t_1) & K(x,t_2) \\ K(t_1,y) & K(t_1,t_1) & K(t_1,t_2) \\ K(t_2,y) & K(t_2,t_1) & K(t_2,t_2) \end{vmatrix} dt_1\, dt_2 - \dots \tag{182.17}$$

The approximate solution (182.11) can be conveniently written in the form

$$\left.\begin{aligned} u_\nu &= \frac{f_\nu A_{\nu\nu}}{A} + \sum_{\mu=1}^{n}{}' \frac{f_\mu A_{\mu\nu}}{A} \\ &= \frac{f_\nu A_{\nu\nu}}{A} + \sum_{\mu=1}^{n}{}' \frac{f_\mu h D_{\nu\mu}}{A} \end{aligned}\right\} \tag{182.18}$$

where $\sum'$ means that $\mu = \nu$ is omitted from the sum. On using the limits (182.15), (182.16), and (182.17) for A, $A_{\nu\nu}$, and $D_{\nu\mu}$, respectively, the limiting form of (182.18) is

$$u(t_\nu) = f(t_\nu) + \int_a^b \frac{f(t)D(t_\nu,t,\lambda)}{D(\lambda)}\,dt$$

or since t_ν is any point of the subdivision it can be replaced by the general variable x in the interval (ab). Thus the inferred solution of the original integral eq. (182.1) is

$$u(x) = f(x) + \frac{1}{D(\lambda)} \int_a^b f(t)D(x,t,\lambda)dt \tag{182.19}$$

It can be shown that if

(a) $D(\lambda) \neq 0$;
(b) $K(x,y)$ is continuous in the rectangle $a \leq x \leq b$; $a \leq y \leq b$;
(c) $f(x)$ is continuous in the interval $a \leq x \leq b$; the equation

$$u(x) = f(x) + \int_a^b K(x,t)u(t)dt \tag{182.20}$$

has one and only one continuous solution $u(x)$, given by

$$u(x) = f(x) + \frac{1}{D(\lambda)} \int_a^b f(t)D(x,t,\lambda)dt \tag{182.21}$$

where $D(x,t,\lambda)$ and $D(\lambda)$ are absolutely convergent series for all values of λ, and $D(x,t,\lambda)$ converges uniformly with respect to x and t in the rectangle of definition.

From the above facts it is at once possible to deduce that if $D(\lambda) \neq 0$, the homogeneous integral equation

$$u(x) = \int_a^b K(x,t)u(t)dt$$

has one and only one continuous solution $u(x)$, which is $u(x) \equiv 0$. It will be seen in the next section that in order for this equation to have a continuous solution not identically zero, $D(\lambda)$ must be zero.

Fredholm established two fundamental relations between $D(x,y,\lambda)$, $D(\lambda)$, and $K(x,y)$, which will be useful in the discussion of the homogeneous integral equation in the next section.

$$D(x,y,\lambda) - \lambda K(x,y)D(\lambda) = \lambda \int_a^b D(x,t,\lambda)K(t,y)dt \tag{182.22}$$

$$= \lambda \int_a^b K(x,t)D(t,y,\lambda)dt \tag{182.23}$$

These relations can readily be verified by comparing the coefficients of like powers of λ on using the series expansions in λ of $D(x,y,\lambda)$, $D(\lambda)$, etc., given by (182.15) and (182.17). In the same way, the relation

$$\int_a^b D(x,x,\lambda)dx = -\lambda \frac{dD(\lambda)}{d\lambda} \tag{182.24}$$

can be established.

Recurrence Formulas for Coefficients of Powers of λ *in* $D(\lambda)$ *and* $D(x,y,\lambda)$. With the help of formulas (182.24) and (182.23), convenient computational recurrence formulas for the coefficients of the various powers of λ in the power series representations of $D(\lambda)$ and $D(x,y,\lambda)$ can be obtained. To do this we first write eqs. (182.15) and (182.17) as

$$D(\lambda) = 1 + d_1\lambda + d_2\lambda^2 + d_3\lambda^3 + \ldots + d_m\lambda^m + \ldots \tag{182.25}$$

$$D(x,y,\lambda) = d_1(x,y)\lambda + d_2(x,y)\lambda^2 + \ldots + d_m(x,y)\lambda^m + \ldots \tag{182.26}$$

On substituting these two series in the relation (182.24) and equating the coefficients of like powers of λ,

$$d_m = -\frac{1}{m}\int_a^b d_m(x,x)dx \tag{182.27}$$

a relation that allows us to compute the constant d_m when $d_m(x,x)$ is known.

On the other hand, if the above two series are substituted in the relation (182.23) and the coefficients of like powers of λ are equated, we find the useful recurrence formula

$$d_m(x,y) = d_{m-1}K(x,y) + \int_a^b K(x,t)d_{m-1}(t,y)dt \tag{182.28}$$

From (182.17),

$$d_1(x,y) = K(x,y) \tag{182.29}$$

Consequently d_1 can be computed from (182.27) and thereafter $d_2(x,y)$ can be computed from (182.28), so that all the terms in the series (182.25) and (182.26) can be computed one after the other.

Examples of the Fredholm Method. Very often it is impractical to obtain the exact solution of a complicated integral equation of the second kind by the use of Fredholm's method because the terms in the series expansions of $D(x,y,\lambda)$ and $D(\lambda)$ become too unwieldy. However, two examples to illustrate the use of the method will be given.

Example 1. Let it be required to solve the integral equation

$$u(x) = x + \int_0^1 xtu(t)dt$$

In accordance with (182.19), we need to calculate $D(\lambda)$ and $D(x,y,\lambda)$, where λ is unity. To compute $D(\lambda)$, it is necessary only to substitute

the kernel $K(x,t) = xt$ in eq. (182.15). On making this substitution it is found that

$$D(\lambda) = 1 - \int_0^1 t^2 dt + \frac{1}{2!}\int_0^1\int_0^1 \begin{vmatrix} t_1^2 & t_1t_2 \\ t_2t_1 & t_2^2 \end{vmatrix} dt_1\,dt_2$$
$$- \frac{1}{3!}\int_0^1\int_0^1\int_0^1 \begin{vmatrix} t_1^2 & t_1t_2 & t_1t_3 \\ t_2t_1 & t_2^2 & t_2t_3 \\ t_3t_1 & t_3t_2 & t_3^2 \end{vmatrix} dt_1\,dt_2\,dt_3 + \ldots$$

It will be noted at once that the above terms involving determinants vanish, since on factoring out common factors at least two rows become identical, thus making the determinant vanish. It may be easily verified that all terms involving higher order determinants vanish for the same reason. Hence

$$D(\lambda) = 1 - \tfrac{1}{3} = \tfrac{2}{3}$$

$D(x,y,\lambda)$ can be calculated by substituting the kernel in eq. (182.17). Then

$$D(x,y,\lambda) = xy - \int_0^1 \begin{vmatrix} xy & xt \\ ty & t^2 \end{vmatrix} dt + \frac{1}{2!}\int_0^1\int_0^1 \begin{vmatrix} xy & xt_1 & xt_2 \\ t_1y & t_1^2 & t_1t_2 \\ t_2y & t_2t_1 & t_2^2 \end{vmatrix} dt_1\,dt_2 - \ldots$$

For the same reasons as above, all the terms involving determinants vanish, so that

$$D(x,y,\lambda) = xy$$

From (182.21), the solution of our integral equation is

$$u(x) = x + \frac{3}{2}\int_0^1 txt\,dt = \frac{3}{2}x$$

Example 2. Let it be required to solve the integral equation

$$u(x) = e^x - \frac{e}{2} + \frac{1}{2} + \frac{1}{2}\int_0^1 u(t)dt$$

Using the Fredholm procedure, we substitute $\lambda = \frac{1}{2}$ and $K(x,t) = 1$ in eq. (175.15), which gives

$$D(\lambda) = 1 - \frac{1}{2}\int_0^1 dt + \left(\frac{1}{2}\right)^2\frac{1}{2!}\int_0^1\int_0^1 \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} dt - \ldots$$
$$= 1 - \frac{1}{2} = \frac{1}{2}$$

Again all terms involving determinants vanish. On substituting in eq. (182.21), the desired solution is

$$u(x) = e^x - \frac{e}{2} + \frac{1}{2} + 2\int_0^1 \frac{1}{2}\left(e^t - \frac{e}{2} + \frac{1}{2}\right)dt$$
$$= e^x$$

Although this simple integral equation can be readily solved by applying the Fredholm method, one would be likely to make use of some ingenuity and solve the equation as follows. Let

$$u(x) = \varphi'(x)$$

whereupon the integral equation may be written

$$\varphi'(x) = e^x - \frac{e}{2} + \frac{1}{2} + \frac{1}{2}[\varphi(1) - \varphi(0)]$$

Integrating this with respect to x gives

$$\varphi(x) = e^x + \frac{1}{2}\{1 + [\varphi(1) - \varphi(0)] - e\}x + c$$

This equation holds for all x in the interval $0 \leq x \leq 1$ and in particular at $x = 0$ and $x = 1$. For these values of x the above equation becomes

$$\varphi(1) = e + \frac{1}{2}\{1 + [\varphi(1) - \varphi(0)] - e\} + c$$

$$\varphi(0) = 1 + c$$

On solving for $\varphi(1)$ in terms of c, the arbitrary integration constant, one has

$$\varphi(1) = e + c$$

so that

$$\varphi(1) - \varphi(0) = e - 1$$

Finally

$$u(x) = \varphi'(x) = e^x$$

and no formidable artillery was needed to solve the equation.

183. Homogeneous integral equations. In the last section it was noted that the homogeneous integral equation

$$u(x) = \lambda \int_a^b K(x,t)u(t)dt \tag{183.1}$$

has no continuous solution other than $u(x) \equiv 0$ if $D(\lambda) \neq 0$. Since $D(\lambda)$ was the limit of the determinant A of the coefficients of a set of n linear algebraic equations as $n \to \infty$, this situation is quite analogous to that encountered in the solution of such a system of homogeneous algebraic equations when $A = 0$. In the case of the homogeneous algebraic equation nonzero solutions exist when $A = 0$ (see Sec. 133). Likewise a continuous solution not identically zero exists for eq. (183.1) when $D(\lambda) = 0$. However, $D(\lambda)$ will vanish only for certain values of λ which are the roots of the equation

$$D(\lambda) = 0 \tag{183.2}$$

The values of λ satisfying the above equation are called the characteristic values or eigen-values of the kernel $K(x,y)$.

In order to obtain the solution in this case, we start by letting $\lambda = \lambda_1$, be a multiple root m times repeated of eq. (183.2). Since $D(\lambda)$ is an analytic function, it may be expanded in a power series about the point $\lambda = \lambda_1$ which may be written

$$D(\lambda) = c_m(\lambda - \lambda_1)^m + c_{m+1}(\lambda - \lambda_1)^{m+1} + \ldots \tag{183.3}$$

where $m > 0$ and $c_m \neq 0$. Also since $D(x,y,\lambda)$ is an analytic function of λ for all values of x and y in the rectangle of definition, we may also expand it in a power series about the point $\lambda = \lambda_1$. This expansion may be written

$$D(x,y,\lambda) = \frac{g_k(x,y)}{k!}(\lambda - \lambda_1)^k + \frac{g_{k+1}(x,y)}{(k+1)!}(\lambda - \lambda_1)^{k+1} + \ldots \tag{183.4}$$

where $k \geq 0$.

If (183.3) and (183.4) are substituted into the relation (182.24), we have the result that

$$\frac{(\lambda - \lambda_1)^k}{k!}\int_a^b g_k(x,x)dx + \frac{(\lambda - \lambda_1)^{k+1}}{(k+1)!}\int_a^b g_{k+1}(x,x)dx + \ldots$$
$$= -\lambda m c_m(\lambda - \lambda_1)^{m-1} - \lambda(m+1)c_{m+1}(\lambda - \lambda_1)^m - \ldots$$

For this equality to hold it is necessary that $k \leq m - 1$, a result shortly needed.

An important result is obtained when the series (183.3) and (183.4) are substituted in the relation (182.23). This result is

$$\frac{g_k(x,y)}{k!}(\lambda - \lambda_1)^k + \frac{g_{k+1}(x,y)}{(k+1)!}(\lambda - \lambda_1)^{k+1} + \ldots$$
$$= \lambda K(x,y)[c_m(\lambda - \lambda_1)^m + c_{m+1}(\lambda - \lambda_1)^{m+1} + \ldots]$$
$$+ \lambda\int_a^b K(x,t)\left[\frac{g_k(t,y)}{k!}(\lambda - \lambda_1)^k + \frac{g_{k+1}(t,y)}{(k+1)!}(\lambda - \lambda_1)^{k+1} + \ldots\right]dt$$

On dividing this equation through by $(\lambda - \lambda_1)^k$ and then passing to the limit $\lambda \to \lambda_1$,

$$g_k(x,y) = \lambda_1\int_a^b K(x,t)g_k(t,y)dt \tag{183.5}$$

since it was just established above that $m > k$. However, for a given value of y this is just the integral equation we started with. Consequently if $g(x,y)$ were known, a variety of solutions could be obtained by assigning y specific values. Of course, a linear combination of these solutions would also be a solution of the equation.

In case λ_1 is a nonrepeated root of $D(\lambda) = 0$, then $m = 1$ and a solution of the integral equation can be found very simply in a formal way. As shown above $k \leq m - 1$, and for this case $k = 0$, so that $g_0(x,y)$ is a solution of the integral equation. From (183.4),

$$g_0(x,y) = \lim_{\lambda \to \lambda_1} D(x,y,\lambda) = D(x,y,\lambda_1)$$

Consequently, we have the result that if $D(\lambda_1) = 0$, but $D'(\lambda_1) \neq 0$, then for a fixed value of $y = y_0$,

$$u_1(x) = D(x,y_0,\lambda_1) \tag{183.6}$$

is a continuous solution not identically zero of the integral equation

$$u(x) = \lambda_1 \int_a^b K(x,t)u(t)dt$$

Along with this result, it is evident that there will exist continuous solutions for a discrete set of values of λ, these values being the roots of $D(\lambda) = 0$, and there will exist a countable set of corresponding solutions.

When $D(\lambda) = 0$ has multiple roots it can be shown that there exist just q independent solutions of the homogeneous integral equation corresponding to the root λ_1, where q is the index of this root. Under these conditions the method becomes very cumbersome and the solutions will not be given here.

Example. Let it be required to obtain a solution of the homogeneous integral equation

$$u(x) = \frac{1}{\pi} \int_0^{2\pi} \sin x \sin t \, u(t)dt \tag{183.7}$$

In the first place we must verify that $\lambda = 1/\pi$ is a root of $D(\lambda)$, or the equation has no continuous solution. From (182.15), we find $D(\lambda)$ to be

$$\begin{aligned} D(\lambda) &= 1 - \lambda \int_0^{2\pi} \sin^2 t \, dt \\ &\quad + \frac{\lambda^2}{2!} \int_0^{2\pi} \int_0^{2\pi} \begin{vmatrix} \sin t_1 \sin t_1 & \sin t_1 \sin t_2 \\ \sin t_2 \sin t_1 & \sin t_2 \sin t_2 \end{vmatrix} dt_1 \, dt_2 \\ &\quad - \cdots \end{aligned}$$

It is easy to see that all terms involving determinants vanish, since on appropriate factoring, two rows will always turn out to be identical. Consequently

$$D(\lambda) = 1 - \pi\lambda$$

and $\lambda = 1/\pi$ is a root of $D(\lambda) = 0$. On using the series (182.17) for the calculation of $D(x,y,1/\pi)$, it is found that

$$\begin{aligned} D\left(x,y,\frac{1}{\pi}\right) &= \frac{1}{\pi} \sin x \sin y - \left(\frac{1}{\pi}\right)^2 \int_0^{2\pi} \begin{vmatrix} \sin x \sin y & \sin x \sin t \\ \sin t \sin y & \sin t \sin t \end{vmatrix} dt \\ &\quad + \cdots \\ &= \frac{1}{\pi} \sin x \sin y \end{aligned}$$

since for the same reason as mentioned above all terms involving determinants vanish. Consequently, from eq. (183.6) the solution of the equation is

$$u(x) = D\left(x,y,\frac{1}{\pi}\right) = \frac{1}{\pi}\sin x \sin y$$

for any specific value of y. Thus there is an infinite number of solutions of the equation.

184. Homogeneous integral equations of the second kind with symmetric kernels. The method of solution to be considered here is that mentioned under 4 in Sec. 179. The general theory of this method, which applies when the kernel is symmetric, is due to Schmidt and Hilbert. As a basis for understanding the method a formal, heuristic development of the theory will be given.

We seek a continuous solution of the integral equation

$$u(x) = \int_a^b K(x,t)u(t)dt \tag{184.1}$$

where $K(x,t)$ is continuous over a certain rectangle and in addition $K(x,t)$ is symmetric, so that

$$K(x,t) = K(t,x)$$

In accordance with the theory of the previous section, it will be supposed that the equation has solutions $u_1(x)$, $u_2(x)$, ..., corresponding to a discrete set of values of λ, namely, λ_1, λ_2, ..., so that

$$u_n(x) = \lambda_n \int_a^b K(x,t)u_n(t)dt \tag{184.2}$$

We shall now make the hypothesis that the kernel $K(x,t)$ can be represented by an expansion in terms of $u_n(x)$ in the form

$$K(x,y) = \sum_n \frac{u_n(x)u_n(y)}{\lambda_n} \tag{184.3}$$

It will be noted that the symmetry of $K(x,y)$ has been maintained. In the above equation, replace y by t, multiply through by $K(t,y)$, and integrate with respect to t over the range ab. This yields

$$\int_a^b K(t,y)K(x,t)dt = \int_a^b K(t,y) \sum_n \frac{u_n(x)u_n(t)}{\lambda_n}\, dt$$

which in turn can be rearranged in the successive forms

$$\begin{aligned}\int_a^b K(x,t)K(t,y)dt &= \sum_n \frac{u_n(x)}{\lambda_n} \int_a^b K(t,y)u_n(t)dt \\ &= \sum_n \frac{u_n(x)}{\lambda_n} \int_a^b K(y,t)u_n(t)dt \\ &= \sum_n \frac{u_n(x)u_n(y)}{\lambda_n^2}\end{aligned} \tag{184.4}$$

In this rearrangement the symmetry property of the kernel has been used as well as the integral eq. (184.2) itself. The term on the left side of the above equation is usually called the first iterated kernel and denoted by $K^2(x,y)$. If the same process is carried out on eq. (184.4) as was performed on eq. (184.3), the result could be written as

$$\int_a^b K(t,y)K^2(x,t)dt = \int_a^b K(t,y) \sum_n \frac{u_n(x)u_n(t)}{\lambda_n^2} dt$$

On rearranging as before, this equation can be written

$$\int_a^b K^2(x,t)K(t,y)dt = \sum_n \frac{u_n(x)}{\lambda_n^2} \int_a^b K(y,t)u_n(t)dt$$

or

$$K^3(x,y) = \sum_n \frac{u_n(x)u_n(y)}{\lambda_n^3} \tag{184.5}$$

Thus it is seen that the pth iterated kernel is

$$K^{p+1}(x,y) = \int_a^b K^p(x,t)K(t,y)dt \tag{184.6}$$

and

$$K^m(x,y) = \sum_n \frac{u_n(x)u_n(y)}{\lambda_n^m} \tag{184.7}$$

If we think of the eigen-values λ_n ordered according to increasing values, then for very large values of m we should expect that the main contribution to the right side of eq. (184.7) would be furnished by the first term, so that approximately

$$K^m(x,y)\lambda_1^m \approx u_1(x)u_1(y) \tag{184.8}$$

It follows from this result that

$$\frac{1}{\lambda_1^m} \approx \frac{\int_a^b K^m(x,x)dx}{\int_a^b [u_1(x)]^2 dx} \tag{184.9}$$

Then since

$$\frac{\lambda_1^m}{\lambda_1^{m+1}} \approx \frac{\int_a^b K^{m+1}(x,x)dx}{\int_a^b K^m(x,x)dx} \approx \frac{1}{\lambda_1} \tag{184.10}$$

we have an approximate value for the first eigen-value. It can be shown that an exact value for λ_1 can be obtained by passing to the limit $m \to \infty$.

Returning to eq. (184.8), an approximate solution can be obtained, since

$$[u_1(x)]^2 \approx K^m(x,x) \cdot \lambda_1^m \tag{184.11}$$

It can be rigorously shown that an exact solution $u_1(x)$ can be obtained by taking the limit as $m \to \infty$, so that

$$u_1(x) = \pm\sqrt{\lim_{m\to\infty} K^m(x,x)\lambda_1^m} \tag{184.12}$$

The procedure outlined above serves to determine the first or smallest eigen-value λ_1 and its associated solution $u_1(x)$. If there are other eigen-values, one could in principle proceed to determine λ_2 and $u_2(x)$ as follows. One starts by rearranging eq. (184.3) so that all known functions are on the left side. Then

$$K(x,y) - \frac{u_1(x)u_1(y)}{\lambda_1} = \sum_{n=2} \frac{u_n(x)u_n(y)}{\lambda_n} \tag{184.13}$$

By following the same procedure as that followed above for obtaining the first iterated kernel, one can easily find that

$$\int_a^b K(x,t)K(t,y)dt - \frac{u_1(x)u_1(y)}{\lambda_1^2} = \sum_{n=2} \frac{u_n(x)u_n(y)}{\lambda_n^2}$$

Forming the mth iterated kernel, we obtain

$$K^m(x,y) - \frac{u_1(x)u_1(y)}{\lambda_1^m} = \sum_{n=2} \frac{u_n(x)u_n(y)}{\lambda_n^m}$$

As before, for large m it is to be expected that

$$K^m(x,y)\lambda_2^m - u_1(x)u_1(y)\frac{\lambda_2^m}{\lambda_1^m} \approx u_2(x)u_2(y) \tag{184.14}$$

From this it can readily be shown that

$$\frac{\lambda_1}{\lambda_2} \approx \frac{\lambda_1^{m+1}\int_a^b K^{m+1}(x,x)dx - \int_a^b [u_1(x)]^2\,dx}{\lambda_1^m\int_a^b K^m(x,x)dx - \int_a^b [u_1(x)]^2\,dx} \tag{184.15}$$

The exact value of the ratio λ_1/λ_2 can be found by passing to the limit as $m \to \infty$.

From eq. (184.14) an approximate result for $u_2(x)$ is

$$u_2(x) \approx \sqrt{\lambda_2^m K^m(x,x) - \left(\frac{\lambda_2}{\lambda_1}\right)[u_1(x)]^2} \tag{184.16}$$

An exact result for $u_2(x)$ can be had by passing to the limit as $m \to \infty$.

While this method of determining the eigen-values and associated solutions of the homogeneous integral equation with symmetric kernel is convenient for some types of kernel, it becomes very difficult to use for others because of the difficulty in evaluating the iterated kernel and passing to the limit as $m \to \infty$. It is often easier to find the eigen-values by determining the root of $D(\lambda)$ as in the previous section.

Example. As an illustrative example, the method will be applied to the integral (183.7) already solved by the Fredholm method. The equation with symmetric kernel is

$$u(x) = \lambda\int_0^{2\pi} \sin x \sin t\, u(t)dt \tag{184.17}$$

We start by calculating the second iterated kernel

$$\begin{aligned} K^2(x,y) &= \int_0^{2\pi} K(x,t)K(t,y)dt \\ &= \pi \int_0^{2\pi} \sin x \sin^2 t \sin y \, dt \\ &= \lambda \sin x \sin y \end{aligned}$$

Proceeding in this way, the mth iterated kernel is

$$K^m(x,y) = \pi^{m-1} \sin x \sin y$$

Using formula (184.10),

$$\frac{1}{\lambda} \approx \frac{\pi^m \int_0^{2\pi} \sin^2 x \, dx}{\pi^{m-1} \int_0^{2\pi} \sin^2 x \, dx} = \pi$$

Passing to the limit as $m \to \infty$, we of course obtain

$$\lambda = \frac{1}{\pi}$$

as found already in Sec. 183.

Using (184.12), the solution is

$$u(x) = \sqrt{\lim_{m\to\infty} \left(\frac{1}{\pi}\right)^m \pi^{m-1} \sin^2 x} = \frac{\sin x}{\sqrt{\pi}}$$

which agrees with the result found in the last section except for an arbitrary multiplicative constant.

185. Orthonormal character of the solutions of the homogeneous integral equation with symmetric kernel. It may readily be demonstrated that the characteristic solutions of a homogeneous integral equation with a symmetric kernel have the important property of being orthogonal in the sense of Sec. 155. This means that

$$\int_a^b u_n(x)u_m(x)dx = 0 \tag{185.1}$$

where

$$u_n(x) = \lambda_n \int_a^b K(x,t)u_n(t)dt \tag{185.2}$$

and

$$u_m(x) = \lambda_m \int_a^b K(x,t)u_m(t)dt \tag{185.3}$$

Let it be supposed that the eigen-values λ_n and λ_m are not equal. We may then multiply the first equation by $\lambda_m u_m(x)$ and the second by $\lambda_n u_n(x)$, subtract the two equations, and integrate the result from a to b. This gives the result

$$\begin{aligned} &(\lambda_m - \lambda_n) \int_a^b u_n(x)u_m(x)dx \\ &\quad = \lambda_n\lambda_m \int_a^b \int_a^b [K(x,t)u_n(t)u_m(x) - K(x,t)u_m(t)u_n(x)]dt \, dx = 0 \end{aligned}$$

That the right side vanishes can be seen when in the second half of the double integral the variables of integration are interchanged, and it is recalled that $K(x,t) = K(t,x)$. Since $\lambda_n \neq \lambda_m$, we have the result expressed in eq. (185.1).

Not only are the characteristic solutions belonging to different eigenvalues orthogonal, but they form a complete set (see Sec. 156) which can obviously be normalized. Such a complete orthonormal set can be used to provide a framework for the solution of a nonhomogeneous integral equation with symmetric kernel, as shown in the next section.

186. Solution of nonhomogeneous integral equation in terms of the characteristic solutions of the associated homogeneous equation. The equation

$$u(x) = f(x) + \lambda \int_a^b K(x,t)u(t)dt \tag{186.1}$$

can be solved with the help of the orthonormal set of solutions of the homogeneous equation mentioned in the preceding section

$$u_n(x) = \lambda_n \int_a^b K(x,t)u_n(t)dt \tag{186.2}$$

In order to obtain the solution, let

$$v(x) = \frac{u(x) - f(x)}{\lambda} = \sum_{n=1}^{\infty} c_n u_n(x) \tag{186.3}$$

Thus $v(x)$ is represented in terms of an expansion in the orthonormal set of functions $u_n(x)$. In order to determine the unknown coefficients c_n, multiply eq. (186.1) through by $u_n(x)$, and integrate with respect to x from a to b. Then

$$\begin{aligned}
&\int_a^b u(x)u_n(x)dx \\
&\quad = \int_a^b f(x)u_n(x)dx + \lambda \int_a^b u_n(x)dx \int_a^b K(x,t)u(t)dt \\
&\quad = \int_a^b f(x)u_n(x)dx + \lambda \int_a^b u(t)dt \int_a^b K(t,x)u_n(x)dx \\
&\quad = \int_a^b f(x)u_n(x)dx + \frac{\lambda}{\lambda_n} \int_a^b u(t)u_n(t)dt
\end{aligned}$$

or

$$\int_a^b u(x)u_n(x)dx = \lambda_n \frac{\int_a^b f(x)u_n(x)dx}{\lambda_n - \lambda} \tag{186.4}$$

Then by using (186.3) in (186.1), multiplying through by $u_n(x)$, and integrating with respect to x from a to b,

$$\begin{aligned}
\int_a^b v(x)u_n(x)dx &= \int_a^b u_n(x)dx \int_a^b K(x,t)u(t)dt \\
&= \int_a^b u(t)dt \int_a^b K(t,x)u_n(x)dx \\
&= \frac{1}{\lambda_n} \int_a^b u(t)u_n(t)dt
\end{aligned} \tag{186.5}$$

In these manipulations the homogeneous eq. (186.2) has been used, as well as the symmetry of the kernel $K(x,t)$. Combining (186.5), (186.3), and (186.4),

$$c_n = \frac{\int_a^b f(x)u_n(x)dx}{\lambda_n - \lambda} \tag{186.6}$$

This result, in addition to eq. (186.3), yields the solution of (186.1), namely,

$$u(x) = f(x) + \sum_{n=1}^{\infty} \left(\frac{\int_a^b f(x)u_n(x)dx}{\lambda_n - \lambda} \right) u_n(x) \tag{186.7}$$

provided none of the eigen-values λ_n is equal to λ. It can be shown that the above series converges and constitutes a unique and continuous solution if $f(x)$ is continuous.

Solution when $\lambda_k = \lambda$. Here it is assumed that one of the eigen-values of the homogeneous eq. (186.2) is equal to λ of eq. (186.1). For this case we rewrite eq. (186.4) in the form

$$\lambda_n \int_a^b f(x)u_n(x)dx = (\lambda_n - \lambda) \int_a^b u(x)u_n(x)dx$$

From this equation we deduce that for $n = k$,

$$\int_a^b f(x)u_k(x)dx = 0$$

This leaves the constant c_k in (186.6) undetermined. For all other n except $n = k$, the result (186.7) holds; consequently the solution in this case is

$$u(x) = f(x) + c_k u_k(x) + \sum_{n=1}^{\infty}{}' \frac{\int_a^b f(x)u_n(x)dx}{\lambda_n - \lambda} u_n(x) \tag{186.8}$$

where $\sum'$ means that the term $n = k$ is omitted in the sum, and c_k is an arbitrary constant.

A similar procedure can be followed if more than one λ_n is equal to λ. The solution will contain just as many arbitrary constants as the number of λ_n's which equal λ.

Example. Let it be required to solve the integral equation

$$u(x) = x(1 - x) + 2\int_0^1 K(x,t)u(t)dt \tag{186.9}$$

where the kernel $K(x,t)$ is defined as

$$K(x,t) = \begin{matrix} (1 - t)x & 0 \leq x \leq t \\ (1 - x)t & t \leq x \leq 1 \end{matrix} \tag{186.10}$$

It can be readily demonstrated that $K(x,t) = K(t,x)$ and is consequently symmetric [see (187.17)].

In order to carry out the Schmidt-Hilbert procedure, it is necessary first to determine the eigen-values λ_n and corresponding solutions $u_n(x)$ of the associated homogeneous equation

$$u_n(x) = \lambda_n \int_0^1 K(x,t)u_n(t)dt \tag{186.11}$$

Perhaps the easiest way to determine λ_n is to determine the roots of $D(\lambda)$ given by eq. (182.15). In this case

$$\begin{aligned} D(\lambda) = 1 &- \lambda \int_0^1 (1-t)t\,dt \\ &+ \frac{\lambda^2}{2!}\int_0^1 dt_2 \left[\int_0^{t_2} \begin{vmatrix} (1-t_1)t_1 & (1-t_2)t_1 \\ (1-t_1)t_2 & (1-t_2)t_2 \end{vmatrix} dt_1 \right. \\ &\left. + \int_{t_2}^1 \begin{vmatrix} (1-t_1)t_2 & (1-t_1)t_2 \\ (1-t_1)t_2 & (1-t_2)t_2 \end{vmatrix} dt_1 \right] - \cdots \end{aligned}$$

On evaluating the integrals contained in $D(\lambda)$ it is found that

$$D(\lambda) = 1 - \frac{\lambda}{3!} + \frac{\lambda^2}{5!} - \frac{\lambda^3}{7!} + \cdots \tag{186.12}$$

The above series closely resembles that of $\sin\lambda$. Actually it is readily seen that this series is

$$D(\lambda) = \frac{\sin\sqrt{\lambda}}{\sqrt{\lambda}} \tag{186.13}$$

so that the roots of $D(\lambda)$ are

$$\lambda_n = n^2\pi^2 \qquad n = 1, 2, \ldots \tag{186.14}$$

It was shown in Sec. 183 that a solution $u_n(x)$ of (186.11) belonging to the root λ_n is $D(x,y,\lambda_n)$. In order to calculate $D(x,y,\lambda)$, it is most convenient to use the recurrence formula (182.28) for the coefficients $d_m(x,y)$ in the expansion

$$\begin{aligned} D(x,y,\lambda) = d_1(x,y) &+ d_2(x,y)\lambda^2 \\ &+ \cdots + d_m(x,y)\lambda^m + \cdots \end{aligned} \tag{186.15}$$

For convenience this formula is rewritten below.

$$d_m(x,y) = d_{m-1}K(x,y) + \int_0^1 K(x,t)d_{m-1}(t,y)dt \tag{186.16}$$

where the d_m's are the coefficients in the expansion for $D(\lambda)$ [see formula (182.27)]. In our case

$$d_m = \frac{(-1)^m}{(2m+1)!} \tag{186.17}$$

In calculating the coefficients $d_m(x,y)$, we shall assume $x < y$. Then from (182.29) and (186.10), we have

$$d_1(x,y) = K(x,y) = (1-y)x \tag{186.18}$$

For the kernel (186.10), the recurrence formula (186.16) can be written

$$\begin{aligned} d_m(x,y) &= d_{m-1}(1-y)x + \int_0^x (1-x)td_{m-1}(t,y)dt \\ &\quad + \int_x^y (1-t)xd_{m-1}(t,y)dt + \int_y^1 (1-t)xd_{m-1}(t,y)dt \\ &= d_{m-1}(1-y)x + \int_0^x (t-x)d_{m-1}(t,y)dt \\ &\quad + \int_0^y (1-t)xd_{m-1}(t,y)dt + \int_y^1 (1-t)xd_{m-1}(t,y)dt \\ &= \int_0^x (t-x)d_{m-1}(t,y)dt + xF_m(y) \end{aligned} \tag{186.19}$$

Proceeding to calculate $d_2(x,y)$, it is readily seen that

$$\begin{aligned} d_2(x,y) &= \int_0^x (t-x)(1-y)t + xF_2(y) \\ &= -(1-y)\frac{x^3}{3!} + xF_2(y) \end{aligned} \tag{186.20}$$

Again using (186.19) to calculate $d_3(x,y)$, we find that

$$\begin{aligned} d_3(x,y) &= -\int_0^x (t-x)(1-y)\frac{t^3}{3!}dt + xF_3(y) \\ &= (1-y)\frac{x^5}{5!} + xF_3(y) \end{aligned} \tag{186.21}$$

By induction from (186.18), (186.20), and (186.21),

$$d_m(x,y) = (1-y)\frac{x^{2m-1}}{(2m-1)!} + xF_m(y)$$

and from (183.6), (186.14), and (186.15), we obtain

$$u_n(x) = D(x,y,n^2\pi^2)$$

$$= (1 - y)\sum_{m=1}^{\infty} (-n^2\pi^2)^m \frac{x^{2m-1}}{(2m-1)!} + \sum_{m=1}^{\infty} F_m(x)$$

(186.11). It can readily be verified that

$$u_n(x) = \sin n\pi x \tag{186.23}$$

is indeed a solution belonging to the eigen-value $= n^2\pi^2$.

Before going on with the complete solution of eq. (186.9), it is worth while to point out the value of experimenting with eq. (186.11) in an

attempt to solve it before applying the formal method already outlined. For instance, with the kernel expressed in explicit form from (186.10), our integral equation can be written

$$u_n(x) = \lambda_n \left[\int_0^x (1 - x)tu_n(t)dt + \int_x^1 (1 - t)xu(t)dt \right]$$

If this equation be differentiated twice with respect to x,

$$u_n''(x) = -\lambda_n u_n(x) \tag{186.24}$$

The general solution of this differential equation is

$$u_n(x) = A \sin \sqrt{\lambda_n}\, x + B \cos \sqrt{\lambda_n}\, x$$

or since we already know from (186.14) that $\lambda_n = n^2\pi^2$,

$$u_n(x) = A \sin n\pi x + B \cos n\pi x$$

Substituting this in the homogeneous integral eq. (186.11), it is readily seen that the above is a solution only if $B = 0$, making the solution correspond to that already found in eq. (186.23).

Making use of (186.7), the unique solution of the integral eq. (186.9) which we set out to solve is

$$u(x) = x(1 - x) + 2 \sum_{n=1}^{\infty} \frac{\int_0^1 t(1 - t)\sin n\pi t\, dt}{n^2\pi^2 - 2} \sin n\pi x \tag{186.25}$$

187. Setting up the integral equation associated with a boundary value problem. As an illustrative example of the connection of integral equations with physical boundary value problems, the integral equation will be set up which describes the lateral vibration of an elastic string which is held fixed at both ends. In particular we shall assume that the string is of uniform density and obeys Hooke's law. The differential equation which describes the transverse displacement $u(x,t)$ is

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \quad \text{with } c \text{ a constant} \tag{187.1}$$

We shall agree to start the string at time $t = 0$ from a configuration at rest specified by $u(x,0) = g(x)$. The initial conditions are then

$$\left.\begin{aligned} u(x,0) &= g(x) \\ \frac{\partial u(x,0)}{\partial t} &= 0 \end{aligned}\right\} \tag{187.2}$$

The length of the string when not vibrating will be taken as unity, and since the end points are not allowed to move, the boundary conditions are

$$\left.\begin{aligned} u(0,t) &= 0 \\ u(1,t) &= 0 \end{aligned}\right\} \tag{187.3}$$

The solution of the differential equation plus boundary conditions is usually handled as follows. One attempts a solution of the differential equation by expressing the solution as a product of a function of x alone, and a function of t alone. Thus

$$u = X(x)T(t) \tag{187.4}$$

Substituting this in the differential equation and dividing through by XT,

$$\frac{T''}{c^2T} = \frac{X''}{X} = -\lambda = \text{constant} \tag{187.5}$$

These functions of t and x have been set equal to a constant, since the only function of t that can equal a function of x for all x and t over a range is a constant. Consequently the functions T and X satisfy the simple ordinary differential equations

$$X'' + \lambda X = 0 \tag{187.6}$$

$$T'' + c^2\lambda T = 0 \tag{187.7}$$

In addition to satisfying the above differential equation, X must satisfy the boundary conditions

$$X(0) = 0 \qquad X(1) = 0 \tag{187.8}$$

From the point of view of the differential equation the problem has been reduced to the solution of the ordinary differential equation

$$\frac{d^2v}{dx^2} + \lambda v = 0 \tag{187.9}$$

subject to the boundary condition

$$v(0) = 0 \qquad v(1) = 0 \tag{187.10}$$

At first sight it might be thought that the pertinent integral equation might be set up by transforming the differential eq. (187.9) into an integral equation by the method outlined in Sec. 178. If indeed this is tried, one obtains the integral equation

$$v(x) = c_1x + c_2 - \lambda\int_0^x (x - t)v(t)dt \tag{187.11}$$

It is true that the solution $v(x)$ of this integral equation is a solution of the differential eq. (187.9), but no choice of the constants c_1 and c_2 will allow v to satisfy the boundary conditions (187.10) for v not identically zero. The trouble lies in the kernel of the integral equation. The nature of the difficulty can be seen by assuming that v satisfies a homogeneous integral equation

$$v(x) = \lambda\int_0^1 K(x,\xi)v(\xi)d\xi \tag{187.12}$$

and then trying to determine the properties of the undetermined kernel. It can readily be seen that if $v(x)$ is to satisfy such an integral equation and be nontrivial, and at the same time satisfy the boundary conditions, $K(x,\xi)$ must vanish at $x = 0$ and $x = 1$, i.e., satisfy the same boundary conditions as $v(x)$ itself. Clearly the kernel in (187.11) does not. The easiest way to try to form an integral equation of the form (187.12) is to start with the two differential equations

$$\frac{d^2v}{dx^2} = -\lambda v \tag{187.13}$$

$$\frac{d^2K(x,\xi)}{dx^2} = 0 \tag{187.14}$$

and multiply the first by $-K(x,\xi)$ and the second by $v(x)$ and add. This gives

$$v(x)\frac{d^2K(x,\xi)}{dx^2} - K(x,\xi)\frac{d^2v(x)}{dx^2} = \frac{d}{dx}(vK' - Kv') = \lambda K(x,\xi)v(x) \tag{187.15}$$

Integration from 0 to 1 would come rather close to our integral equation, but as yet we have not made $K(x,\xi)$ satisfy the boundary conditions. The solution of the differential eq. (187.14) is

$$K(x,\xi) = \alpha(\xi)x + \beta(\xi)$$

which cannot be made to vanish at $x = 0$ and $x = 1$ unless $K(x,\xi) \equiv 0$. This dilemma can be averted by considering solutions of (187.14) which are continuous but which do not possess a continuous derivative. Such a solution would be

$$\begin{aligned} K(x,\xi) &= (1 - \xi)x \qquad 0 \le x \le \xi \\ K(x,\xi) &= (1 - x) \qquad \xi \le x \le 1 \end{aligned} \tag{187.16}$$

These solutions are of the same form as that given above, but the first derivative is discontinuous at a point $x = \xi$ where $0 < \xi < 1$. However, all derivatives are continuous at other points in the interval. Under these conditions eq. (187.15) can be integrated over each interval, so that

$$(vK' - Kv')_{x=\xi-0} - (vK' - Kv')_{x=0} = \lambda\int_0^{\xi-0} K(x,\xi)v(x)dx$$

$$(vK' - Kv')_{x=1} - (vK' - Kv')_{x=\xi+0} = \lambda\int_{\xi+0}^{1} K(x,\xi)v(x)dx$$

Since v and K vanish at $x = 0$ and $x = 1$ and v, v', and K are continuous functions of x over the interval, adding the two equations gives

$$v(\xi)[K'(\xi-0) - K'(\xi+0)] = \lambda\int_0^1 K(x,\xi)v(x)dx$$

but from (187.16),

$$K'(\xi-0) = 1 - \xi$$
$$K'(\xi+0) = -\xi$$

Consequently their difference is unity, so that the above integral equation becomes

$$v(\xi) = \lambda\int_0^1 K(x,\xi)v(x)dx$$

with $K(x,\xi)$ defined by (187.16). The terms involving ξ in (187.16) were chosen so that the difference in the derivative for $x < \xi$ and $x > \xi$ would be unity.

From (187.16) it can be readily shown that $K(x,\xi)$ is symmetric. To see this let $0 \leq y_1 \leq y_2 \leq 1$.

$$K(y_1,y_2) = (1 - y_2)y_1$$
and $$K(y_2,y_1) = (1 - y_2)y_1$$
so that $$K(y_1,y_2) = K(y_2,y_1) \tag{187.17}$$

As a result of the symmetry, x and ξ can be interchanged in the above integral equation, so that our boundary value problem is entirely equivalent to the integral equation

$$v(x) = \lambda\int_0^1 K(x,\xi)v(\xi)d\xi \tag{187.18}$$

The solution of this integral equation is discussed in the example at the end of Sec. 186.

SUGGESTED REFERENCES

W. V. Lovitt, *Linear Integral Equations*, New York, McGraw-Hill Book Company, Inc., 1924.

M. Bôcher, *An Introduction to the Study of Integral Equations*, London, Cambridge University Press, 1926.

T. Lalesco, *Theorie des equations integrales*, Paris, A. Hermann et Fils, 1912.

R. Courant and D. Hilbert, *Methoden der mathematischen Physik*, I., Berlin, Julius Springer, 1931.

PROBLEMS

Change the following differential equations into integral equations.

178.1. $\frac{d^2y}{dx^2} + y = 0.\ y = 0, \frac{dy}{dx} = 1$ at $x = 0$.

178.2. $\frac{d^2y}{dx^2} + y = \cos x.\ y = 0, \frac{dy}{dx} = 1$ at $x = 0$.

178.3. $\frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 6y = 0.\ y = 0, \frac{dy}{dx} = -1$ at $x = 0$.

180.1. Solve $\frac{1}{1 - x^2/a^2} = \int_{-\infty}^{\infty} \frac{u(t)dt}{1 - (x - t)^2/b^2}$ and discuss result.

180.2. Find the curve along which a particle falls if it starts at the origin and takes a time

$$t(y) = -\frac{1}{\sqrt{g}}\left[F\left(\frac{1}{\sqrt{2}}, \arccos\sqrt{y}\right) - F\left(\frac{1}{\sqrt{2}}, \frac{\pi}{2}\right)\right]$$

to fall a distance y.

180.3. Find the curve along which a particle falls if it starts at the origin and takes a time $t(y) = \sqrt{\frac{2}{g}} \arcsin\sqrt{y}$ to fall a distance y.

181.1. Solve $u = 1 - 2\int_0^x tu(t)dt$.

181.2. Find an approximation to the solution of

$$u = x + \lambda\int_{-1}^{1} (1 - |x - t|)u(t)dt.$$

181.3. Solve

$$u(x) = \sin x - \frac{x}{4} + \frac{1}{4}\int_0^{\pi/2} txu(t)dt.$$

182.1. Solve by the Fredholm method

$$u(x) = x + \lambda\int_{-\pi}^{\pi} \cos(x - t)\, u(t)dt.$$

182.2. Solve by the Fredholm method

$$u(x) = x^3 + \lambda\int_0^1 (x + t)u(t)dt.$$

183.1. Use the Fredholm method to find the eigen-values and solutions of

$$u(x) = \lambda\int_{-1}^{1} (x - t)^2 u(t)dt.$$

183.2. Find eigen-values and solutions of

$$u(x) = \lambda\int_{-\pi}^{\pi} \sin^2 x \sin^2 t\, u(t)dt.$$

184.1. Using the Schmidt-Hilbert method, find an eigen-value and eigen-function of the kernel

$$K(x,y) = (x + y)^3 \qquad \begin{matrix} -1 \le x \le 1, \\ -1 \le y \le 1. \end{matrix}$$

184.2. Using the Schmidt-Hilbert method, find an eigen-value and eigen-function of

$$K(x,y) = \cos(x - y) \qquad \begin{matrix} -\pi \le x \le \pi, \\ -\pi \le y \le \pi. \end{matrix}$$

186.1. Use the Schmidt-Hilbert method to find the solution of

$$u(x) = x^2 - 1 + \lambda\int_{-1}^{1} (x - t)^2 u(t)dt.$$

Hint. Note that

$$\int_{-1}^{1} x^l P_n(x)dx = 0 \quad \text{for} \quad l < n.$$

187.1. Find the integral equation corresponding to

$$\frac{d^2u}{dx^2} + \lambda u = 0$$

with boundary conditions $u(-l) = 0$, $u(l) = 0$.

187.2. Find the integral equation corresponding to

$$\frac{1}{r}\frac{d}{dx}\left(x\frac{du}{dx}\right) + \lambda u = 0$$

with boundary conditions $u(0) = 1$, $u(1) = 0$.

Chapter 15

VARIATIONAL METHODS

188. Typical problems of the calculus of variations. Many of the laws of physics can be expressed in terms of variational principles. For example, in classical mechanics, it has been shown that the motion of a mechanized system is always such that the time integral of the Lagrangian function $L(q,\dot{q},t)$ taken between two configurations of the system has an extreme value. Actually, in general the motion is such that

$$J = \int_{t_1}^{t_2} L(q,\dot{q},t)dt \tag{188.1}$$

is a minimum, so that this principle has come to be known as that of least action. Thus from a mathematical point of view, the problem of determining the motion of a system is that of determining the position coordinates as those functions of the time which will make the above integral a minimum. This is a typical problem in the calculus of variations.

Another such principle is that of Fermat, which states that the propagation of light always takes place in such a way that the actual optical path, i.e., length of geometric path multiplied by the index of refraction of the medium, is an extreme value. If the index of refraction varies continuously from point to point in the medium, Fermat's principle states that

$$J = \int_{p_1}^{p_2} n(x,y,z)ds = \text{extremum} \tag{188.2}$$

From the standpoint of the calculus of variations, functions $y(x)$, $z(x)$ must be found which make the above integral take on an extreme value, which is usually its minimum value.

Many other physical laws can be expressed in terms of a variational principle, but the examples above are sufficient to illustrate the general mathematical problem involved.

Many problems receive their initial formulation directly in the domain of the calculus of variations. Examples of simple ones in this category include determination of the form of the curve in a plane, which

will make the distance between points in the plane a minimum, and determination of the meridian curve connecting two points in a plane which, when revolved around an axis, gives the surface of revolution with the minimum area. In these two problems we require those particular functions $y(x)$ which make the following two integrals a minimum.

$$J = \int_{x_1}^{x_2} \sqrt{1 + y'^2}\, dx = \text{minimum} \tag{188.3}$$

$$J = 2\pi \int_{x_1}^{x_2} y\sqrt{1 + y'^2}\, dx = \text{minimum} \tag{188.4}$$

subject to the conditions that $y_1 = y(x_1)$ and $y_2 = y(x_2)$.

189. Nature of the simplest variational problem. Many variational problems are included under the particularly simple type in which we seek to determine a function $y(x)$ which makes the integral

$$J[y] = \int_{x_1}^{x_2} F(x,y,y')dx \tag{189.1}$$

a minimum. The notation $J[y]$ indicates that the value of the integral J depends on the argument function $y(x)$ for a specified function F. The problem with which we are confronted bears some resemblance to the problem of finding the minimum or maximum of a simple function $f(x)$ (see Sec. 21). In the latter case, a necessary condition for an extremum is that the differential df vanish at the point. The problem encountered in the calculus of variations, however, is much more difficult, because instead of finding a point where a simple function $f(x)$ has a minimum, we are required to determine a function $y(x)$ out of an infinity of possible functions defined over an interval of x which makes an integral J a minimum. A further very fundamental difference between the maximum and minimum problem and the calculus of variations problem lies in the fact that a theorem of Weierstrasz always insures the solution of the straight maximum-minimum problem, while no such general fundamental existence theorem exists for the variational calculus problem. Consequently, in the calculus of variations the existence of a solution to any given extremum problem requires a special proof.

In those cases where a solution exists, it will turn out that a necessary condition for an extremum is that the first variation of $J[y]$ must be zero. This is written

$$\delta J[y] = 0 \tag{189.2}$$

This requirement is formally analogous to the requirement for the straight minimum problem that the differential df vanish. It remains for us to define the meaning of the first variation of an integral.

190. The first variation and the Euler equation. We shall proceed to determine the necessary condition for the integral

$$J[y] = \int_{x_1}^{x_2} F(x,y,y')dx \tag{190.1}$$

to have an extremum. It will be assumed that $x_1,x_2,y_1 = y(x_1)$, and $y_2 = y(x_2)$ are given, and that the function F possesses continuous second-order partial derivatives with respect to x, y, and y'. It will also be supposed that $y''(x)$ is to be continuous.

It is convenient to start by assuming that we know the function which actually makes J a minimum and call it $y_0(x)$ so that

$$J[y_0] = \text{a minimum}$$

and consider a group of functions $y(x)$ which vary somewhat from $y_0(x)$, in which case

$$J[y] \geq J[y_0]$$

The particular group of functions $y(x)$ will be taken to be those defined by

$$y(x) = y_0(x) + \epsilon\eta(x) = y_0 + \delta y \tag{190.2}$$

where $\eta(x)$ possesses a continuous second derivative and vanishes at x_1 and x_2, so that $y(x) = y_0(x)$ at the end point of the curve $y(x)$. Here ϵ is regarded as a small quantity, so that δy denotes a small variation of the function from the minimal function $y_0(x)$. Placing (190.2) in the integral of (190.1) gives a function of ϵ only, namely,

$$\Phi(\epsilon) = \int_{x_1}^{x_2} F(x,y+\epsilon\eta,y'+\epsilon\eta')dx \tag{190.3}$$

Since by definition the integral takes on its minimum value when $\epsilon = 0$, it follows that

$$\Phi'(0) = 0 = \int_{x_1}^{x_2} \left(\frac{\partial F}{\partial y}\eta + \frac{\partial F}{\partial y'}\eta'\right)dx$$

Since $\eta(x_1) = \eta(x_2) = 0$, an integration by parts of the second term can be carried out, and the above integral becomes

$$\Phi'(0) = \int_{x_1}^{x_2} \left(\frac{\partial F}{\partial y} - \frac{d}{dx}\frac{\partial F}{\partial y'}\right)\eta\, dx = 0 \tag{190.4}$$

If this equation be multiplied through by ϵ, then on the right side $\epsilon\eta$ by (190.2) can be written as δy, the variation of the function y. If, then, the quantity $\epsilon\Phi(0)$ on the left side be termed δJ, this defines the first variation of the original integral corresponding to the variation of the function $y(x)$. The necessary condition (190.4) for an extremum for the integral J is that the first variation of J vanish, i.e.,

$$\delta J = \int_{x_1}^{x_2} \left(\frac{\partial F}{\partial y} - \frac{d}{dx}\frac{\partial F}{\partial y'}\right)\delta y\, dx = 0 \tag{190.5}$$

It is readily seen that if $\Phi(\epsilon)$ defined in (190.3) is expanded in a power

series in ϵ, the term containing the first power of ϵ, i.e., $\epsilon \left\{\frac{d\Phi}{d\epsilon}\right\}_{\epsilon=0}$ is just δJ. This is the reason for calling it the first variation of J. Terms in higher powers of ϵ are called second, third, etc., variations of J.

Since $\eta(x)$ is quite arbitrary except for its values at x_1 and x_2, the quantity δy is also arbitrary, and it follows that the integral in (190.4) or (190.5) can vanish for all η or δy only if

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\frac{\partial F}{\partial y'} = 0 \tag{190.6}$$

When the total derivative of $\partial F/\partial y'$ with respect to x is taken, the above equation becomes

$$y'' \frac{\partial^2 F}{\partial y'^2} + y' \frac{\partial^2 F}{\partial y\, \partial y'} + \frac{\partial^2 F}{\partial x\, \partial y'} - \frac{\partial F}{\partial y} = 0 \tag{190.7}$$

This is Euler's differential equation. That this equation must be satisfied by $y(x)$ constitutes a necessary condition for the existence of an extremum. Every solution of this equation fitting the boundary conditions is called an extremal or stationary function of the minimum problem. This differential equation arising out of the first variation of $J[y]$ plays the same role in the calculus of variations minimum problem as the ordinary differential quotient plays in the simple minimum problem in ordinary calculus.

There are several circumstances in which it is simple to obtain the solution of the Euler equation. These cases will be pointed out in the following.

Case 1; F Explicitly Independent of x. In this case the following identity holds

$$\frac{d}{dx}\left(F - y' \frac{\partial F}{\partial y'}\right) \equiv y'\left(\frac{\partial F}{\partial y} - \frac{d}{dx}\frac{\partial F}{\partial y'}\right)$$

However, from condition (190.6), the right side is identically zero, so that

$$F - y' \frac{\partial F}{\partial y'} = \text{constant} \tag{190.8}$$

Thus in this case a first integral of Euler's equation can be obtained directly. To proceed further, the actual function $F(y,y')$ must be known.

Case 2; F Independent of y. In this case, eq. (190.6) reduces to

$$\frac{d}{dx}\frac{\partial F}{\partial y'} = 0$$

or

$$\frac{\partial F}{\partial y'} = \text{constant} \tag{190.9}$$

so that a first integral can be obtained directly here also.

Case 3; F Independent of y'. In this case, we would have the algebraic equation

$$\frac{\partial F}{\partial y} = 0 \tag{190.10}$$

to solve.

Case 4; F Independent of x and y. In this case, the Euler eq. (190.7) reduces to

$$y'' \frac{\partial^2 F}{\partial y'^2} = 0$$

In case $\partial^2 F/\partial y'^2$ does not vanish identically, we have the simple differential equation

$$y'' = 0 \tag{190.11}$$

to solve.

Example 1. The foregoing method will be applied to the problem of determining the plane curve passing through the points (x_1,y_1) and (x_2,y_2) which shall have the shortest length. The length of such a curve [formula (188.3)] is

$$J[y] = \int_{x_1}^{x_2} \sqrt{1 + y'^2}\, dx \tag{190.12}$$

This corresponds to Case 4 above, since $F = \sqrt{1 + y'^2}$ is independent of x and y. Further, since $\partial^2 F/\partial y'^2$ is not identically zero, our problem reduces to that of solving the differential equation

$$y'' = 0$$

subject to the conditions that the curve pass through the points specified above. The solution of the above equation is obviously

$$y = c_1 x + c_2$$

where the constants c_1 and c_2 are determined so that the straight line represented by the above equation passes through the points y_1 and y_2. Thus

$$y_1 = c_1 x_1 + c_2 \qquad y_2 = c_1 x_2 + c_2$$

Finally

$$y = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) + y_1 \tag{190.13}$$

Consequently, the shortest distance between the indicated points is a straight line connecting them, and verification of this fact, as seen above, is a simple problem in the calculus of variations.

Example 2. *The Brachystochrone.* The time required for a particle to descend from the origin to the point (x_1,y_1) on a frictionless curved path lying in a plane is

$$t = \int_0^{x_1} \sqrt{\frac{1 + y'^2}{2gy}}\, dx \tag{190.14}$$

This relation was essentially derived in Sec. 177. The above formula can be obtained from that in Sec. 177 by replacing the element of length ds by $\sqrt{1 + y'^2}\,dx$, where the curve is specified by $y(x)$. The

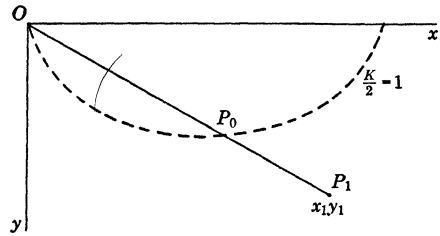


FIG. 190.1

problem considered here is that of determining the function $y(x)$ for which the time of descent will be a minimum when the curve must pass through the points (0,0) and (x_1,y_1) (see Fig. 190.1).

In this case

$$F(x,y,y') = F(y,y') = \sqrt{\frac{1 + y'^2}{2gy}} \tag{190.15}$$

and consequently our problem falls under Case 1 above, so that $y(x)$ must satisfy the equation

$$y' \frac{\partial F}{\partial y'} - F = -c_1$$

or, from (190.15), this equation reduces to

$$c_1\sqrt{2gy(1 + y'^2)} = 1$$

Upon further reduction, this equation reduces to

$$y' = \pm\sqrt{\frac{k - y}{y}} \quad \text{where} \quad k = \frac{1}{c_1^2 2g}$$

This can be integrated, and the result is

$$x = +\int \frac{y\,dy}{\sqrt{ky - y^2}} \tag{190.16}$$

The proper sign to choose here is, of course, the positive one. By using the method of Sec. 28, the above integral can be readily integrated. Perhaps the simplest way to integrate the above integral is to write it

$$x = -\frac{1}{2}\int_0^y \frac{k - 2y}{\sqrt{ky - y^2}}\,dy + \frac{k}{2}\int_0^y \frac{dy}{\sqrt{ky - y^2}}$$

The first integral can be integrated at once, and the second one can be integrated easily by making the substitution

$$\frac{k}{2} - y = \frac{k}{2}\cos\theta$$

The final result is

$$x = -\sqrt{ky - y^2} + \frac{k}{2}\cos^{-1}\frac{k - 2y}{k} \tag{190.17}$$

This is the equation of the extremal or the curve of minimum time of descent. It still remains for us to select k so that the curve goes through the point x_1,y_1. To do this it is simplest to use the parametric equations of the curve, which are found by letting $k/2 - y = (k/2)\cos\theta$. Then

$$x = \frac{k}{2}(\theta - \sin\theta) \tag{190.18}$$

$$y = \frac{k}{2}(1 - \cos\theta) \tag{190.19}$$

These are the parametric equations of a cycloid. First draw the cycloid for which $k/2 = 1$, as shown in the dotted curve of Fig. 190.1. Then draw a straight line cutting it and connecting the origin and the point (x_1,y_1). Since all cycloids are similar, the one passing through (x_1,y_1) will be that one for which

$$\frac{k}{2} = \frac{OP_1}{OP_0}$$

Consequently, we have found the path down which a particle would slide without friction to the point (x_1,y_1) in the shortest time.

191. Variational problems involving several unknown functions. In the above discussion the variational problem involving only one unknown function of a single independent variable was considered. The required generalization to cover problems wherein more than one function of a single independent variable is involved can readily be made. If the unknown functions are $y_1(x)$, $y_2(x)$, ..., the problem before us is to determine these functions such that the integral

$$J[y_1,y_2,\ldots] = \int_{x_1}^{x_2} F(x,y_1,y_2,\ldots,y_1',y_2',\ldots)dx \tag{191.1}$$

will have a stationary value.

As in the preceding section, one assumes that $y_{01}(x)$, $y_{02}(x)$, etc. are the functions which yield the extremum, and consider the variations

$$\begin{aligned} y_1 &= y_{01} + \epsilon_1\eta_1(x) & y_2 &= y_{02} + \epsilon_2\eta_2(x) \ldots \\ &= y_{01} + \delta y_1 & &= y_{02} + \delta y_2 \ldots \end{aligned} \tag{191.2}$$

Then the integral (184.1) becomes

$$\begin{aligned} \Phi(\epsilon_1,\epsilon_2,\ldots) &= J[y_{01}+\delta y_1, y_{02}+\delta y_2,\ldots] \\ &= \int_{x_1}^{x_2} F(x, y_{01}+\epsilon_1\eta_1, y_{02}+\epsilon_2\eta_2,\ldots, \\ &\qquad y_{01}'+\epsilon_1\eta_1', y_{02}'+\epsilon_2\eta_2',\ldots)dx \end{aligned}$$

Here Φ is a function of several variables ϵ_1, ϵ_2, etc., and by hypothesis Φ has an extremum when $\epsilon_1 = \epsilon_2 = \ldots = 0$. Consequently, Φ has an extremum when

$$\frac{\partial \Phi}{\partial \epsilon_1}\Big]_{\epsilon_1=0} = \frac{\partial \Phi}{\partial \epsilon_2}\Big]_{\epsilon_2=0} = \ldots = 0$$

Proceeding just as for the one-functional case, this leads to the following system of Eulerian equations, the simultaneous satisfaction of which constitutes a necessary condition for the required extremum

$$\left.\begin{aligned} \frac{d}{dx}\frac{\partial F}{\partial y_1'} - \frac{\partial F}{\partial y_1} &= 0 \\ \frac{d}{dx}\frac{\partial F}{\partial y_2'} - \frac{\partial F}{\partial y_2} &= 0 \end{aligned}\right\} \tag{191.3}$$

In terms of the first variation of J, which in this case is

$$\delta J = \epsilon_1 \left(\frac{\partial \Phi}{\partial \epsilon_1}\right)_{\epsilon_1=0} + \epsilon_2 \left(\frac{\partial \Phi}{\partial \epsilon_2}\right)_{\epsilon_2=0} + \ldots$$

the necessary condition for an extremum can be expressed by the condition that

$$\delta J = 0$$

as in the one-functional case.

192. Variational problems involving several independent variables. Many physical problems require us to determine a function of several independent variables which will give rise to an extremum in the case of a multiple integral. The following two-variable cases will serve to illustrate the type of problem which is to be handled. The integral to be considered is

$$J[u] = \int_D \int F(x,y,u,u_x,u_y)dx\,dy \tag{192.1}$$

where $u(x,y)$ is the function of two variables which is to be determined such that J is an extremum and $u_x = \partial u/\partial x$ and $u_y = \partial u/\partial y$. The function F is defined over the domain D of x and y, and is assumed to have continuous second-order partial derivatives.

As in the preceding sections, one starts with a function $u_0(x,y)$ which makes $J[u]$ a minimum and then considers a variation of this function of the form

$$\begin{aligned} u(x,y) &= u_0(x,y) + \epsilon\eta(x,y) \\ &= u_0(x,y) + \delta u \end{aligned} \tag{192.5}$$

where $\eta(x,y)$ is arbitrary except that it vanishes on a curve C forming the boundary of the domain D. In accordance with the preceding sec-

tions, we form δJ, the first variation of J, and set it equal to zero. Thus

$$\begin{aligned}\delta J &= \epsilon\left[\frac{d}{d\epsilon}\Phi(\epsilon)\right]_{\epsilon=0} = \epsilon\left[\frac{d}{d\epsilon}J(u_0+\epsilon\eta)\right] \\ &= \epsilon\int_D\int\left[\frac{\partial F}{\partial u}\eta + \frac{\partial F}{\partial u_x}\frac{\partial \eta}{\partial x} + \frac{\partial F}{\partial u_y}\frac{\partial g}{\partial y}\right]dx\,dy = 0\end{aligned} \tag{192.3}$$

We shall assume that the curve C bounding D has a piecewise continuously turning tangent, so that Green's theorem (Sec. 59) can be used to make a useful transformation of the above integral. Upon application of Green's theorem to the last two terms of the above integral, it becomes

$$\begin{aligned}\delta J = \epsilon\int_D\int \eta\left[\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}\frac{\partial F}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial F}{\partial u_y}\right]dx\,dy \\ + \epsilon\int_C \eta\left[\frac{\partial F}{\partial u_x}dy - \frac{\partial F}{\partial u_y}dx\right] = 0\end{aligned} \tag{192.4}$$

Since η is assumed to vanish on the curve C, the right integral vanishes. Also since η is otherwise arbitrary, it follows that the part of the integrand in brackets in the left integral must vanish identically. Consequently,

$$\frac{\partial}{\partial x}\frac{\partial F}{\partial u_x} + \frac{\partial}{\partial y}\frac{\partial F}{\partial u_y} - \frac{\partial F}{\partial u} = \frac{0}{0} \tag{192.5}$$

When this equation is written out in detail, we have the following partial differential equation.

$$\begin{aligned}\frac{\partial^2 F}{\partial u_x^2}\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 F}{\partial u\,\partial u_x}\frac{\partial u}{\partial x} + \frac{\partial^2 F}{\partial u_y\,\partial u_x}\frac{\partial^2 u}{\partial x\,\partial y} + \frac{\partial^2 F}{\partial x\,\partial u_x} + \frac{\partial^2 F}{\partial u_y^2}\frac{\partial^2 u}{\partial y^2} \\ + \frac{\partial^2 F}{\partial u\,\partial u_y}\frac{\partial u}{\partial y} + \frac{\partial^2 F}{\partial u_x\,\partial u_y}\frac{\partial^2 u}{\partial y\,\partial x} + \frac{\partial^2 F}{\partial y\,\partial u_y} - \frac{\partial F}{\partial u} = 0\end{aligned} \tag{192.6}$$

This is Euler's partial differential equation for the variational problem involving two independent variables. A necessary condition for an extremum of $J[u]$ is that u satisfy this equation. If $u(x,y)$ is required by conditions of the original problem to have specified values on the boundary C, a solution of the Euler equation must be obtained which satisfies the given boundary conditions.

Example. The Vibrating String. An example of this type of variational problem is provided by using the principle of least action mentioned at the beginning of Sec. 188 to derive the differential equation governing the motion of a vibrating string. According to eq. (188.1), the motion of the system will be that for which the integral

$$J[q] = \int_{t_0}^{t_1} L(q,\dot{q},t)dt \tag{192.7}$$

has a stationary value, where L is the Lagrangian function or merely the difference between the kinetic and potential energies of the system, i.e.,

$$L = T - U \tag{192.8}$$

In order to determine these quantities for a string, it will be supposed that initially an elastic string is stretched along the x axis from $x = 0$ to $x = l$. The string will be given a small lateral displacement which at each point along the x axis will be denoted by $u(x,t)$. If ρ is the mass of the string per unit length, the kinetic energy of the string is given by

$$T = \frac{1}{2}\int_0^l \rho u_t^2 \, dx \qquad u_t = \frac{\partial u}{\partial t} \tag{192.9}$$

when lengthwise motion of the string is neglected.

If the displacement of the string is small, the change in length of the string occasioned by the displacement is given by

$$\begin{aligned}\int_0^l \sqrt{1 + u_x^2}\, dx - l &\approx \int_0^l \left(1 + \frac{1}{2}u_x^2\right)dx - l \\ &= \frac{1}{2}\int_0^l u_x^2 \, dx \end{aligned} \tag{192.10}$$

Since the potential energy will be proportional to the elongation, the potential energy is

$$U = \frac{k}{2}\int_0^l u_x^2 \, dx$$

Thus

$$L = T - U = \frac{1}{2}\int_0^l (\rho u_t^2 - k u_x^2)dx$$

and eq. (192.7) becomes

$$J[u] = \int_{t_0}^{t_1}\int_0^l \left(\frac{1}{2}\rho u_t^2 - \frac{1}{2}k u_x^2\right)dx\, dt \tag{192.11}$$

In this case the functional F does not depend explicitly on x, t, or u, since

$$F(u_x,u_t) = \frac{1}{2}(\rho u_t^2 - k u_x^2)$$

The first variation δJ will be zero when the Euler differential eq. (192.6) is satisfied. On using the above function $F(u_x,u_t)$ this Euler equation reduces to

$$\rho\frac{\partial^2 u}{\partial t^2} - k\frac{\partial^2 u}{\partial x^2} = 0 \tag{192.12}$$

which is the familiar differential equation for the vibrating string.

193. Potential theory and the associated variational problem. Potential problems involving Laplace's equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{193.1}$$

can be replaced by a variational problem. To find the equivalent variational problem it is necessary only to determine the function F in the integral (192.1) for which the Euler equation is the same as eq. (193.1). It is readily seen that if

$$F(x,y,u,u_x,u_y) = \frac{1}{2}(u_x^2 + u_y^2)$$

the Euler partial differential eq. (192.6) reduces to Laplace's eq. (193.1) in two independent variables. Consequently, the potential function u is that function which makes the integral

$$J[u] = \int\int_D \frac{1}{2}(u_x^2 + u_y^2)dx\,dy$$

an extremum, i.e., makes the first variation of J vanish. Hence u is the function for which $\delta J = 0$.

194. Variational problems with an auxiliary condition. It is often required to determine the function $u(x)$ which will make the integral

$$J[y] = \int_{x_1}^{x_2} F(x,y,y')dx \tag{194.1}$$

an extremum subject to the auxiliary condition that

$$\int_{x_1}^{x_2} G(x,y,y')dx = C \tag{194.2}$$

It may be shown that this problem is equivalent to the problem already considered in Sec. 190, namely, that of finding the function $y(x)$ which will yield an extremum for the integral of a modified function $F^*(x,y,y')$.

$$J^*[y] = \int_{x_1}^{x_2} F^*(x,y,y')dx \tag{194.3}$$

without regard to the auxiliary condition (194.2). The modified function F^* is given by

$$F^* = F(x,y,y') + \lambda G(x,y,y') \tag{194.4}$$

where the constant λ must be determined from the auxiliary condition (194.2). This method breaks down for the exceptional case where

$$\frac{d}{dx}\frac{\partial G}{\partial y'} - \frac{\partial G}{\partial y} = 0$$

Example. Probably one of the simplest examples which will illustrate the operations indicated above is that of determining the curve of shortest length between two points (x_1,y_1) and (x_2,y_2) such that the area under the curve is A.

From Sec. 188, the length of the curve is

$$J[y] = \int_{x_1}^{x_2} \sqrt{1 + y'^2}\, dx \tag{194.5}$$

and the area under it is

$$\int_{x_1}^{x_2} y\, dx = A \tag{194.6}$$

According to the previous part of this section our problem reduces to that of finding the extremum of

$$J^*[y] = \int_{x_1}^{x_2} (\sqrt{1 + y'^2} + \lambda y) dx \tag{194.7}$$

From Sec. 190 the necessary condition for the first variation of $J^*[y]$ to vanish is that eq. (190.6) be satisfied for this case, or

$$\frac{d}{dx}\frac{\partial}{\partial y'}(\sqrt{1 + y'^2} + \lambda y) = \frac{\partial}{\partial y}(\sqrt{1 + y'^2} + \lambda y)$$

or

$$\frac{d}{dx}\frac{y'}{\sqrt{1 + y'^2}} = \lambda$$

This differential equation can be integrated twice, after which y is given by the equation

$$(y - c_2)^2 + \left(x - \frac{c_1}{\lambda}\right)^2 = \frac{1}{\lambda^2} \tag{194.8}$$

Thus the curve of shortest length, the area under which is A, is a circle with center at the point $(c_1/\lambda, c_2)$ and a radius of $1/\lambda$. The constants of integration c_1 and c_2, together with the constant λ, can be determined from the condition that the curve passes through the points (x_1,y_1) and (x_2,y_2) and the condition furnished by eq. (194.6).

The variational problem handled in Sec. 191 can be generalized to include an auxiliary condition also. Specifically it is required to determine the functions $y_1(x)$, $y_2(x)$, ..., which yield an extremum for the integral

$$J[y_1,y_2,\ldots] = \int_{x_1}^{x_2} F(x,y_1,y_2,\ldots,y_1',y_2',\ldots)dx \tag{194.9}$$

subject to the auxiliary condition

$$G(x,y_1,y_2,\ldots) = 0 \tag{194.10}$$

where the values of $y_1(x_1)$, $y_1(x_2)$, $y_2(x_1)$, $y_2(x_2)$, etc., are prescribed. The straightforward procedure would be to solve for one of the unknown functions in terms of x and the remaining y's, and substitute in the integrand of (194.9), thus eliminating one of the unknown functions in the original integral and proceeding exactly as outlined in Sec. 191. However, this is not always easy to do, and it is more convenient to replace F in (194.9) by the modified function F^* defined as follows.

$$F^* = F + \lambda G$$

The resulting variational problem is just that handled in Sec. 191 without an auxiliary condition. In this case λ is generally not constant but a function of x.

Example. There are many examples of this type of variational problem. A simple example is the determination of the curve of shortest length between two points on a sphere when the whole curve is confined to the surface of a sphere. The length of a space curve when the coordinates y and z are expressed as functions of x is given by

$$J[y,z] = \int_{x_1}^{x_2} \sqrt{1 + y'^2 + z'^2}\, dx \tag{194.11}$$

and if this curve is to lie on a sphere of radius r, then x, y, and z are related by the equation

$$x^2 + y^2 + z^2 - r^2 = 0 \tag{194.12}$$

Our minimum problem is then equivalent to finding the minimum of $J^*[y,z]$ where

$$J^*[y,z] = \int_{x_1}^{x_2} [\sqrt{1 + y'^2 + z'^2} + \lambda(x^2 + y^2 + z^2 - r^2)]dx \tag{194.13}$$

The method of handling this variational problem is given in Sec. 191. From eq. (191.3), the Euler differential equations are

$$\begin{aligned} \frac{d}{dx}\frac{y'}{\sqrt{1 + y'^2 + z'^2}} - 2\lambda y &= 0 \\ \frac{d}{dx}\frac{z'}{\sqrt{1 + y'^2 + z'^2}} - 2\lambda z &= 0 \end{aligned} \tag{194.14}$$

From these two equations and eq. (194.12), the unknown functions $y(x)$, $z(x)$, and $\lambda(x)$ can be determined. There will also be enough arbitrary constants so that the curve can be made to pass through the points x_1,y_1,z_1 and x_2,y_2,z_2 on the surface of the sphere.

195. Direct methods of obtaining solutions of variational problems. In the preceding sections, the problem of determining a function which makes a certain integral a minimum has been replaced by the problem of obtaining a solution of the Euler differential equation. It often happens, however, that the Euler equation or equations, as the case may be, are difficult to solve and it is easier to come to grips with the determination of the argument function which makes the integral a minimum in a more direct way. In fact, for some differential equations especially associated with characteristic value problems which can be replaced by a variational problem, it is possible at least to obtain an approximate solution by direct methods of solving the equivalent variational problem.

For the purpose of illustrating a method of procedure for obtaining a direct solution of a variational problem we shall consider the variational problem discussed in Sec. 189. We are required to determine the function $y(x)$ which makes the integral

$$J[y] = \int_{x_1}^{x_2} F(x,y,y')dx \tag{195.1}$$

a minimum. The method to be used can equally well be generalized to apply to integrals involving functions with more derivatives or to multiple integrals and the like.

In all direct methods of solving variational problems of this type one attempts to set up a sequence of functions $y_1, y_2, y_3, \ldots$, such that

$$\lim_{n\to\infty} J[y_n] = d \tag{195.2}$$

an extremum of J, and

$$\lim_{n\to\infty} y_n = y_0 \tag{195.3}$$

the actual solution of the variational problem. If d is the minimum value of $J[y]$, the sequence $y_1, y_2, y_3, \ldots$ which satisfies (195.2) is called a minimal sequence. It is usually possible to set up such a minimal sequence by a method outlined below, but whether the minimal sequence thus set up will converge to the actual solution y_0 remains a question which must be investigated in particular cases. Examples can be given where the variational problem has a definite solution, but the minimal sequence does not converge to it. Nevertheless the method is useful for numerical calculations of the minimum value d and for determining the solution y_0 approximately, even though the convergence of the minimal sequence to y_0 has not been proved (see second example in Sec. 196).

It is hardly necessary to point out that the functions y_n must be defined over the same domain as y and satisfy the same boundary or auxiliary conditions as y. It is characteristic of the individual functions y_k of the minimal sequence that

$$J[y_k] = d_k \geq d \tag{195.4}$$

when the sequence is established for the solution of a minimum problem.

196. Construction of the minimal sequence. A particularly convenient method of setting up a minimal sequence is known as the Ritz procedure. The first step is to select a set of linearly independent functions $\varphi_1, \varphi_2, \ldots, \varphi_n$ which have the same boundary or auxiliary conditions which y is required to satisfy. One then represents the function $y(x)$ approximately by a linear combination

$$y_n = c_1\varphi_1 + c_2\varphi_2 + \ldots + c_n\varphi_n \tag{196.1}$$

of the functions $\varphi_1, \varphi_2, \ldots, \varphi_n$ in accordance with Sec. 156. These functions might be a part of a complete orthonormal set for the domain of definition of y. The cleverness with which this set of functions is chosen with respect to their adaptability to the problem determines the rapidity of convergence of the method. In a physical problem, a knowledge of the general nature of the solution is of enormous benefit in selecting the most appropriate set of functions.

If each φ satisfies the boundary conditions irrespective of the value of the constants c_1, c_2 in (196.1), then y_n satisfies the imposed boundary conditions. To determine the constants c_1 to c_n, the term y_n as given by (196.1) is substituted for y in $F(x,y,y')$ of eq. (195.1). The integral can be evaluated, and the result $J_n(c_1,c_2,\ldots,c_n)$ will be a known function of the constants $c_1, c_2, \ldots, c_n$. One is then confronted with an ordinary problem in maxima and minima to determine the values of the c's such that $J[y_n]$ will be an extremum. Consequently the c's are to be determined from the n equations.

$$\frac{\partial J_n}{\partial c_1} = 0, \quad \frac{\partial J_n}{\partial c_2} = 0, \quad \ldots, \quad \frac{\partial J_n}{\partial c_n} = 0 \tag{196.2}$$

as outlined in Sec. 21. If this system of equations possesses the unique solution $c_1 = \gamma_1, c_2 = \gamma_2, \ldots, c_n = \gamma_n$, the best possible representation of $y(x)$, using the set of n functions $\varphi_1, \varphi_2, \ldots, \varphi_n$ will be

$$y(x) \approx y_n(x) = \gamma_1\varphi_1 + \gamma_2\varphi_2 + \ldots + \gamma_n\varphi_n \tag{196.3}$$

In many physical problems $J_n(c_1,c_2,\ldots,c_n)$ turns out to be a quadratic form in the constants c, so that eqs. (196.2) are ordinary linear equations.

The minimal set of functions in this case is made up of the functions

$$\left.\begin{aligned} y_1 &= \gamma_1\varphi_1 \\ y_2 &= \gamma_1\varphi_1 + \gamma_2\varphi_2 \\ &\cdots\cdots\cdots\cdots \\ y_n &= \gamma_1\varphi_1 + \gamma_2\varphi_2 + \ldots + \gamma_n\varphi_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}\right\} \tag{196.4}$$

Also $\lim_{k\to\infty} J_k[y_k] = d =$ the minimum of $J[y]$. It would further be expected that $\lim_{k\to\infty} y_k = y_0$ would yield the solution to the minimum problem, but as pointed out in the preceding section this cannot always be guaranteed.

The above method of setting up the minimal set for the direct solution of the variational problem (195.1) can readily be extended to cover the case where the argument function y must satisfy certain auxiliary conditions such as those considered in Sec. 194. In this case the constants $c_1, c_2, \ldots, c_n$ must be made to satisfy the auxiliary equation in

addition to eqs. (196.2). This minimum problem with auxiliary conditions has been treated in Sec. 22.

Example 1. As an example of the typical procedure used in applying the above method, we shall consider a variational problem discussed by Ritz. It will be required to determine the function $y(x)$, which makes the integral

$$J[y] = \int_a^b [p(x)y'^2 + 2q(x)yy' + r(x)y^2]dx \tag{196.5}$$

a minimum, subject to the conditions that $y'(x)$ shall be finite and continuous in the interval ab, and $y(a) = A$ and $y(b) = B$. It will be assumed that p, q, and r, together with their first derivatives, are finite and continuous in the interval ab, and that the quadratic form

$$py'^2 + 2qyy' + ry^2$$

is positive definite.

It will be more convenient to transform the problem to one where the unknown function has simple boundary conditions. This may be done by letting

$$y(x) = \frac{x-b}{a-b}A + \frac{x-a}{b-a}B + u(x) \tag{196.6}$$

By requiring $u(x)$ to satisfy the boundary conditions $u(a) = u(b) = 0$, the function $y(x)$ automatically satisfies the proper boundary condition $y(a) = A$ and $y(b) = B$ as required. Replacing y in (196.5) by its expression in terms of u given above,

$$\begin{aligned} J[u] = \int_a^b \Bigg\{ & pu'^2 + 2quu' + ru^2 \\ & + \left[2q\left(\frac{B-A}{b-a}\right) + 2p\left(\frac{x-b}{a-b}A + \frac{x-a}{b-a}B\right)\right]u' \\ & + \left[2q\left(\frac{B-A}{b-a}\right) + 2r\left(\frac{x-b}{a-b}A + \frac{x-a}{b-a}B\right)\right]u \\ & + p\left(\frac{B-A}{b-a}\right)^2 + 2q\left(\frac{x-b}{a-b}A + \frac{x-a}{b-a}B\right)\frac{B-A}{b-a} \\ & + r\left(\frac{x-b}{a-b}A + \frac{x-a}{b-a}B\right)^2\Bigg\}dx \end{aligned}$$

If the first bracket of the integrand is integrated by parts, keeping in mind the boundary conditions on u, and the result is combined with the second bracket of the integrand, the above integral may be written as follows.

$$J[u] = \int_a^b (pu'^2 + 2quu' + ru^2 + 2Fu)dx + \text{constant} \tag{196.7}$$

where

$$F(x) = (r - q')\frac{x-b}{a-b}A + \frac{x-a}{b-a}B - p'\frac{B-A}{b-a} \tag{196.8}$$

and the constant term in (196.7) arises from integrating the subsequent terms of the integrand, which include all known functions of x. The constant term is of no consequence as far as the variational problem is concerned, so that the original variational problem is replaced by one which requires us to find $u(x)$ which makes the integral

$$J[u] = \int_a^b (\rho u'^2 + 2quu' + ru^2 + 2Fu)dx \tag{196.9}$$

a minimum.

Since $u(a) = u(b) = 0$, it would appear convenient to use the sine functions to form the minimal set. With this choice

$$\varphi_k = \sin \frac{k\pi(x - a)}{b - a}$$

would satisfy the boundary conditions for all k. Using these linearly independent functions over the interval ab, the prospective minimal set becomes

$$u_1 = c_1 \sin \frac{\pi(x - a)}{b - a}$$

$$u_2 = c_1 \sin \frac{\pi(x - a)}{b - a} + c_2 \sin \frac{2\pi(x - a)}{b - a}$$

. . . . . .

$$u_n = c_1 \sin \frac{\pi(x - a)}{b - a} + c_2 \sin \frac{2\pi(x - a)}{b - a} + \ldots + c_n \sin \frac{n\pi(x - a)}{b - a}$$

When these functions are substituted in (196.9) and the constants $c_1 \ldots c_n$ determined so as to make $J[u]$ a minimum, it can be proved that the resulting functions form a minimal set, and

$$\lim_{k \to \infty} J_k[u_k] = d = \text{minimum}$$

and

$$\lim_{k \to \infty} u_k = u_0$$

is the unique solution of the variational problem.

This particular example will not be pursued further, since it would be necessary to know the actual functions p, q, and r in order to proceed. The above discussion, however, will indicate how one goes about using the Ritz method of solving a variational problem directly.

Example 2. As another example of the use of the minimal set of functions to obtain a direct solution of a variational problem, we shall consider the variational problem associated with the motion of a vibrating string as discussed in Sec. 192. The integral that must have a stationary value is

$$J[u] = \int_{t_0}^{t_1} \int_0^l \left(\frac{1}{2}u_t^2 - \frac{1}{2}ku_x^2\right)dt\,dx \tag{196.10}$$

as given by (192.11). For our present purpose we shall attempt to determine the amplitude of vibration of a string as a function of the distance along it when the ends are held fixed and the remainder vibrates harmonically with the time. Our problem will therefore be that of determining the function $y(x)$ when

$$u(x,t) = y(x)\sin \omega t \tag{196.11}$$

and

$$y(0) = y(l) = 0 \tag{196.12}$$

for all values of t.

When one substitutes (196.11) in (196.10) and performs an integration with respect to t over one period of the motion, i.e., from $t = 0$ to $t = 2\pi/\omega$, one obtains

$$J[u] = J[y] = \frac{\pi}{2\omega}\int_0^l [\omega^2 u^2(x) - ky'^2(x)]dx \tag{196.13}$$

It will be convenient to consider a string of length 2 by letting $\xi = 2x/l - 1$, whereupon

$$J[y] = \frac{\rho\omega l\pi}{4}\int_{-1}^{1}\left[y^2(\xi) - \frac{k}{\rho\omega^2}y'^2(\xi)\right]d\xi \tag{196.14}$$

or

$$J[y] = \frac{\rho\omega l\pi}{4}J'[y] \tag{196.15}$$

where

$$J'[y] = \int_{-1}^{1}[y^2 - K^2y'^2]d\xi \tag{196.16}$$

and

$$K^2 = \frac{k}{\rho\omega^2} \tag{196.17}$$

The boundary condition (196.12) must of course be replaced by the condition

$$y(-1) = y(1) = 0 \tag{196.18}$$

The linearly independent set of functions φ_n we shall attempt to use for setting up the minimal sequence of functions will be powers of ξ, so that the minimal sequence will be a sequence of polynomials. It is convenient to set the minimal sequence up so that each function of the sequence satisfies the boundary condition (196.18). This may readily be done by choosing the nth member of the sequence as

$$y_n(\xi) = (1 - \xi^2)(c_0 + c_1\xi^2 + c_2\xi^4 + \ldots + c_n\xi^{2n}) \tag{196.19}$$

The choice of even powers of ξ limits us to the determination of an amplitude function $y(\xi)$ which is even. Thus even solutions of the problem can be determined, and any odd solutions can be found by choosing a polynomial with odd powers of ξ.

To further illustrate the method we shall compute the function $y_1(\xi)$ which gives the stationary value of $J_1'[y]$, i.e., J_1'. Replacing y in (196.16) by $y_1(\xi)$, one obtains

$$J_1' = 8\left[\left(\frac{2}{15} - \frac{1}{3}K^2\right)c_0^2 + \left(\frac{4}{105} - \frac{2}{15}K^2\right)c_0c_1 + \left(\frac{2}{315} - \frac{11}{105}K^2\right)c_1^2\right] \tag{196.20}$$

The values of c_0 and c_1 which will make J_1' an extremum are determined from eqs. (196.2) or

$$\frac{\partial J_1'}{\partial c_0} = 2\left(\frac{2}{15} - \frac{1}{3}K^2\right)c_0 + \left(\frac{4}{105} - \frac{2}{15}K^2\right)c_1 = 0$$

$$\frac{\partial J_1'}{\partial c_1} = \left(\frac{4}{105} - \frac{2}{15}K^2\right)c_0 + 2\left(\frac{2}{315} - \frac{11}{105}K^2\right)c_1 = 0$$

These equations can be simplified somewhat numerically, so that

$$\left(1 - \frac{5}{2}K^2\right)c_0 + \left(\frac{1}{7} - \frac{1}{2}K^2\right)c_1 = 0$$

$$\left(1 - \frac{7}{2}K^2\right)c_0 + \left(\frac{1}{3} - \frac{11}{2}K^2\right)c_1 = 0 \tag{196.21}$$

For nontrivial solutions of these equations to exist, the determinant of the coefficients must vanish (see Sec. 133). This determines K. The determinant in question is

$$\begin{vmatrix} \left(1 - \frac{5}{2}K^2\right) & \left(\frac{1}{7} - \frac{1}{2}K^2\right) \\ \left(1 - \frac{7}{2}K^2\right) & \left(\frac{1}{3} - \frac{11}{2}K^2\right) \end{vmatrix} = 0$$

or

$$63K^4 - 28K^2 + 1 = 0 \tag{196.22}$$

Solving this equation yields the values

$$K^2 = \frac{14 \pm 11.5325}{63}$$

For the larger of the two values of K^2,

$$\frac{1}{K^2} = 2.46744$$

For this particular problem one may obtain the exact solution by solving the Euler differential equation corresponding to this variational problem. This differential equation has been given [see eq. (192.12)], and is

$$\frac{\rho\partial^2 u}{\partial t^2} - \frac{k\partial^2 u}{\partial x^2} = 0$$

Substituting the type of solution already used, namely, (196.11) gives the differential equation

$$y'' + \frac{k}{\rho\omega^2}y = 0 \quad \text{or} \quad y'' + \frac{1}{K^2}y = 0 \tag{196.23}$$

The solution of this equation which satisfies the boundary conditions (196.18) is obviously

$$y = \cos\frac{x}{K}$$

where $$\frac{1}{K} = \frac{n\pi}{2} \qquad n \text{ an odd integer}$$

The numerical value of $1/K^2$ from this exact result for $n = 1$ is 2.46740. Comparing this with the value found above by the minimal sequence method, it is seen that even with the first minimal function, the value of K^2 is determined with remarkable accuracy. The reader should go ahead and determine the ratio c_1/c_0 from (196.21) and compare the resulting function $y_1(\xi)$ with the exact solution $c_0 \cos(x/K)$. To obtain values of K corresponding to $n = 3, 5, \ldots$, etc., it would be necessary to use minimal functions of higher order to obtain equivalent accuracy.

SUGGESTED REFERENCES

R. Courant and D. Hilbert, *Methoden der mathematischen Physik*, I., Berlin, Julius Springer, 1931.

A. R. Forsyth, *Calculus of Variations*, London, Cambridge University Press, 1927.

Oskar Bolza, *Lectures on the Calculus of Variations*, Chicago, The University of Chicago Press, 1904.

G. A. Bliss, *Calculus of Variations*, La Salle, Ill., Open Court Publishing Company, 1925.

PROBLEMS

190.1. The distance between two infinitely near points on a sphere is given by

$$ds = \sqrt{a^2\,d\theta^2 + a^2\sin^2\theta\,d\varphi^2}.$$

Show that the shortest curve between two points on a sphere is a great circle.

190.2. Find the curve of shortest length between two points on the surface of a cone, and show that if the surface is rolled out onto a plane, the curve becomes a straight line.

190.3. Show that the shortest curve from the point (0,0) to the curve $y = f(x)$ must satisfy the equation

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0,$$

and the conditions

$$F - y'\frac{\partial F}{\partial y'} + \frac{\partial F}{\partial y'}f'\Big|_0^x = 0.$$

190.4. A light ray travels in a region whose index of refraction is given by $n = e^x$. Find the path taken by the ray in going between two points on the y axis.

191.1. The kinetic and potential energies of a spherical pendulum are given by

$$T = \frac{1}{2}\left[l^2\left(\frac{d\theta}{dt}\right)^2 + l^2\sin^2\theta\left(\frac{d\varphi}{dt}\right)^2\right],$$

$$V = mgl(1 - \cos\theta).$$

Find the equations of motion.

191.2. If the kinetic energy of a symmetrical top is given in terms of Euler's angles by

$$T = \frac{1}{2}\left[I\left(\frac{d\theta}{dt}\right)^2 + I\left(\sin\theta\,\frac{d\varphi}{dt}\right)^2 + J\left(\frac{d\chi}{dt} + \cos\theta\,\frac{d\varphi}{dt}\right)^2\right],$$

find the equation of motion of (1) a top rotating in the absence of external forces, (2) a top rotating under the force of gravity, so that its potential energy is $U = mgh\cos\theta$.

191.3. If the distance between two points in a space of n dimensions is given by

$$ds^2 = g_{11}(dx_1)^2 + 2g_{12}\,dx_1\,dx_2 + \ldots = \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}\,dx_i\,dx_j,$$

show that the geodesics in this space must satisfy

$$\frac{d^2x_i}{ds^2} + \sum_{k=1}^{n}\sum_{l=1}^{n}\left\{\begin{matrix} i \\ kl \end{matrix}\right\}\frac{dx_l}{dx}\frac{dx_k}{ds},$$

where

$$\left\{\begin{matrix} i \\ kl \end{matrix}\right\} = \frac{1}{2}\sum_{j=1}^{n} g^{ij}\left(\frac{dg_{jl}}{dx_k} + \frac{dg_{jk}}{dx_l} - \frac{dg_{lk}}{dx_j}\right),$$

and where g^{ij} is defined by $\sum_{j=1}^{n} g^{ij}g_{jk} = \delta_{ik}$.

192.1. The potential energy per unit length of a beam is given by $EI/2R^2$, where E is Young's modulus, I is the moment of inertia of the cross section of the beam, and R is its radius of curvature. Find, in the linear approximation, the equation of motion and boundary conditions of a conical mast which is set firmly in the ground.

192.2. Verify that the Lagrangian

$$L = \frac{1}{8\pi}(E^2 - B^2) - \rho\varphi + \frac{1}{c}J\cdot A$$

leads to Maxwell's equations.

194.1. A cylinder containing a volume V of liquid rotates about its axis with an angular velocity ω. The liquid assumes a configuration which makes its potential energy a minimum. Find the shape of the surface of the liquid.

194.2. Find the shape of a uniform string of fixed length hanging under its own weight from two points on a horizontal plane.

194.3. Find the equation, and the boundary conditions for the shape of a bow of uniform cross section when strung. See Problem 192.1.

196.1. Obtain, by a direct method, using about three functions, an approximation to the lowest frequency of vibration of the mast of Problem 192.1.

196.2. By using a sequency of polynomials, suitably normalized, find an approximation to the shape of the bow of Problem 194.3.

Chapter 16

ELEMENTS OF PROBABILITY THEORY

197. Preliminaries. The occurrence or nonoccurrence of most events cannot be predicted with certainty. The small uncontrollable variations in the conditions under which seemingly identical experiments are performed introduce a degree of uncertainty which cannot be eliminated. The theory of probability recognizes the existence of this uncertainty and attempts to attach numerical "weights" (probabilities) to various events. From the purely mathematical point of view the theory of probability is nothing but a consistent system whereby certain numbers are assigned to certain events. The applicability of probability theory to physical reality rests on the empirically verifiable fact that often the theoretically computed probabilities approximate very well the observed frequencies with which the events actually occur.

198. Probabilities assigned to events. The procedure adopted in probability theory is the following. Probabilities are assigned to certain basic events and then rules are set up whereby probabilities of more complicated events can be calculated. Theoretical predictions can then be tested experimentally. If our predictions are not borne out by experiment, we might be forced to reconsider the assignments of probabilities to basic events and the rules underlying our calculations.

As an illustration, let us consider an ideal gas composed of N identical particles. Let the mass of each particle be m, and let $v_x^{(i)}$, $v_y^{(i)}$, $v_z^{(i)}$ denote the velocity components (at a given time instant) of the ith particle. Assuming that the gas is an isolated system, we can assert that its energy E must remain constant. Furthermore, the assumption that the gas is ideal implies that the total energy E is the kinetic energy:

$$\frac{m}{2}\sum_{i=1}^{N}[v_x^{(i)2} + v_y^{(i)2} + v_z^{(i)2}] = E$$

The state of the gas is now characterized by a point P on the surface of the $3N$-dimensional sphere of radius $\sqrt{2E/m}$. We now assume that the probability that the representative point P should lie in a region

(on the sphere) of "area" S is proportional to this area. Since it is customary to normalize the probabilities in such a way that the probability of the point falling somewhere on the surface is 1, we obtain

$$\text{probability of } P \text{ falling into a region of "area" } S = \frac{S}{\text{total "area" of the sphere}}$$

By this assumption we have assigned probabilities to certain basic events. These basic events are the events that the representative point falls within regions on the sphere.

The rules for manipulating with these probabilities are simply the rules of manipulating with areas of regions on the sphere.

Let us, for instance, calculate the probability that a specified velocity component, say $v_x^{(1)}$, should lie between α and β. We must calculate the area of the zone

$$\alpha < v_x^{(1)} < \beta$$

For a sphere this area can be calculated by taking the derivative with respect to the radius of the volume of the portion of the sphere in which

$$\alpha < v_x^{(1)} < \beta$$

The volume can be obtained by integrating from α to β the volume of the section $v_x^{(1)} = v$. This section is a $(3N - 1)$-dimensional sphere of radius $\sqrt{2E/m - v^2}$, and its volume* is

* The volume of an n-dimensional sphere can be calculated as follows.

$$\begin{aligned} V_n(r) &= \int \cdots \int_{x_1^2+\cdots+x_n^2 \le r^2} dx_1 dx_2 \cdots dx_n = \\ &= \int_{-r}^{r} dx_n \int \cdots \int_{x_1^2+\cdots+x_{n-1}^2 \le r^2 - x_n^2} dx_1 \cdots dx_{n-1} \\ &= \int_{-r}^{r} V_{n-1}(\sqrt{r^2 - x_n^2}) dx_n \\ &= 2\int_0^r V_{n-1}(\sqrt{r^2 - x_n^2}) dx_n \end{aligned}$$

It follows by introducing new variables $y_1 = rx_1, \ldots, y_n = rx_n$, that $V_n(r) = \gamma_n r^n$ and consequently $V_{n-1}\sqrt{r^2 - x^2} = \gamma_{n-1}(r^2 - x^2)^{(n-1)/2}$. Substituting these expressions in the recurrence formula above,

$$\gamma_n = 2\gamma_{n-1}\int_0^1 (1 - x^2)^{(n-1)/2} dx = 2\gamma_{n-1}\int_0^{\pi/2} \cos^n \theta \, d\theta$$

Since the area of a circle of radius r is πr^2, we have $\gamma_2 = \pi$. Thus

$$\gamma_n = \gamma_{n-1}\sqrt{\pi}\frac{\Gamma[(n+1)/2]}{\Gamma(n/2+1)}, \qquad \gamma_n = \pi$$

and it follows easily that

$$\gamma_n = \frac{\pi^{(n/2)}}{\Gamma(n/2+1)}$$

$$C_N\left(\frac{2E}{m} - v^2\right)^{(3N-1)/2}$$

Consequently the desired volume is

$$C_N\int_\alpha^\beta \left(\frac{2E}{m} - v^2\right)^{(3N-1)/2} dv$$

Differentiating with respect to the radius $\sqrt{2E/m}$, we obtain that the area of the zone is

$$C_N \frac{3N-1}{2}\sqrt{\frac{2E}{m}}\int_\alpha^\beta \left(\frac{2E}{m} - v^2\right)^{(3N-3)/2} dv$$

Finally, the desired probability is

$$\text{prob. } [\alpha < v_x^i < \beta] = \frac{\int_\alpha^\beta (2E/m - v^2)^{(3N-3)/2}\, dv}{\int_{-\sqrt{2E/m}}^{\sqrt{2E/m}} (2E/m - v^2)^{(3N-3)/2}\, dv}$$

Assuming E to be proportional to N, that is, $E = \kappa N$, we have

$$\left(\frac{2E}{m} - v^2\right)^{(3N-3)/2} \backsim \left(\frac{2E}{m}\right)^{(3N-3)/2} e^{-3mv^2/2\kappa}$$

and consequently

$$\text{prob. } [\alpha < v_x^{(i)} < \beta] \backsim \frac{\sqrt{3m}}{2\sqrt{\pi\kappa}}\int_\alpha^\beta e^{-3mv^2/2\kappa}\, dv$$

We thus obtain the well-known Maxwell-Boltzmann distribution.

199. Mutually exclusive events. One of the fundamental rules of probability theory is the following. If $E_1, E_2, \ldots$ are *mutually exclusive* events, the probability that at least one of them will occur is the sum of their individual probabilities. Symbolically,

$$\begin{aligned}&\text{prob. } \{E_1 \text{ or } E_2 \text{ or } E_3 \text{ or etc.}\}\\ &= \text{prob. } \{E_1\} + \text{prob. } \{E_2\} + \text{prob. } \{E_3\} + \ldots\end{aligned}$$

In the foregoing example, probability was identified with surface area, and this rule is clearly satisfied. In various special situations our rule can be similarly justified. In general, however, we must accept this rule as an axiom.

200. Independence. A very important notion is that of *independence* of events. Intuitively speaking, two (or more) events are independent if the occurrence or nonoccurrence of one does not influence the occurrence or nonoccurrence of the other.

We should now like to discuss the following problem. If the events E_1 and E_2 are independent and if we know the probabilities prob. $\{E_1\}$

and prob. $\{E_2\}$ assigned to them, what is the probability prob. $\{E_1$ and $E_2\}$ of the simultaneous occurrence of E_1 and E_2?

Since independence precludes the existence of any "interaction" between the events E_1 and E_2, we can assume that prob. $\{E_1$ and $E_2\}$ should be completely determined by the knowledge of the individual probabilities prob. $\{E_1\}$ and prob. $\{E_2\}$. If there were any "interaction" between E_1 and E_2, then clearly we would have to know something about its nature before we could hope to calculate prob. $\{E_1$ and $E_2\}$. We can thus postulate the existence of an *universal* function $F(u,v)$ such that whenever E_1 and E_2 are independent,

$$\text{prob. } \{E_1 \text{ and } E_2\} = F[\text{prob. } \{E_1\}, \text{prob. } \{E_2\}]$$

Imagine now, that a die and a coin are tossed independently and that the probabilities associated with the numbers 1, 2, ... , 6 on the die are $p_1, p_2, \ldots, p_6$, and those associated with heads or tails, α and β. We have, for instance,

$$\begin{aligned} &\text{prob. } \{1 \text{ and } h\} = F(p_1,\alpha) \\ &\text{prob. } \{2 \text{ and } h\} = F(p_2,\alpha) \\ &\text{prob. } \{1 \text{ or } 2 \text{ and } h\} = F(p_1+p_2,\alpha) \end{aligned}$$

On the other hand, the event $\{1$ or 2 and $h\}$ is equivalent to the event $\{1$ and h or 2 and $h\}$, and since the events $\{1$ and $h\}$ and $\{2$ and $h\}$ are mutually exclusive, we have by the axiom of Sec. 199,

$$\begin{aligned} &\text{prob. } \{1 \text{ or } 2 \text{ and } h\} \\ &= \text{prob. } \{1 \text{ and } h\} + \text{prob. } \{2 \text{ and } h\} \end{aligned}$$

Thus
$$F(p_1+p_2,\alpha) = F(p_1,\alpha) + F(p_2,\alpha) \tag{200.1}$$

Apart from the conditions

$$p_i \geq 0, \quad \alpha \geq 0, \quad \beta \geq 0, \quad \sum_1^6 p_i = 1, \quad \alpha + \beta = 1$$

the numbers p_1 and p_2 and α can be quite arbitrary (we can think of different "loaded" dice and coins). Keeping α fixed and differentiating with respect to p_1,

$$\frac{\partial F(p_1+p_2,\alpha)}{\partial p_1} = \frac{\partial F(p_1,\alpha)}{\partial p_1}$$

Thus, as long as $p_1 \geq 0$, $p_2 \geq 0$, $\alpha \geq 0$, $\alpha \leq 1$, $p_1 + p_2 \leq 1$, we have

$$\frac{\partial F}{\partial p_1} = f(\alpha)$$

or
$$F(p_1,\alpha) = f(\alpha)p_1 + C$$

Substituting back into (200.1),

$$(p_1 + p_2)f(\alpha) + C = (p_1 + p_2)f(\alpha) + 2C$$

or $C = 0$.

If we assume that our universal function F is also symmetric, i.e., $F(u,v) = F(v,u)$, we get

$$F(\alpha,p_1) = \alpha f(p_1) = F(p_1,\alpha) = p_1 f(\alpha)$$

or $f(\alpha) = D\alpha$, where D is a constant. Thus, $F(u,v) = Duv$, and the normalizing constant D can be easily seen to be 1. In fact,

$$\begin{aligned} Dp_1 = F(p_1,1) &= \text{prob. } \{1 \text{ and } h \text{ or } t\} \\ &= \text{prob. } \{1\} = p_1 \end{aligned}$$

Finally, $$F(u,v) = uv$$

Generalizing the above considerations to an arbitrary number of events, we are led to the following important rule. If the events E_1, E_2, ... , E_n are independent,

$$\begin{aligned} &\text{prob. } \{E_1 \text{ and } E_2 \text{ and } \dots \text{ and } E_n\} \\ &= \text{prob. } \{E_1\} \text{ prob. } \{E_2\} \dots \text{ prob. } \{E_n\} \end{aligned} \tag{200.2}$$

From the purely mathematical point of view, (200.2) constitutes a definition of independent events. The considerations of this section show that the physical notion of independence leads in a natural manner to this definition which embodies the basic law of multiplication of probabilities.

201. Random variables. Suppose that an experiment can lead to k results which are characterized by the numbers a_1, a_2, ... , a_k (for instance, the outcomes of a single throw of a die are 1, 2, ... , 6). Suppose furthermore that probabilities p_1, p_2, ... , p_k are assigned to the respective results. We have

$$p_i \geq 0, \qquad \sum_1^k p_i = 1$$

The outcome of an experiment can be denoted by a symbol X, where X can be any number a_1, a_2, ... , a_k. Furthermore we have

$$\text{prob. } \{X = a_i\} = p_i$$

The symbol X is called a "random variable."* Consider now another experiment which can have outcomes b_1, b_2, ... , b_k with corresponding probabilities q_1, q_2, ... , q_k. The outcome of this experiment can be symbolized by the random variable Y, for which

$$\text{prob. } \{Y = b_j\} = q_j \tag{201.1}$$

If the experiments are performed independently, the random variables

* This, of course, is not a precise definition. A precise definition cannot be given without appealing to the notion of measure. For the purposes of our exposition an intuitive notion will suffice.

associated with them are called independent. Mathematically, this means that

$$\text{prob. } \{X = a_i, Y = b_j\} = \text{prob. } \{X = a_i\} \text{ prob. } \{Y = b_j\}$$
$$= p_i q_j$$

If the experiments are not independent, the probability of the event $\{X = a_i, Y = b_j\}$ is some number p_{ij} which is not necessarily the product $p_i q_j$. The numbers p_{ij} are not quite arbitrary. Apart from the usual requirement $p_{ij} \geq 0$, we must clearly have also

$$\sum_{j=1}^{l} p_{ij} = p_i, \qquad \sum_{i=1}^{k} p_{ij} = q_j \tag{201.2}$$

It is possible to define operations on random variables. If, for instance, X is a random variable capable of assuming values $a_1, a_2, \ldots, a_k$ with respective probabilities $p_1, p_2, \ldots, p_k$, then X^2 is a random variable capable of assuming the values $a_1^2, a_2^2, \ldots, a_k^2$. To determine the corresponding probabilities we write

$$\text{prob. } \{X = a_j^2\} = \text{prob. } \{X = a_j\} + \text{prob. } \{X = -a_j\}$$

If $-a_j$ is not among $a_1, a_2, \ldots, a_k$, we have

$$\text{prob. } \{X = a_j^2\} = p_j$$

If $-a_j = a_s$,

$$\{\text{prob. } \{X = a_j^2\} = p_j + p_s$$

If X and Y are the random variables defined above and if

$$\text{prob. } \{X = a_i, Y = b_j\} = p_{ij}$$

it is easy to define the random variable $X + Y$. In fact, $X + Y$ can assume the values $a_i + b_j$; and

$$\text{prob. } \{X + Y = \alpha\} = \sum p_{ij}$$

where the summation is extended over all indexes i, j for which $a_i + b_j = \alpha$. The above considerations apply, of course, to an arbitrary number of random variables.

202. Expectations. Let X be a random variable capable of assuming the values $a_1, \ldots, a_k$ with respective probabilities $p_1, p_2, \ldots, p_k$. The mathematical expectation $E\{X\}$ of X is defined by the formula

$$E(X) = \sum_{i=1}^{k} a_i p_i$$

We see that $E\{X\}$ is just a weighted average. One often denotes the mathematical expectation by $\langle X \rangle_{av}$ and refers to it just as the average of X. If $X_1, X_2, \ldots, X_n$ are random variables (not necessarily independent),

$$E\{X_1 + X_2 + \ldots + X_n\} = E\{X_1\} + \ldots + E\{X_n\} \tag{202.1}$$

We carry out the proof for the two random variables X and Y considered in Sec. 201. For the sake of simplicity let us assume that all sums $a_i + b_j$ are distinct. Then

$$\text{prob. } \{X + Y = a_i + b_j\} = \text{prob. } \{X = a_i,\ Y = b_j\} = p_{ij}$$

and consequently

$$E\{X + Y\} = \sum_{i,j} (a_i + b_j)p_{ij}$$

Using (201.2), we readily obtain

$$E\{X + Y\} = E\{X\} + E\{Y\}$$

It is also easy to see that if $X_1, \ldots, X_n$ are *independent*,

$$E\{X_1, X_2, \ldots, X_n\} = E\{X_1\} \ldots E\{X_n\} \tag{202.2}$$

203. Binomial distribution. Suppose that in a single trial an event E has probability p of occurring and probability $1 - q$ of not occurring. Suppose furthermore that we perform n independent trials and ask for the probability that E will occur exactly k times ($0 \leq k \leq n$). Let us associate with the trials random variables $X_1, X_2, \ldots, X_n$ such that

$$\text{prob. } \{X_i = 1\} = p, \qquad \text{prob. } \{X_i = 0\} = q = 1 - p$$

Corresponding to the assumption of the independence of trials we assume that $X_1, X_2, \ldots, X_k$ are independent. The random variable

$$X_1 + X_2 + \ldots + X_k$$

symbolizes the number of times the event E occurs in k trials. We want to find

$$P_n(k) = \text{prob. } \{X_1 + X_2 + \ldots + X_n = k\}$$

Let z be a complex number and consider

$$E\{z^{X_1 + X_2 + \ldots + X_n}\}$$

On the one hand we have

$$E\{z^{X_1 + \ldots + X_n}\} = \sum_{k=0}^{n} P_n(k)z^k$$

and on the other hand, since $X_1, \ldots, X_n$ are independent,

$$E\{z^{X_1 + \ldots + X_n}\} = E\{z^{X_1}\} \ldots E\{z^{X_n}\} = (pz + q)^n$$

Consequently,

$$(pz + q)^n = \sum_{k=0}^{n} P_n(k)z^k$$

or

$$P_n(k) = \binom{n}{k} p^k q^{n-k} \tag{203.1}$$

we need the following inequality derived originally by Tchebysheff. If Y is a random variable,

$$\text{prob. } \{ | Y | \geq \alpha \} \leq \frac{E\{Y^2\}}{\alpha^2} \tag{203.3}$$

Suppose that Y can assume the values $a_1, a_2, \ldots$ with respective probabilities $p_1, p_2, \ldots$. Then

$$\text{prob. } \{ | Y | \geq \alpha \} = \sum_{|a_i| \geq \alpha} p_i$$

We now write

$$E\{Y^2\} = \sum_i a_i^2 p_i \geq \sum_{a_i^2 \geq \alpha^2} a_i^2 p_i \geq \alpha^2 \sum_{|a_i| \geq \alpha} p_i = \alpha^2 \text{ prob. } \{ | Y | \geq \alpha \}$$

and (203.3) follows. Applying (203.3) to the random variable

$$Y = K - \langle K \rangle_{av}$$

we obtain after noticing that

$$E\{[K - E\{K\}]^2\} = E\{K^2\} - [E\{K\}]^2 = \frac{nv}{V}\left(1 - \frac{v}{V}\right)$$

$$\text{prob. } \left\{ \left| K - n\frac{v}{V} \right| \geq \alpha \right\} \leq \frac{n(v/V)(1 - v/V)}{2}$$

Putting $\alpha = \beta n$ we obtain

$$\text{prob. } \left\{ \left| \frac{K}{n} - \frac{v}{V} \right| \geq \beta \right\} \leq \frac{(v/V)(1 - v/V)}{n\beta^2}$$

We thus see that the probability that the ratio of the actual number K of points in v to n will deviate from the average ratio (v/V) by more than β approaches 0 as $n \to \infty$. This is true regardless of the size of β.

An ideal gas can be thought of as a collection of n points distributed randomly (in the above sense) in a volume V. The density of the gas in a volume $v < V$ is a random variable, namely, $d = Km/v$, where m denotes the mass of one gas particle. The average density is

$$\langle d \rangle_{av} = \frac{m\langle K \rangle_{av}}{v} = \frac{nm}{V}$$

By (203.3),

$$\text{prob. } \{ | d - \langle d_{av} \rangle | \geq \delta \langle d \rangle_{av} \}$$

$$= \text{prob. } \left\{ \left| \frac{K}{n} - \frac{v}{V} \right| \geq \delta \frac{v}{V} \right\} \leq \frac{(1 - v/V)}{\delta^2 nv/V}$$

Due to the enormous size of n (which is of order 10^{20} per cm^3) the probability that the density in a volume v will deviate from the average density by more than say $10^{-3}\%$ ($\delta = 10^{-5}$) of the average density will be very small as long as v/V is not too small. This is the reason why it still makes sense to speak of the density in v even though the number of particles in v is not fixed but is subject to fluctuations. Of course, if

v/V is very small, the fluctuations become sufficiently pronounced and the ordinary notion of density loses its meaning. The "density fluctuations" in small volumes can actually be detected experimentally inasmuch as they cause scattering of sufficiently short wavelengths.

204. The normal (Gaussian) distribution as a limiting case. The inequality (203.3) is too rough, especially if the right side is not small. We shall now derive a more accurate result. We shall prove that

$$\lim_{n\to\infty} \text{prob. } \{K - np < \alpha\sqrt{pqn}\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-u^2/2} du \qquad (204.1)$$

The proof is based on a principle which is due to Markoff. Let $\psi(x)$ be defined as follows.

$$\psi(x) = \begin{cases} 0 \text{ if } x < \alpha \\ 1 \text{ if } x > \alpha \\ \frac{1}{2} \text{ if } x = \alpha \end{cases}$$

and consider the random variable

$$\psi\left(\frac{K - np}{\sqrt{pqn}}\right) \qquad (204.2)$$

Since the random variable (204.2) assumes only values 1, 0, and $\frac{1}{2}$, it is clear that

$$E\left\{\psi\left(\frac{K - np}{\sqrt{pqn}}\right)\right\}$$

$$= \text{prob. } \{K - np < \alpha\sqrt{pqn}\} + \frac{1}{2} \text{ prob. } \{K - np = \alpha\sqrt{pqn}\} \qquad (204.3)$$

It is easy to show that

$$\psi(x) = \frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin(\alpha - x)\xi}{\xi} d\xi$$

and consequently

$$E\left\{\psi\left(\frac{K - np}{\sqrt{pqn}}\right)\right\} = \frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{E\left\{\sin\left(\alpha - \frac{K - np}{\sqrt{pqn}}\right)\xi\right\}}{\xi} d\xi \qquad (204.4)$$

Now,

$$\begin{aligned} E\left\{\sin\left(\alpha - \frac{K - np}{\sqrt{pqn}}\right)\xi\right\} &= \mathcal{I}m E\{e^{i\xi\alpha} e^{-i\xi(K-np)/\sqrt{pqn}}\} \\ &= \mathcal{I}m e^{i\xi\alpha}\{E[e^{-i\xi(X-p)/\sqrt{pqn}}]\}^n \\ &= \mathcal{I}m e^{i\xi\alpha}[pe^{-i\xi(1-p)/\sqrt{pqn}} + qe^{i\xi p/\sqrt{pqn}}]^n \end{aligned}$$

As $n \to \infty$, we find

$$\lim_{n\to\infty} E\left\{\sin\left(\alpha - \frac{K - np}{\sqrt{pqn}}\right)\xi\right\} = \mathcal{I}m e^{i\xi\alpha} e^{-\xi^2/2} = \sin \alpha\xi e^{-\xi^2/2}$$

and thus by (204.4),

$$\lim_{n\to\infty} E\left\{\psi\left(\frac{K-np}{\sqrt{pqn}}\right)\right\} = \frac{1}{2\pi} + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\sin \alpha\xi}{\xi} e^{-\xi^2/2} d\xi$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-\xi^2/2} d\xi$$

Using (204.4) and the intuitively clear fact that

$$\lim_{n\to\infty} \text{prob. } \{K - np = \alpha\sqrt{pqn}\} = 0$$

we get

$$\lim_{n\to\infty} \text{prob. } \{K - np < \alpha\sqrt{pqn}\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-\xi^2/2} d\xi \qquad (204.5)$$

In the above derivation we failed to justify rigorously two rather subtle steps. The first step consisted in interchanging the order of operation E and integration (over an infinite range), and the second in taking the limit under the integral sign. In our particular case the first step is easily justifiable because the operation E is nothing but a finite summation. The second step is far more delicate, since some of the integrals involved are not absolutely convergent. Markoff himself was aware of the incompleteness of his argument and he used it only heuristically. The first rigorous justification was given in 1925 by Paul Lévy.

Lévy actually proved a more general theorem, which we state in a form most suitable for our purposes: If $\{X_n\}$ is a sequence of random variables such that for every real ξ

$$\lim_{n\to\infty} E\{e^{i\xi X_n}\} = \int_{-\infty}^{\infty} e^{i\xi x} \zeta(x) dx$$

where $\zeta(x) \geq 0$ and $\int_{-\infty}^{\infty} \zeta(x) dx = 1$, then for every α,

$$\lim_{n\to\infty} \text{prob. } \{X_n < \alpha\} = \int_{-\infty}^{\alpha} \zeta(x) dx$$

A heuristic derivation of this important result follows the procedure outlined above.

We note here a closely related result. If, for every real ξ,

$$\lim_{n\to\infty} E\{e^{i\xi X_n}\} = f(\xi) \qquad (204.6)$$

and $f(\xi)$ is continuous at $\xi = 0$, then

$$\lim_{n\to\infty} \text{prob. } \{X_n < \alpha\} = \int_{-\infty}^{\alpha} \zeta(x) dx \qquad (204.7)$$

where

$$\zeta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\xi x} f(\xi) d\xi \qquad (204.8)$$

It follows, in particular, that $\zeta(x) \geq 0$ and $\int_{-\infty}^{\infty} \zeta(x) dx = 1$. The following corollary of this general theorem is often used. If $\{X_n\}$ is a sequence of random variables such that

$$\lim_{n\to\infty} E\{X_n^k\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-x^2/2} dx$$

for $k = 1, 2, \ldots$, then

$$\lim_{n\to\infty} \text{prob. } \{X_n < \alpha\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx$$

The quantity $E\{X_n^k\}$ is called the kth moment of X_n.

205. The continuous case. So far, we have been dealing with random variables which were capable of assuming only a finite number of values. However, in many important applications one is led to random variables whose values may fill an interval. The modifications in treatment which thus become necessary are reasonably obvious. In defining a random variable X we must be able to assign probabilities to events like $\{\alpha < X < \beta\}$. This is often done by postulating a "probability density" $\zeta(x)$ and requiring that

$$\text{prob. } \{\alpha < X < \beta\} = \int_{\alpha}^{\beta} \zeta(x) dx \tag{205.1}$$

A probability density must satisfy the conditions

$$\zeta(x) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} \zeta(x) dx = 1$$

Independence of random variables $X_1, X_2, \ldots, X_n$ with respective density functions $\zeta_1(x), \zeta_2(x), \ldots, \zeta_n(x)$ is defined by the equation

$$\text{prob. } \{\alpha_1 < X_1 < \beta_1, \alpha_2 < X_2 < \beta_2, \ldots, \alpha_n < X_n < \beta_n) = \int_{\alpha_2}^{\beta_2} \zeta(x_1) dx, \int_{\alpha_2}^{\beta_2} \zeta(x_2) dx_2, \ldots, \int_{\alpha_n}^{\beta_n} \zeta(x_n) dx_n \tag{205.2}$$

From this definition it follows that if Ω is a region in the n-dimensional Euclidean space, the probability that the "random point" $(X_1, X_2, \ldots, X_n)$ will fall within Ω is given by the integral

$$\int \cdots \int_{\Omega} \zeta(x_1)\zeta(x_2) \ldots \zeta(x_n) dx_1 dx_2 \ldots dx_n \tag{205.3}$$

In fact, definition (205.2) is completely equivalent with this statement if Ω is the "parallelepiped" $(a_1 \leq x_1 \leq l_1)$, $(a_2 \leq x_2 \leq l_2)$, $\ldots$, $(a_n \leq x_n \leq l_n)$. If Ω is not a "parallelepiped" it can (in most cases of physical interest) be decomposed into a finite or denumerably infinite number of nonoverlapping "parallelepipeds." Since nonoverlapping "parallelepipeds" correspond to mutually exclusive events, we see that in order to justify (205.3) one needs an extension of the axiom of Sec. 199 to a denumerable infinitude of mutually exclusive events. Such an extension is indeed postulated in the theory of probability. If there are no grounds for assuming that $X_1, X_2, \ldots, X_n$ are independent,

formula (205.3) must be modified. We must, in general, assume the existence of a function $\zeta(x_1,x_2,\ldots,x_n)$ satisfying the conditions

$$\zeta \geq 0,$$

$$\int_{-\infty}^{\infty} \cdots \int \zeta(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = 1$$

and postulate that the probability that the "random point" $(X_1,X_2,\ldots,X_n)$ should fall within Ω in

$$\int \cdots \int_{\Omega} \zeta(x_1,x_2,\ldots,x_n)dx_1 \ldots dx_n \tag{205.4}$$

The function $(x_1,x_2,\ldots,x_n)$ is called the joint probability density of $X_1, X_2, \ldots, X_n$.

206. An example. To illustrate the notions discussed in Sec. 205, let us consider the following problem. A point P is chosen at random in a volume V (bounded by the surface S), and then it is given a velocity $\boldsymbol{v}$ of random magnitude and direction. What is the probability that P will be outside V after time t? We assume that the position and velocity are independent. If $\boldsymbol{r}$ and $\boldsymbol{v}$ denote, respectively, the position and velocity of P, the desired probability can be written as the sixtuple integral

$$\int \cdots \int_{\Omega} \zeta(\boldsymbol{r},\boldsymbol{v})d\tau_1\, d\tau_2 \tag{206.1}$$

where $d\tau_1$ and $d\tau_2$ denote the respective volume elements in the configuration $(\boldsymbol{r})$ and velocity $(\boldsymbol{v})$ spaces. If Cartesian coordinates are used throughout, we have $d\tau_1 = dx\, dy\, dz$ and $d\tau_2 = dv_x\, dv_y\, dv_z$. The region Ω is defined by the requirement that $\boldsymbol{r} + \boldsymbol{v}t$ be outside V. The assumption that $\boldsymbol{r}$ and $\boldsymbol{v}$ are independent means that

$$\zeta(\boldsymbol{r},\boldsymbol{v}) = \zeta_1(\boldsymbol{r})\zeta_2(\boldsymbol{v}) \tag{206.2}$$

Assuming furthermore $\zeta_1(\boldsymbol{r})$ to be constant, all possible positions are equiprobable.

$$\zeta_1(\boldsymbol{r}) = \frac{1}{V} \tag{206.3}$$

It is also natural to assume that

$$\zeta_2(\boldsymbol{v}) = f(\omega)d\omega\,\frac{d\alpha}{4\pi} \tag{206.4}$$

where ω is the magnitude of $\boldsymbol{v}$, and $d\alpha$ is the element of the solid angle. The assumption (206.4) means that the magnitude and the direction of the velocity are independent, the probability density of the magnitude is $f(\omega)$, and that all directions in space are equiprobable.

With the simplifying assumptions (206.2), (206.3), and (206.4), we can rewrite (206.1).

$$\frac{1}{V}\iiint_V d\tau_1 \int_0^\infty f(\omega)\,\frac{\alpha_P(\omega t)d\omega}{4} \tag{206.5}$$

where $\alpha_P(\omega t)$ is the solid angle at P subtended by the portion of the sphere with center at P and radius ωt, which lies outside V. In general, an explicit evaluation of (206.5) is impossible. However, a simple formula can be obtained which is valid for small t. Interchanging the order of integrations in (206.5), we are led to the integral

$$\iiint_V \alpha_P(\omega t)d\tau_1 \tag{206.6}$$

At each point of the surface S (bounding the volume V) we draw an inward normal and locate on it the point whose distance from the surface is ωt. These new points form a surface S_1.* If P is not in the shell between S_1 and S it is clear that $\alpha_p(\omega t) = 0$. Consequently, integration in (206.6) is extended only over the shell between S_1 and S. Let S_x, for $0 \leq x \leq \omega t$, be the surface constructed in the same way as S_1 except that ωt is replaced by x. ($S_0 = S$, $S_{\omega t} = S_1$). If P is a point on S_x and if S is sufficiently smooth we get easily

$$\alpha_P(\omega t) \backsim \frac{1}{2}\left(1 - \frac{x}{\omega t}\right)$$

and consequently

$$\iiint_V \alpha_p(\omega t)d\tau_1 \backsim \frac{1}{2}\int_0^{\omega t} S_x\left(1 - \frac{x}{\omega t}\right)dx$$

$$\frac{S}{2}\int_0^{\omega t}\left(1 - \frac{x}{\omega t}\right)dx = \frac{S}{4}\omega t$$

Finally, for small t, the desired probability is

$$\frac{St}{4V}\int_0^\infty \omega f(\omega)d\omega = \frac{S}{4V}\bar{\omega}t$$

where

$$\bar{\omega} = \int_0^\infty \omega f(\omega)d\omega$$

is the mean magnitude of velocity. The result just derived is often used in kinetic theory of gases.

207. Expectations in the continuous case. The extension of the definition of the mathematical expectation to continuously distributed random variables consists in replacing sums by integrals. If X is a random variable with probability density $\rho(x)$, the mathematical expectation $E(X)$ is defined by the formula

* This surface may be extremely complicated even though S may be quite simple. However, for *small* ωt, surface S_1 will, for sufficiently regular S, be also regular.

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)dx \tag{207.1}$$

provided the integral exists.

If $X_1, X_2, \ldots, X_n$ are random variables whose joint probability density is $\rho(x_1,x_2,\ldots,x_n)$ and if $Y = f(X_1,X_2,\ldots,X_n)$, where f is a sufficiently regular function of n variables,

$$E\{Y\} = \int_{-\infty}^{\infty} \cdots \int f(x_1,x_2,\ldots,x_n)\rho(x_1,x_2,\ldots,x_n)dx_1\,dx_2 \ldots dx_n \tag{207.2}$$

assuming again that the integral exists.

208. Characteristic functions. The characteristic function of the joint probability density of $X_1, X_2, \ldots, X_n$ is defined by the formula

$$F(\xi_1,\xi_2,\ldots,\xi_n) = E\{e^{i\sum_1^n \xi_k X_k}\}$$
$$= \int_{-\infty}^{\infty} \cdots \int e^{i\sum_1^n \xi_k x_k}\rho(x_1,x_2,\ldots,x_n)dx_1\,dx_2 \ldots dx_n \tag{208.1}$$

where $\xi_1, \xi_2, \ldots, \xi_n$ are real.

If $X_1, X_2, \ldots, X_n$ are independent, we clearly have

$$E\{e^{i\sum_1^n \xi_k X_k}\} = \prod_1^n E\{e^{i\xi_k X_k}\}$$

and the characteristic function $F(\xi_1,\xi_2,\ldots,\xi_n)$ becomes a product of n functions of the individual variables $\xi_1, \xi_2, \ldots, \xi_n$.

From (208.1) we obtain by the Fourier integral formula

$$\rho(x_1,x_2,\ldots,x_n)$$
$$= \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \cdots \int e^{-i\sum_1^n \xi_k x_k} F(\xi_1,\xi_2,\ldots,\xi_n)d\xi_1\,d\xi_2 \ldots d\xi_n \tag{208.2}$$

To justify (208.2) one needs some regularity conditions on ρ. In most cases of practical interest these conditions are easily satisfied. The case $n = 1$ is of particular importance and we record it for future reference

$$\rho(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\xi x} F(\xi)d\xi \tag{208.3}$$

The importance of formulas (208.2) and (208.3) is due to the fact that often it is easy to calculate the characteristic function directly. Once this is done the corresponding density functions can be immediately written as Fourier integrals.

209. Random flights. To illustrate the use of characteristic functions let us first consider the following problem, first proposed and solved by Lord Rayleigh. A point starting from the origin of a Cartesian coordinate system moves in steps of constant length l, but the direction is chosen and time at random (all directions in space being equiprobable).

Assuming that individual steps are independent, we want to find the probability $\rho_n(x, y, z)dx\,dy\,dz$, that after n steps the point is in the volume element $dx\,dy\,dz$.

If we symbolize the steps by random vectors $r_1, r_2, \ldots, r_n$ and denote by X_k, Y_k, Z_k the x, y, and z components of r_k, we see that our problem is to find the joint density function of the random variables

$$\sum_1^n X_k, \quad \sum_1^n Y_k, \quad \sum_1^n Z_k$$

we have

$$F_n(\xi_1,\xi_2,\xi_3) = E\{e^{i(\xi_1\sum_1^n X_k+\xi_2\sum_1^n Y_k+\xi_3\sum_1^n Z_k)}\}$$
$$= \prod_{k=1}^n E\{e^{i(\xi_1X_k+\xi_2Y_k+\xi_3Z_k)}\}$$

Since $X_k^2 + Y_k^2 + Z_k^2 = l^2$ (the length of each step is l) we see that the joint density function of X_k, Y_k, Z_k is singular, and in view of the spherical symmetry of the problem (all directions in space being equiprobable) we can represent this density by the formula

$$\rho_1(x,y,z) = \frac{\delta(l - \sqrt{(x^2 + y^2 + z^2)})}{4\pi l^2}$$

Thus

$$E\{e^{i(\xi_1X_k+\xi_2Y_k+\xi_3Z_k)}\}$$
$$= \frac{1}{4\pi l^2}\int_{-\infty}^{\infty}\int\int e^{i(\xi_1x+\xi_2y+\xi_3z)}(l - \sqrt{(x^2 + y^2 + z^2)})dx\,dy\,dz$$

If ξ is the vector whose components are ξ_1, ξ_2, ξ_3, we can write the above integral in an invariant form

$$\frac{1}{4\pi l^2}\int_{-\infty}^{\infty}\int\int e^{i\xi\cdot r}\delta(l - r)dt$$

Introducing polar coordinates with the z axis along the vector ξ we see that the above integral is equal to

$$\frac{1}{4\pi l^2}\int_0^{2\pi}\int_0^{\pi}\int_0^{\infty} e^{i\xi r\cos\varphi}\delta(l - r)r^2\sin\varphi\,dr\,d\varphi\,d\theta$$

where

$$\xi = \sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}$$

Integrating with respect to r first, we get finally

$$\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} e^{i\xi l\cos\varphi}\sin\varphi\,d\varphi\,d\theta = \frac{\sin l\sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}}{l\sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}}$$

Thus

$$F_n(\xi_1,\xi_2,\xi_n) = \left(\frac{\sin l\sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}}{l\sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}}\right)^n$$

and using (208.2),

$$\rho_n(x,y,z) = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}\int\int e^{-i(\xi_1 x+\xi_2 y+\xi_3 z)} \left(\frac{\sin l\sqrt{\xi_1^2+\xi_2^2+\xi_3^2}}{l\sqrt{\xi_1^2+\xi_2^2+\xi_3^2}}\right)^n d\xi_1\, d\xi_1\, d\xi_2 \qquad (209.1)$$

It is difficult to reduce (209.1) to a form which would be amenable to numerical calculations. However, one limiting case is of particular importance and interest. If

$$l = \frac{\alpha\sqrt{3}}{\sqrt{n}} \qquad (209.2)$$

we see that

$$\left(\frac{\sin l\sqrt{\xi_1^2+\xi_2^2+\xi_n^2}}{l\sqrt{\xi_1^2+\xi_2^2+\xi_n^2}}\right)^n \to e^{-(\alpha^2/2)(\xi_1+\xi_2^2+\xi_n^2)}$$

and consequently

$$\rho_n(x,y,z) \to \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}\int\int e^{-i(\xi_1 x+\xi_2 y+\xi_3 z)} e^{-(\alpha^2/2)(\xi_1^2+\xi_2^2+\xi_3^2)} d\xi_1\, d\xi_2\, d\xi_3 \qquad (209.3)$$

$$= \frac{1}{(\sqrt{2\pi}\alpha)^3} e^{-(x^2+y^2+z^2)/2\alpha^2}$$

A rigorous justification of this limiting process could be based on a suitable generalization of the theorem stated at the end of Sec. 204. The assumption (209.2) may seem artificial, but it arises quite naturally in physical problems.

For instance, a satisfactory model of Brownian motion of a free particle is obtained by treating it as a random walk problem of the type considered in this section. The steps are assumed to be of length Δ, and the duration of each step is τ. Both Δ and τ are small but one requires the ratio $\Delta^2/\tau = D$ to be finite.

If we are interested in statistical predictions after time t (assuming that the Brownian particle was observed at the origin at $t = 0$) we see that

$$n \backsim \frac{t}{\tau}$$

and consequently

$$l = \Delta = \sqrt{D\tau} = \frac{\sqrt{Dt}}{\sqrt{n}}$$

We are thus led to (209.2) with

$$\alpha\sqrt{3} = \sqrt{Dt}$$

Formula (202.3) assumes now the form

$$\rho_t(x,y,z) = \left(\frac{3}{2\pi Dt}\right)^{3/2} e^{-3(x^2+y^2+z^2)/2Dt}$$

which is one of the fundamental results of the theory of Brownian motion.

210. The Holtzmark distribution. As another example we consider the following problem.

Suppose that N material particles, each of mass m, are distributed at random in the interval $(-R,R)$. We want to find the probability distribution of the field at O in the limit as $N \to \infty$, $R \to \infty$ but

$$\frac{N}{2R} = \nu$$

remains fixed. Let the position of the ith particle be symbolized by the random variable X_i. We interpret the assumption that the particles are distributed at random in $(-R,R)$ as meaning that each X_i is uniformly distributed in $(-R,R)$, and the X_i's are independent.

Our problem is to find the limiting distribution of

$$F = m \sum_{j=1}^{N} \frac{X_j}{|X_j|^3}$$

We find the characteristic function

$$E\{\exp(i\xi F)\}$$

Because of independence we have

$$E\{\exp(i\xi F)\} = \left\{E\left\{\exp\left(i\xi m \frac{X}{|X|^3}\right)\right\}\right\}^N$$

and since X is uniformly distributed in $(-R,R)$ we have

$$\begin{aligned} E\left\{\exp\left(i\xi m \frac{X}{|X|^3}\right)\right\} &= \frac{1}{2R}\int_{-R}^{R} \exp\left(i\xi m \frac{X}{|X|^3}\right)dx \\ &= \frac{1}{R}\int_0^R \cos\frac{m\xi}{X^2}\,dx \\ &= 1 - \frac{1}{R}\int_0^R \left(1 - \cos\frac{m\xi}{x^2}\right)dx \end{aligned}$$

It is easily seen that

$$\int_0^\infty \left(1 - \cos\frac{m\xi}{x^2}\right)dx$$

exists and consequently, in the limit

$$N \to \infty, \qquad R \to \infty, \qquad \frac{N}{2R} = \nu$$

We obtain

$$\begin{aligned} E\{\exp(i\xi F)\} &\to \exp\left\{-2\nu\int_0^\infty \left(1 - \cos\frac{m\xi}{x^2}\right)dx\right\} \\ &= \exp(-\gamma\,|\,\xi\,|^{1/2}) \end{aligned}$$

where

$$\gamma = 2\nu\sqrt{m}\int_0^\infty \left(1 - \cos\frac{1}{y^2}\right)dy.$$

The limiting density function of F is thus

$$\rho(F) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-i\xi F)\exp(-\gamma \mid \xi \mid^{1/2})d\xi$$

This integral is not elementary. The above problem was first solved in the more realistic, three-dimensional case (points distributed at random in a sphere of radius R) by Holtzmark. The same problem is encountered in stellar statistics, and in this connection it was discussed by von Neumann and Chandrasekhar.

211. The Poisson distribution. The number of electrons emitted by a heated cathode in a given time interval is a random variable whose distribution is of considerable interest.

The theory of this, and related phenomena, is based on the following assumptions.

1. There exists a constant λ such that the probability that an electron will be emitted between t and $t + dt$ is $\lambda\, dt$ up to terms of higher order. The probability that more than one electron will be emitted in the time interval $(t,t+dt)$ is of order higher than dt.

2. The number of electrons emitted in nonoverlapping time intervals are independent. From these assumptions it is easy to find the probability $p_k(t)$ that *exactly* k electrons will be emitted between 0 and t. We have

$$p_k(t+dt) = p_k(t)(1 - \lambda\, dt) + p_{k-1}(t)\lambda\, dt, \qquad k > 1$$

and

$$p_0(t+dt) = p_0(t)(1 - \lambda\, dt)$$

From these equations we obtain

$$\frac{dp_k(t)}{dt} = -\lambda p_k(t) + \lambda p_{k-1}(t), \qquad k > 1$$

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t)$$

Thus

$$p_0(t) = e^{-\lambda t}$$

since $p_0(0) = 1$, and

$$p_k(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

since $p_k(0) = 0$ for $k > 1$. If a random variable X can assume only the values 0, 1, 2, ..., and if

$$\text{prob. } \{X = k\} = e^{-h}\frac{h^k}{k!} \quad (h > 0,\ k = 0, 1, 2, \ldots)$$

we say that it has the Poisson distribution with mean h.

The fact that h is the mean of X follows from the equation

$$E\{X\} = e^{-h} \sum_{k=0}^{\infty} \frac{kh^k}{k!} = h$$

An interesting property of the Poisson's distribution is

$$E\{(X - h)^2\} = h$$

Coming back to our original problem of electron emission, we can say that the number of electrons emitted during a time interval of duration t is distributed according to Poisson's distribution with mean λt. Thus λ can be interpreted as the mean number of electrons emitted per unit time. The above derivation of Poisson's distribution can be applied to various situations. For instance, if $p_0(x)$ denotes the probability that a molecule of an ideal gas will travel distance x without suffering a collision,

$$p_0(x + dx) = p_0(x)(1 - \lambda\, dx)$$

and consequently

$$p_0(x) = e^{-\lambda x}$$

Thus the probability $\omega(x)dx$ that the length of the free path should be between x and $x + dx$ is

$$\omega(x)dx = p_0(x) - p_0(x + dx) = \lambda e^{-\lambda x} dx$$

This is a classical formula due to Clausius. Defining the mean free path as

$$\bar{l} = \int_0^{\infty} x\omega(x)dx$$

we see that $\bar{l} = 1/\lambda$. From our derivation it is clear that λ is the number of collisions suffered by a particle per unit length.

The Poisson distribution is also encountered in problems of a different nature. We have seen that if n points are distributed at random in a volume V, the probability that exactly k points will fall within a volume v is

$$\binom{n}{k}\left(\frac{v}{V}\right)^k\left(1 - \frac{v}{V}\right)^{n-k}$$

In the limit as $n \to \infty$, $V \to \infty$ but

$$\frac{n}{V} = \nu$$

remains constant, we see that the above expression becomes

$$e^{-\nu}\frac{\nu^k}{k!}$$

Since molecules of an ideal gas can be considered as randomly distrib-

uted in a container, if v is small compared to V we can say that the number of molecules in v is distributed according to Poisson's law.

If ν is large and $k - \nu$ small compared to ν, that is $(k - \nu)/\nu \ll 1$, one can approximate

$$e^{-\nu}\frac{\nu^k}{k!}$$

conveniently as follows. We write [using Stirling's formula, see formula (113.7)]

$$e^{-\nu}\frac{\nu^k}{k!} \approx \frac{e^{-\nu}\nu^k}{\sqrt{2\pi k}k^k e^{-k}} = \frac{e^{-(\nu-k)}}{\sqrt{2\pi k}}\left(1 - \frac{k-\nu}{k}\right)^k$$

and note that

$$\left(1 - \frac{k-\nu}{k}\right)^k = \exp\left\{k \log\left(1 - \frac{k-\nu}{k}\right)\right\} \approx$$

$$\exp\left\{-(k-\nu) - \frac{1}{2}\frac{(k-\nu)^2}{k}\right\}$$

Finally,

$$e^{-\nu}\frac{\nu^k}{k!} \approx \frac{1}{\sqrt{2\pi k}} \exp\left\{-\frac{1}{2}\frac{(k-\nu)^2}{k}\right\}$$

212. Generating functions. One of the oldest and most powerful methods of probability theory is the method of generating functions. We illustrate this method by the following example. Suppose that in an elementary time interval Δt a particle can either disappear without producing any effect (probability p_0) or disappear producing 1, 2, 3, ... new particles (respective probabilities being p_1, p_2, p_3, ...). Each of the new particles, independently of each other, behaves like the parent particle, etc.

The problem is to calculate the probability that after r steps (time $r\Delta t$) there will be exactly k particles ($k = 0, 1, 2, \ldots$). This is a simplified version of a problem encountered in the theory of chain reactions.

Let $X_1, X_2, \ldots$ be independent random variables and

$$\text{prob. } \{X_j = l\} = p_l, \quad j = 1, 2, \ldots; \quad l = 0, 1,$$

and let N_r denote the random variable which symbolizes the number of particles at time $r\Delta t$. We have $N_0 = 1$, and

$$N_r = \sum_{j=1}^{N_{r-1}} X_j$$

It is clear that the X's are independent from N_{r-1} (this means that what each individual particle does is independent of the number of particles in the given generation). Let

$$P_k^{(r)} = \text{prob. } \{N_r = k\}$$

Then
$$f_r(z) = \sum_{k=0}^{\infty} P_k^{(r)} z^k$$
is called the generating function of the sequence $\{P_k^{(r)}\}$. Clearly,
$$f_r(z) = E\{z^{N_r}\} \tag{212.1}$$
Now,
$$E\{z^{N_r}\} = E\{z^{\sum_1^{N_{r-1}} X_i}\}$$
$$= \sum_{l=0}^{\infty} \text{prob. } \{N_{r-1} = l\} E\{z_1^{\sum^l X_i}\}$$
$$= \sum_{l=0}^{\infty} P_l^{(r-1)} g^l(z)$$
where
$$g(z) = E\{z^{X_i}\} = \sum_{l=0}^{\infty} p_l z^l$$
Consequently,
$$E\{z^{N_r}\} = f_{r-1}[g(z)]$$
or
$$f_r(z) = f_{r-1}[g(z)] \tag{212.2}$$
Finally,
$$f_1(z) = g(z)$$
$$f_2(z) = g[g(z)]$$
$$f_3(z) = g\{g[g(z)]\}$$
etc. The process of interaction leads, in general, to extremely complicated expressions for $f_r(z)$, and it is almost impossible to get a general formula for P_k^r.

It is easy, however, to obtain the average number of particles $E\{N_r\}$. In fact, from (212.1),
$$E\{N_r\} = \left(\frac{df_r(z)}{dz}\right)_{z=1} = f_r'(1)$$
and from (212.2),
$$f_r'(1) = f_{r-1}'[g(1)]g'(1)$$
But $g(1) = 1$ (since $p_0 + p_1 + \ldots = 1$) and consequently
$$f_r'(1) = f_{r-1}'(1)g'(1)$$
or
$$E\{N_r\} = f_r'(1) = [g'(1)]^r$$
It is easy to see that $g'(1) = \kappa$ is the average number of particles produced by a particle in time Δt. If $\kappa > 1$ we have an exponential increase; if $\kappa < 1$, an exponential decay; and if $\kappa = 1$, a state of "uneasy" equilibrium. By similar means one can calculate $E\{N_r^2\}$ and thus the fluctuation $E\{N_r^2\} - [E\{N_r\}]^2$.

213. An example. As another illustration of the method of generating functions we consider the following problem. A particle, starting

from the origin, can move along the x axis in steps $+1$ or -1, the sign being chosen at random with equal probabilities. We want to calculate the probability that the particle will return to the origin, *for the first time*, after s steps. Let $P'(s)$ denote the desired probability and let $P(s)$ denote the probability that the particle will be at the origin after s steps (not necessarily for the first time). A moment's reflection will convince us that the following relation holds.

$$P(s) = P'(s) + \sum_{k=1}^{s-1} P'(k)P(s-k) \tag{213.1}$$

Introducing the generating functions

$$f(z) = \sum_{1}^{\infty} P(s)z^s$$

$$g(z) = \sum_{1}^{\infty} P'(s)z^s$$

we see that (213.1) is equivalent to

$$f(z) = g(z) + f(z)g(z) \tag{213.2}$$

Thus

$$g(z) = \frac{f(z)}{1+f(z)} \tag{213.3}$$

On the other hand, $P(s)$ is the probability that in s games of "heads or tails," the gain of a player should be 0, and consequently

$$P(s) = \begin{cases} \frac{1}{2^s}\binom{s}{\frac{s}{2}} & \text{if } s \text{ is even} \\ 0 & \text{if } s \text{ is odd} \end{cases}$$

We can write

$$P(s) = \frac{1}{2\pi}\int_0^{2\pi} \cos^s \xi \, d\xi$$

and hence

$$f(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{d\xi}{1 - z\cos\xi} - 1 = \frac{1}{\sqrt{1-z^2}} - 1$$

Finally,

$$g(z) = 1 - \sqrt{1-z^2} = \sum_{r=1}^{\infty} \binom{\frac{1}{2}}{r} (-1)^{r+1} z^{2r}$$

and

$$P'(s) = \begin{cases} 0 & \text{if } s \text{ is odd} \\ \binom{\frac{1}{2}}{r}(-1)^{s/2+1} & \text{if } s \text{ is even} \end{cases}$$

SUGGESTED REFERENCES

J. V. Uspensky, *Introduction to Mathematical Probability*, New York, McGraw-Hill Book Company, Inc., 1937.

Harold Cramer, *Random Variables and Probability Distributions*, London, Cambridge University Press, 1937.

Richard V. Mises, *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Leipzig und Wien, Franz Deuticke, 1931.

William Feller, *Introduction to Probability Theory and Its Applications*, Wiley & Sons, 1950.

PROBLEMS

204.1. Prove that

$$\lim_{h\to\infty} e^{-h} \sum_{k<h+\alpha\sqrt{h}} \frac{h^k}{k!} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-x^2/2}\, dx.$$

Note that

$$\sum_{k<h+\alpha\sqrt{h}} \frac{h^k}{k!} = \sum_{k=0}^{\infty} \frac{h^k}{k!} \psi\left(\frac{k-h}{\sqrt{h}}\right)$$

and follow the derivation of Sec. 197.

204.2. A particle starting from the origin on the x axis moves in independent steps of length 1, the direction at each step being determined at random with equal probabilities. Find an asymptotic expression for the expected number of times the particle returns to the origin during the first n steps.

Let $X_1, X_2, \ldots, X_n$ be the random variables representing successive steps. The successive displacements are then $s_1 = X_1$, $s_2 = X_1 + X_2, \ldots, s_n = X_1 + \ldots + X_n$. Let, for integral k,

$$\varphi(k) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k > 0. \end{cases}$$

The expected number of returns is then

$$E\{\varphi(s_1) + \varphi(s_2) + \ldots + \varphi(s_n)\}.$$

Use the results of Sec. 203 and Stirling's formula to arrive at the desired result.

205.1. The noise output of a quadratic detector in a radio receiver can be represented as $X^2 + Y^2$, where X and Y are independent random variables each having the density function

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

Find the probability density of the noise output, i.e.,

$$\frac{d}{dx} \text{prob. } \{X^2 + Y^2 < x\}.$$

If, in addition to noise, there is a signal present, the output is represented by $(X + \alpha)^2 + (Y + \beta)^2$. Solve the corresponding problem in this case.

208.1. Let $X_1, X_2, \ldots, X_n$ be independent random variables each having the density function

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

Find the density function of

$$\frac{X_1 + \ldots + X_n}{n}.$$

[First calculate the characteristic function

$$E\{e^{it(X_1+\ldots+X_n)/n}\}$$

and then apply formula (208.3).]

208.2. Solve Problem 208.1 if the density

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}$$

is replaced by

$$\frac{1}{\pi\sigma}\frac{1}{1+(x/\sigma)^2}.$$

209.1. Carry out the calculations of Sec. 209 for random flights in the plane.

210.1. Let $X_1, X_2, \ldots$ be independent random variables each having the density function

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}$$

Calculate the limiting distribution of

$$\frac{1}{n}\sum_{k=1}^{n}\frac{1}{X_k}$$

211.1. An electron going through matter has probability $\lambda\, dl$ (up to terms of higher order in dl) of producing a pair while traveling distance dl. Show that the probability that n electrons will be observed at distance l is

$$e^{-\lambda l}(1 - e^{-\lambda l})^{n-1}.$$

(Furry's model)

213.1. A player whose capital is A dollars plays "heads or tails," betting one dollar each time. Assuming that the coin is "honest" (i.e., both outcomes are equiprobable) calculate the probability that after n trials the player has still some money left. (This is the classical "ruin problem." Interpret the problem in terms of random walk and apply the method of Sec. 213.)

INDEX

Abel's integral equation, solution of, 376, 377
Approximation in the mean, 311
Associated Laguerre polynomials:
  differential equation for, 335
  generating function for, 334
  integral representation for, 336
  integrals involving, 335
  recurrence formulas for, 334, 335
Associated Legendre polynomials, 342
  differential equation for, 343
  generating function for, 342
  integral representation of, 343
  integrals involving, 343, 345
  recurrence formulas for, 343
Asymptotic expansions:
  as solution of a differential equation, 210
  calculations with, 194
  differentiation of, 194
  for the gamma function, 198
  integration of, 194
  method of steepest descents, 201
  methods of obtaining, 195
  properties of, 192
  saddle-point method, 201

Bernoulli equation, 244
Bernoulli numbers, 183
Bessel functions:
  addition formula for, 357
  differential equation for, 350
  generating function for, 348
  integral representations of, 350, 351
  integrals involving, 355
  of half odd-integer order, 351
  of the second kind, 352
  recurrence formulas for, 349, 350
Bessel's inequality, 312
Beta function, defined, 109
Bilinear forms, 272
  matrix of, 272
Binomial distribution, 430
Bound of a function, 3
Boundary-value problems, 217
Brachystochrone, The, 407
Brownian motion, model of, 440

Calculus of variations:
  Brachystochrone problem in the, 407
  direct methods of solving problems in the, 415
  Euler's equation for simple problem in the, 406
Calculus of variations (*Cont.*)
  Euler's partial differential equation in the, 411
  potential problems as problems in the, 412
  problem of vibrating string, 411
  problems involving several functions in the, 409
  problems involving several variables in the, 409
  problems with auxiliary conditions in the, 413
  systems of Euler equations in the, 410
  the first variation in the, 404, 405
  the minimal function in the, 405
  the simplest problem in the, 404
  typical problems of the, 403
Cauchy's integral formula, 151
Cauchy's integral theorem, 149
Characteristic functions, 438
Circulation of a fluid, 238
Complex numbers:
  calculation with, 129
  defined, 128
  infinite series of, 132
  representations of, 129, 130, 131
  sphere, 131
Conformal mapping, 216
  applied to fluid flow, 239
  as solution of potential problems, 218
  examples of, 218, 219
  for flow around a cylinder, 241
  for fluid with free surfaces, 245
  for potential distributed over a circle, 222
  for regions defined by certain curves, 220
  for stream lines, 239
  function for cyclotron dees, 232
  function for polygonal boundaries, 225
  function for rectangular poles, 228
  hodographic method, 245
  of polygon on a circle, 236
Conjugate:
  of complex number, 129
  potential function, 215
Continuous functions:
  definition of, 8
  theorems concerning, 9
Convergence in the mean, 312
Coordinates:
  curvilinear, 87
  elliptic, 89
  of a point, 2
  parabolic, 88

Correspondence, one-to-one, 1
Cramer's rule for solution of linear equations, 259
Curves, rectifiable, 93

Derivative, defined by a Cauchy integral, 152
Derivatives:
  definition of partial, 22
  directional, 23
  formula for change variable in, 16
  formulas for change of variable in partial, 23
  formulas for higher-order, 16
  geometrical significance of, 15
  of complex functions, 145
  of functions in parametric form, 17
  of functions of one variable, 14
  of inverse functions, 16, 17
Determinants:
  certain properties of, 253
  definition of, 252
  evaluation of by minors, 256
  Gramian, 315
  Laplace's development of, 255
  minors of, 255
  multiplication of, 255
  principal and complementary minors of, 256
  traces of, 278
  value of, 253
Differential equations, Cauchy-Riemann, 147
Differentials:
  definition of, 15
  errors estimated by total, 28
  line integral of total, 98
  total, 26
Differentiation:
  of functions in parametric form, 29
  of implicit functions, 30
Dirichlet formula, 377

Eigen-differentials, 369
Element of area:
  on a surface, 110
  on a transformed plane, 106
Element of length:
  in transformed space, 90
  on a surface, 110
Element of volume, transformation of, 118
Elliptic integrals, 54
  linear transformation of, 56
  recurrence formula for, 56
  reduction to 1st, 2nd, or 3rd kinds, 55
  reduction to normal forms, 59, 60, 61, 62
Equipotential lines, 217
Euler's equation:
  for a simple variational problem, 406
  some solutions of, 406, 407
Expectations, 429
  for continuous random variables, 437
Extrema, criteria for, 37

Fermat's principle, 403
Field intensity, 215
Flow:
  of fluid with free surfaces, 245
  of heat and Hankel functions, 356
  of ideal fluid, 237
  of ideal fluid around a cylinder, 241
Fluid flow:
  through a jet, 246
  use of hodographic method in, 245
Force of fluid on immersed obstacle, 242
Fourier coefficients, 311
Fourier integrals, examples of, 366, 367, 368
Fourier series, 317
  data represented as, 323
  number of terms required in a, 320
  over an arbitrary interval, 319
  spectral extent for a function expanded in, 322
Fourier transforms, 362
  examples of, 366, 367, 368
  mean spectral extent of, 364
  relation between functions and their, 364
Functions, analytic, 146, 149, 150
  approximation by means of orthonormal, 311
  beta, 109
  branch points of, 158
  classes of, 12
  class of integrable, 47
  complete set of, 312
  contour integral representations of, 178
  criterion for linear independence of, 315
  criterion for unique inverses of, 35
  criterion of dependence and independence, 32
  derivative of complex, 145
  explicit, 30, 31, 32
  Fourier coefficients of, 311
  graph of, 3
  implicit, 30
  infinite product representation of, 188
  limit of, 5
  lower and upper bound of, 3
  many-valued, 158, 159
  maxima and minima of, 3, 36
  nature of singularities of, 156
  odd or even, 4
  of a complex variable, continuity of, 134
  of a single variable defined, 3
  of bounded variation, 93
  of several variables, limit and continuity of, 9
  orthogonalization of, 313
  orthonormal systems of, 309
  periodic, 4
  representation as Fourier integral, 362
  residues of, 164
  singular points of, 146
  variation of, 3
  with an infinite number of poles, 186
Furry's model, 448

Gamma function:
asymptotic expansion of, 198
contour integral representation of, 179
of a real variable, 72
relations satisfied by, 191, 192
representation by infinite product, 190
Gauss method of solution of linear equations, 260
Gaussian distribution, 433
Generating functions, 444
Gradient of a function, 25
Gramian determinant, 315
Green's theorem:
for double integrals, 102
for triple integrals, 117

Hankel functions, 353
and flow of heat, 356
relation to Bessel functions, 354
Harmonic analysis, 323
Hermitian form, 273
Hermitian polynomials, 317, 326
differential equation for, 327
generating function for, 326
integrals involving, 328
recurrence formulas for, 327
Hilbert-Schmidt treatment of integral equations, 394
Holtzmann distribution, 441
Hypergeometric polynomials, 317

Impulse function, 176
Independence of events, 426
Indeterminate forms, evaluation of, 18
Infinite products, representation of a function by, 190
Infinite series (*see* Series)
Integral equations:
Abel type of, 372
and the vibrating string, 398
characteristic values of homogeneous, 387
characteristic values of when kernel is symmetric, 391
definition of, 371
forms of solution of, 374
Fredholm's solution of, 380
Hilbert-Schmidt solution of, 394
homogeneous, 373
of the first kind, 371
of the second kind, 373
orthogonality of solution of homogeneous, 393
relation of differential equations to, 373
solution by method of inversion, 375, 376
solution by successive approximations, 378
solution of homogeneous, 387
the Volterra type of, 373
transformation of differential equations to, 373
with symmetric kernel, solution of homogeneous, 390
Integral representation of a series, 196

Integrals:
absolutely convergent, 68
Cauchy principal value of, 75
change of variable in, 49
comparison test for convergence of, 68
containing a parameter, 64
convergence of alternating, 68
convergence of infinite, 67, 68
differentiation of with respect to a parameter, 65
double integrals defined, 99
double, over general domains, 101
double, properties of, 100
elliptic (*see* Elliptic integrals)
evaluation of by double integration, 108
evaluation of by residues, 166, 167, 169, 170, 172, 176
existence of line, 97
having infinite integrands, convergence of, 74
improper, 66
indefinite, 49
inversion of order of integration in, 66
limit test for convergence of, 68
line defined, 95
line integral of total differential, 98
mean-value theorem for, 48
methods of evaluating definite, 71
methods of evaluation of, 49, 50, 52
principal value of, 174
properties of line, 97
properties of Riemann, 48
repeated, 100
Riemann, 46
Stieltjes, 76
surface, 113
surface transformation of coordinates in, 115
triple, defined, 105
with infinite integrands, tests for convergence of, 76
with many-valued integrands, 176
with parameter, inversion of order of integration and differentiation in, 71
with parameter, inversion of order of integration in, 71
with parameter, tests for uniform convergence of, 70
with parameter, uniform convergence of, 69
Integration:
complex, 148
complex, over a closed path, 149
Interval, open and closed, 2
Inverse of a function, 4
Inversions, definition of, 252

Jacobian determinant, 30
Jacobians, properties of, 34
Jacoby polynomials, 317

Lagrangian functions, 403
Lagrangian multipliers, use of, 40
Laguerre polynomials, 317

Laguerre polynomials (*Cont.*)
differential equation for, 332
generating function for, 330
integral representations of, 330, 331
integrals involving, 332, 333
recurrence relations for, 331
Laplace's equation, solution by analytic functions, 147
Legendre polynomials, 316, 337
addition formula for, 346
differential equation for, 341
generating function for, 337
integral representations for, 338, 339
integrals involving, 339, 340
recurrence formulas for, 340
Rodrigues' formula for, 339
Levy, theorem of, 434
Limit point, 2
Limits of functions, theorems on, 6, 7
Linear equations:
solutions of, 259
solutions of homogeneous, 260
Linear transformations, 265
by operational methods, 271
matrix of, 267
Lines of force, 217

Maclaurin's formula, 36
Matrices:
addition of, 269
and linear equations, 259
augmented, 259
Hermitian, 273
inverse, 269
inverse, determination of, 269
multiplication of, 268
operations with, 267
orthogonal, 274
raised to powers, 270
transposed, 267
unit, 269
unitary, 275
Matrix:
definition of, 33
rank of, 33
Maxima and minima:
criteria for, 38
with auxiliary conditions, 39
Maxwell-Boltzmann distribution, 426
Mean-value theorems for functions of one variable, 17
Mutually exclusive events, 426

Normal coordinates, 276
Normal distribution, 433

Orthogonal functions, 309
as solutions of a homogeneous integral equation, 393
Legendre polynomials, 316
Orthogonalization of functions, 313
Orthonormal functions:
involving associated Laguerre polynomials, 335
involving Bessel functions, 354
Orthonormal Functions (*Cont.*)
involving Hermitian polynomials, 317, 329
involving Legendre polynomials, 340, 344
Jacoby polynomials, 317
Laguerre polynomials, 317
special sets of, 317
trigonometric functions, 317
with continuous spectrum, 361, 363
Orthonormal functions with continuous spectra, expansion of functions in series of, 368

Permanence, definition of, 252
Permutation, definition of, 252
Piecewise continuity, definition of, 9
Poisson distribution, 442
Poisson integral formula, 223
Polynomials:
associated Laguerre (*see* Associated .. 334
Laguerre (*see* Laguerre polynomial)
Legendre, 337
Potential:
for case of cyclotron dees, 232
for rectangular poles, 228
Potential theory:
and analytic functions, 215
as a variational problem, 412
Power series:
calculations with, 141
continuation of, 155
convergence of, 139
inversion of, 185
radii of convergence of, 140
radius of convergence related to singularities, 154
reciprocal of, 142
transformation of, 142
Principle of least action, 403
Probabilities, basic events and, 424

Quadratic form, 272
matrix criterion for definiteness of, 272
matrix of, 272

Random flights, 438
Random variables, 428
continuous, 435
Region:
multiply connected, 150
simply connected, 150
Residue, 164
determination of, 165
Riemann surface, 159
Rodrigues' formula, 339

Saddle point, 39
Schwarz's inequality, 309
Schwarz-Christoffel formula, 225
Series:
absolute convergence of, 133
addition of, 133 (*also see* Power series)
asymptotic (*see* Asymptotic expansions)

Series (*Cont.*)
Burmann-Lagrange, 184
Cauchy test for absolute convergence, 138
coefficients in a Laurent, 157
convergence of, 132
convergence of Laurent, 144
differentiation of infinite, 149
example of Laurent, 182
for functions with an infinite number of poles, 186
Fourier (*see* Fourier series)
integration of infinite, 149
Laurent, 143
of complex numbers, 132
of functions of a complex variable, 135
of powers of a function, 183
products of, 133
semi-convergent (*see* Asymptotic expansions)
tests for convergence of, 136, 137
uniform convergence of, 135
Set:
countable, 1
definition of, 1
point, 2
Singularities, nature of, 156
Sources and sinks, 240
Spaces, definition of n-dimensional, 2
Space curves:
definition of, 82
direction cosines of, tangent to, 83
equation of the, tangent to, 83, 84, 85
Spherical harmonic, 344
Stokes' theorem, 123
Stream lines of a fluid, 239
Surfaces:
area of, 111
definition of, 86
sides of, 86

Taylor's theorem, 35, 153
Tensors:
addition of, 303
associate, 305
concept and definition of, 300
contraction of, 303
contravariant (order-one), 301
covariant derivative of, 306
covariant (order-one), 301
fundamental, 304
higher-order, 302
inner multiplication of, 304
magnitude of, 306
mixed, 303
product of, 303
Transformations:
conformal and orthogonal, 91
direct and inverse, 106
linear (*see* Linear transformations)
of bilinear, quadratic and Hermitian forms, 273
of coordinates, 87
of coordinates in a surface integral, 115
of Laplacian to curvilinear coordinates, 121
Transformations (*Cont.*)
of quadratic forms to principal axes, 276
of several quadratic forms to principal axes, 282
of vectors, 267
of vectors in different coordinate systems, 273
unitary, 275

Uniformly continuous function, definition of, 8
Unitary spaces, 264

Variational principle, 403
Variational problems:
and potential theory, 412
construction of the minimal sequence in, 416
direct methods of solving, 415
method of Ritz for the solution of, 416
minimal sequence for solving, 415 (*see also* Calculus of variations)
with auxiliary conditions, 413
Vector fields, 293
Vectors, characteristic or eigen-, 279
Vectors in an infinite dimensional space, 284
Vectors in n dimensions, 261
addition and subtraction of, 261
completeness of, 264
components of, 261
length or magnitude of, 262
linear independence of, 262
orthogonality of, 261
orthogonalization of, 263
scalar product of, 261
Schwarz inequality for, 262
transformations by matrices, 271
with complex components, 265
Vectors in 3 dimensions:
addition and subtraction of, 290
continuity of, 294
definition of, 288
derivatives of, 294
derivatives of scalar and vector products of, 294
divergence of, 296
gradient of, 295
integral of, 294
length of, 289, 291
relations between differential operators for, 299
relations between products of, 293
scalar product of, 290
the curl of, 298
vector product of, 291
Vector space, 262
basic vectors in, 262
Velocity potential, for an ideal fluid, 239
Vibrating string:
as a variational problem, 411
integral equation for a, 398
Ritz solution of the variational problem for, 419
Vibrations of circular membrane, 354
Volterra's integral equation, 373