August 30, 2013

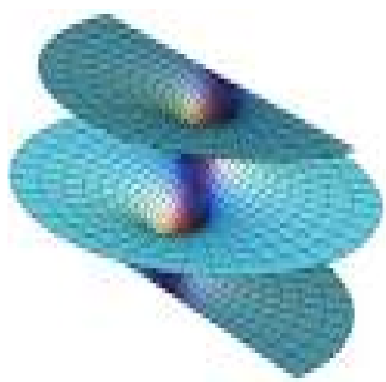
Lecture Notes on Mathematical Methods of Theoretical Physics

Giampiero Passarino *

Dipartimento di Fisica Teorica, Università di Torino, Italy
INFN, Sezione di Torino, Italy

Lecture Notes for 2013−2014

PRELIMINARY VERSION



*EMAIL: giampiero@to.infn.it

Contents

Chapter 1

Hilbert Spaces

1.1 Notations

Notation

Logic

$A \Longrightarrow B$	A implies B
$A \Longleftarrow B$	A is implied by B
iff	if and only if
$A \Longleftrightarrow B$	A implies B and is implied by B
$\forall x \in X$	for all x in X
$\exists x \in X$	there exists an x in X such that

Sets and Functions (Mappings)

$x \in X$	x is an element of the set X
$x \notin X$	x is not in X
$\{x \in X \mid P(x)\}$	the set of elements x of the set X obeying the property $P(x)$
$A \subset X$	A is a subset of X
$X \setminus A$	complement of A in X
$\overline{A}$	closure of set A
$X \times Y$	Cartesian product of X and Y
$f : X \to Y$	mapping (function) from X to Y
$f(X)$	range of f
χ_A	characteristic function of the set A
$\emptyset$	empty set
$\mathbb{N}$	set of natural numbers (positive integers)
$\mathbb{Z}$	set of integer numbers
$\mathbb{Q}$	set of rational numbers
$\mathbb{R}$	set of real numbers
$\mathbb{R}_+$	set of positive real numbers
$\mathbb{C}$	set of complex numbers

Vector Spaces

$H \oplus G$	direct sum of H and G
H^*	dual space
$\mathbb{R}^n$	vector space of n-tuples of real numbers
$\mathbb{C}^n$	vector space of n-tuples of complex numbers
l^2	space of square summable sequences
l^p	space of sequences summable with p-th power

Normed Linear Spaces

$\|\|x\|\|$	norm of x
$x_n \longrightarrow x$	(strong) convergence
$x_n \xrightarrow{w} x$	weak convergence

Function Spaces

$\operatorname{supp} f$	support of f
$H \otimes G$	tensor product of H and G
$C_0(\mathbb{R}^n)$	space of continuous functions with bounded support in $\mathbb{R}^n$
$C(\Omega)$	space of continuous functions on Ω
$C^k(\Omega)$	space of k-times differentiable functions on Ω
$C^\infty(\Omega)$	space of smooth (infinitely diffrentiable) functions on Ω
$\mathcal{D}(\mathbb{R}^n)$	space of test functions (Schwartz class)
$L^1(\Omega)$	space of integrable functions on Ω
$L^2(\Omega)$	space of square integrable functions on Ω
$L^p(\Omega)$	space of functions integrable with p-th power on Ω
$H^m(\Omega)$	Sobolev spaces
$C_0(V, \mathbb{R}^n)$	space of continuous vector valued functions with bounded support in $\mathbb{R}^n$
$C^k(V, \Omega)$	space of k-times differentiable vector valued functions on Ω
$C^\infty(V, \Omega)$	space of smooth vector valued functions on Ω
$\mathcal{D}(V, \mathbb{R}^n)$	space of vector valued test functions (Schwartz class)
$L^1(V, \Omega)$	space of integrable vector valued functions functions on Ω
$L^2(V, \Omega)$	space of square integrable vector valued functions functions on Ω
$L^p(V, \Omega)$	space of vector valued functions functions integrable with p-th power on Ω
$H^m(V, \Omega)$	Sobolev spaces of vector valued functions

Linear Operators

D^α	differential operator
$L(H, G)$	space of bounded linear transformations from H to G
$H^* = L(H, \mathbb{C})$	space of bounded linear functionals (dual space)

1.2 Hilbert Spaces: Introduction

The concept of a Hilbert space is seemingly technical and special. For example, the reader has probably heard of the space ℓ^2 (or, more precisely, $\ell^2(Z)$) of square-summable sequences of real or complex numbers.

In what follows, we mainly work over the reals in order to serve intuition, but many infinite-dimensional vector spaces, especially Hilbert spaces, are defined over the complex numbers. Hence we will write our formulae in a way that is correct also for C instead of R. Of course, for $z \in \mathbb{R}$ the expression $\mid z \mid^2$ is just z^2. We will occasionally use the fancy letter K, for Körper, which in these notes stands for either $K = \mathbb{R}$ or $K = \mathbb{C}$.

That is, ℓ^2 consists of all infinite sequences $\{\ldots, c_{-1}, c_0, c_1, \ldots\}$, $c_k \in K$, for which

$$\sum_{k=-\infty}^{\infty} \mid c_k \mid^2 < \infty \tag{1.1}$$

Another example of a Hilbert space one might have seen is the space $\mathrm{L}^2(\mathbb{R})$ of square-integrable complex-valued functions on $\mathbb{R}$, that is, of all functions $f : \mathbb{R} \to K$ for which

$$\int_{-\infty}^{\infty} dx \mid f(x) \mid^2 < \infty \tag{1.2}$$

The elements of $\mathrm{L}^2(\mathbb{R})$ are, strictly speaking, not simply functions but equivalence classes of Borel functions.

In view of their special nature, it may therefore come as a surprise that Hilbert spaces play a central role in many areas of mathematics, notably in analysis, but also including

- (differential) geometry,
- group theory,
- stochastics,
- and even number theory.

In addition, the notion of a Hilbert space provides the mathematical foundation of quantum mechanics. Indeed, the definition of a Hilbert space was first given by von Neumann (rather than Hilbert!) in 1927 precisely for the latter purpose. However, despite his exceptional brilliance, even von Neumann would probably not have been able to do so without the preparatory work in pure mathematics by Hilbert and others, which produced numerous constructions (like the ones mentioned above) that are now regarded as examples of the abstract notion of a Hilbert space.

In what follows, we shall separately trace the origins of the concept of a Hilbert space in mathematics and physics. As we shall see, Hilbert space theory is part of functional analysis, an area of mathematics that emerged between approximately $1880-1930$. Functional analysis is almost indistinguishable from what is sometimes called “abstract analysis” or “modern analysis”, which marked a break with classical analysis. The latter involves, roughly speaking, the study of properties of a single function, whereas the former deals with spaces of functions. The modern concept of a function as a map $f : [a,b] \to \mathbb{R}$ was only arrived at by Dirichlet as late as 1837, following earlier work by notably Euler and Cauchy. But Newton already had an intuitive graps of this concept, at least for one variable.

One may argue that classical analysis is tied to classical physics, whereas modern analysis is associated with quantum theory. Of course, both kinds of analysis were largely driven by intrinsic mathematical arguments as well. 5 The jump from classical to modern analysis was as discontinuous as the one from classical to quantum mechanics. The following anecdote may serve to illustrate this. G.H. Hardy was one of the masters of classical analysis and one of the most famous mathematicians altogether at the beginning of the 20th century. John von Neumann, one of the founders of modern analysis, once gave a talk on this subject at Cambridge in Hardy's presence. Hardy's comment was: Obviously a very intelligent man. But was that mathematics?

The final establishment of functional analysis and Hilbert space theory around 1930 was made possible by combining a concern for rigorous foundations with an interest in physical applications [2].

Classical analysis grew out of the calculus of Newton, which in turn had its roots in both geometry and physics. (Some parts of the calculus were later rediscovered by Leibniz.) In the 17th century, geometry was a practical matter involving the calculation of lenths, areas, and volumes. This was generalized by Newton into the calculus of integrals. Physics, or more precisely mechanics, on the other hand, had to do with velocities and accellerations and the like. This was abstracted by Newton into differential calculus. These two steps formed one of the most brilliant generalizations in the history of mathematics, crowned by Newton's insight that the operations of integration and differentiation are inverse to each other, so that one may speak of a unified differential and integral calculus, or briefly calculus. Attempts to extend the calculus to more than one variable and to make the ensuing machinery mathematically rigorous in the modern sense of the word led to classical analysis as we know it today. (Newton used theorems and proofs as well, but his arguments would be called "heuristic" or "intuitive" in modern mathematics.)

1.2.1 Origins in mathematics

The key idea behind functional analysis is to look at functions as points in some infinite-dimensional vector space. To appreciate the depth of this idea, it should be mentioned that the concept of a finite-dimensional vector space only emerged in the work of Grassmann between 1844 and 1862 (to be picked up very slowly by other mathematicians because of the obscurity of Grassmann's writings), and that even the far less precise notion of a "space" (other than a subset of $\mathbb{R}^n$) was not really known before the work of Riemann around 1850.

Indeed, Riemann not only conceived the idea of a manifold (albeit in embryonic form, to be made rigorous only in the 20th century), whose points have a status comparable to points in $\mathbb{R}^n$, but also explicitly talked about spaces of functions (initially analytic ones, later also more general ones). However, Riemann's spaces of functions were not equipped with the structure of a vector space.

In 1885 Weierstrass considered the distance between two functions (in the context of the calculus of variations), and in 1897 Hadamard took the crucial step of connecting the set-theoretic ideas of Cantor with the notion of a space of functions.

Finally, in his PhD thesis of 1906, which is often seen as a turning point in the development of functional analysis, Hadamard's student Fréchet defined what is now called a metric space (i.e., a possibly infinite-dimensional vector space equipped with a metric, see below), and gave examples of such spaces whose points are functions. Fréchet's main example was $C[a,b]$, seen as a metric space in the supremum-norm, i.e., $d(f,g) = \| f-g \|$ with $\| f \| = \sup\{f(x) \mid x \in [a,b]\}$

After 1914, the notion of a topological space due to Hausdorff led to further progress, eventually leading to the concept of a topological vector space, which contains all spaces mentioned below as special cases.

To understand the idea of a space of functions, we first reconsider $\mathbb{R}^n$ as the space of all functions $f : \{1,2,\ldots,n\} \to \mathbb{R}$, under the identification $x_1 = f(1),\ldots,x_n = f(n)$. Clearly, under this identification the vector space operations in $\mathbb{R}^n$ just correspond to pointwise operations on functions (e.g., $f+g$ is the function defined by $(f+g)(k) = f(k)+g(k)$, etc.). Hence $\mathbb{R}^n$ is a function space itself, consisting of functions defined on a finite set. The given structure of $\mathbb{R}^n$ as a vector space may be enriched by defining the length f of a vector f and the associated distance $d(f,g) = \| f-g \|$ between two vectors f and g. In addition, the angle θ

between f and g in $\mathbb{R}^n$ is defined. Lengths and angles can both be expressed through the usual inner product

$$(f,g) = \sum_{k=1}^{n} \overline{f(k)}\, g(k) \tag{1.3}$$

through the relations $\|f\| = \sqrt{(f,f)}$ and $(f,g) = \|f\|\,\|g\|\cos\theta$ ($\overline{f}$ is the complex conjugate of f).

In particular, one has a notion of orthogonality of vectors, stating that f and g are orthogonal whenever $(f,g) = 0$, and an associated notion of orthogonality of subspaces: we say that $V \subset \mathbb{R}^n$ and $W \subset \mathbb{R}^n$ are orthogonal if $(f,g) = 0$ for all $f \in V$ and $g \in W$. This, in turn, enables one to define the (orthogonal) projection of a vector on a subspace of $\mathbb{R}^n$. Even the dimension n of $\mathbb{R}^n$ may be recovered from the inner product as the cardinality of an arbitrary orthogonal basis. Now replace $\{1,2,\ldots,n\}$ by an infinite set. In this case the corresponding space of functions will obviously be infinite-dimensional in a suitable sense. The simplest example is $\mathbb{N} = \{1,2,\ldots,\}$, so that one may define $\mathbb{R}^\infty$ as the space of all functions $f : \mathbb{N} \to \mathbb{R}$, with the associated vector space structure given by pointwise operations. However, although $\mathbb{R}^\infty$ is well defined as a vector space, it turns out to be impossible to define an inner product on it, or even a length or distance.

Fréchet's main example was $C[a,b]$, seen as a metric space in the supremum-norm, i.e., $d(f,g) = \|f - g\|$ with $\|f\| = \sup\{f(x) \mid x \in [a,b]\}$

A subspace of a vector space is by definition a linear subspace.

This is most easily done by picking a basis e_i of the particular subspace V. The projection pf of f onto V is then given by $pf = \sum_i (e_i, f)\, e_i$.

This is the same as the cardinality of an arbitrary basis, as any basis can be replaced by an orthogonal one by the Gram-Schmidt procedure.

The dimension of a vector space is defined as the cardinality of some basis. The notion of a basis is complicated in general, because one has to distinguish between algebraic (or Hamel) and topological bases. Either way, the dimension of the spaces described below is infinite, though the cardinality of the infinity in question depends on the type of basis. The notion of an algebraic basis is very rarely used in the context of Hilbert spaces (and more generally Banach spaces), since the ensuing dimension is either finite or uncountable. The dimension of the spaces below with respect to a topological basis is countably infinite, and for a Hilbert space all possible cardinalities may occur as a possible dimension. In that case one may restrict oneself to an orthogonal basis.

Indeed, defining

$$(f,g) = \sum_{k=1}^{\infty} \overline{f(k)}\, g(k) \tag{1.4}$$

it is clear that the associated length f is infinite for most f. This is hardly surprising, since there are no growth conditions on f at infinity. The solution is to simply restrict $\mathbb{R}^\infty$ to those functions with $\|f\| < \infty$. These functions by definition form the set $\ell^2(N)$, which is easily seen to be a vector space. Moreover, it follows from the Cauchy-Schwarz inequality

$$(f,g) \leq \|f\|\,\|g\| \tag{1.5}$$

that the inner product is finite on $\ell^2(N)$. Consequently, the entire geometric structure of $\mathbb{R}^n$ in so far as it relies on the notions of lengths and angles (including orthogonality and orthogonal projections) is available on $\ell^2(N)$. Running ahead of the precise definition, we say that $\mathbb{R}^n \cong \ell^2(\{1,2,\ldots,n\})$ is a finite-dimensional Hilbert space, whereas $\ell^2(N)$ is an infinite-dimensional one. Similarly, one may define $\ell^2(Z)$ (or indeed $\ell^2(S)$ for any countable set S) as a Hilbert space in the obvious way.

From a modern perspective, $\ell^2(N)$ or $\ell^2(Z)$ are the simplest examples of infinite-dimensional Hilbert spaces, but historically these were not the first to be found. From the point of view of most mathematicians around 1900, a space like $\ell^2(N)$ would have been far to abstract to merit consideration.

The initial motivation for the concept of a Hilbert space came from the analysis of integral equations of the type

$$f(x)+\int_a^b dyK(x,y)f(y)=g(x) \tag{1.6}$$

where f, g, and K are continuous functions and f is unknown. Integral equations were initially seen as reformulations of differential equations. For example, the differential equation $Df=g$ or $f'(x)=g(x)$ for unknown f is solved by $f=\int g$ or $f(x)=\int_0^x dyg(y)=\int_0^1 dyK(x,y)g(y)$ for $K(x,y)=\theta(x-y)$ (where $x\leq 1$), which is an integral equation for g.

Such equations were first studied from a somewhat modern perspective by Volterra and Fredholm around 1900, but the main break-through came from the work of Hilbert between 1904−1910. In particular, Hilbert succeeded in relating integral equations to an infinite-dimensional generalization of linear algebra by choosing an orthonormal basis $\{e_k\}$ of continuous functions on $[a,b]$ (such as $e_k(x)=\exp(2\pi kix)$ on the interval $[0,1]$), and defining the (generalized) Fourier coefficents of f by $\hat{f}_k=(e_k,f)$ with respectto the inner product

$$(f,g)=\int_a^b dx\overline{f(x)}g(x) \tag{1.7}$$

The integral equation (I.6) is then transformed into an equation of the type

$$\hat{f}_k=\sum_l \hat{K}_{kl}\hat{f}_l=\hat{g}_l \tag{1.8}$$

Hilbert then noted from the Parseval relation (already well known at the time from Fourier analysis and more general expansions in eigenfunctions)

$$\sum_{k\in Z}|\hat{f}_k|^2=\int_a^b dx\,|f(x)|^2 \tag{1.9}$$

that the left-hand side is finite, so that $\hat{f}\in\ell^2(Z)$. This, then, led him and his students to study ℓ^2 also abstractly. E. Schmidt should be mentioned here in particular. Unlike Hilbert, already in 1908 he looked at ℓ^2 as a "space" in the modern sense, thinking of seqences (c_k) as point in this space. Schmidt studied the geometry of ℓ^2 as a Hilbert space in the modern sense, that is, empasizing the inner product, orthogonality, and projections, and decisively contributed to Hilbert's work on spectral theory. The space $\mathrm{L}^2(a,b)$ appeared in 1907 in the work of F. Riesz and Fischer as the space of (Lebesgue) integrable functions on (a,b) for which

$$\int_a^b dx\,|f(x)|^2<\infty \tag{1.10}$$

of course, this condition holds if f is continuous on $[a,b]$. 14 More precisely, the elements of $\mathrm{L}^2(a,b)$ are not functions but equivalence classes thereof, where $f\sim g$ when $\|f-g\|_2=0$.

Equipped with the inner product (I.7), this was another early example of what is now called a Hilbert space. 15 The term "Hilbert space" was first used by Schoenflies in 1908 for ℓ^2, and was introduced in the abstract sense by von Neumann in 1927; see below. The context of its appearance was what is now called the Riesz-Fischer theorem:

Given any sequence (c_k) of real (or complex) numbers and any orthonormal system (e_k) in $\mathrm{L}^2(a,b)$, there exists a function $f\in\mathrm{L}^2(a,b)$ for which $(e_k,f)=c_k$ if and only if $c\in\ell^2$, i.e., if $\sum_k|c_k|^2<\infty$. The notion of an orthonormal system of functions on the interval $[a,b]$ was as old as Fourier, and was defined abstractly by Hilbert in 1906. At the time, the Riesz-Fischer theorem was completely unexpected, as it proved that two seemingly totally different spaces were "the same" from the right point of view. In modern terminology, the theorem establishes an isomorphism of ℓ^2 and L^2 as Hilbert spaces, but this point of view was only established twenty years later, i.e., in 1927, by von Neumann. Inspired by quantum mechanics, in that year von Neumann gave the definition of a Hilbert space as an abstract mathematical structure, as follows. First, an inner product on a vector space V over a field K (where $K=\mathbb{R}$ or $K=\mathbb{C}$), is a map $V\times V\to K$, written as $\langle f,g\rangle\mapsto(f,g)$, satisfying, for all $f,g\in V$ and $t\in K$,

1. $(f,f)\geq 0$;
2. $(g,f)=\overline{(f,g)}$;
3. $(f,tg)=t\,(f,g)$;
4. $(f,g+h)=(f,g)+(f,h)$;
5. $(f,f)=0\Rightarrow f=0$.

Given an inner product on V , one defines an associated length function or norm $\|\,.\,\|:V\to\mathbb{R}^+$ by (I.2). A Hilbert space (over K) is a vector space (over K) with inner product, with the property that Cauchy sequences with respect to the given norm are convergent (in other words, V is complete in the given norm). A sequence (f_n) is a Cauchy sequence in V when $\|f_n-f_m\|\to 0$ when $n,m\to\infty$; more precisely, for any $\varepsilon>0$ there is $l\in\mathbb{N}$ such that $\|f_n-f_m\|<\varepsilon$ for all $n,m>l$. A sequence (f_n) converges if there is $f\in V$ such that $\lim_{n\to\infty}\|f_n-f\|=0$.

Hilbert spaces are denoted by the letter H rather than V. Thus Hilbert spaces preserve as much as possible of the geometry of $\mathbb{R}^n$.

It can be shown that the spaces mentioned above are Hilbert spaces. Defining an isomorphism of Hilbert spaces $U:H_1\to H_2$ as an invertible linear map preserving the inner product (i.e., $(Uf,Ug)_2=(f,g)_1$ for all $f,g\in H_1$), the Riesz-Fischer theorem shows that $\ell^2(Z)$ and $L^2(a,b)$ are indeed isomorphic.

In a Hilbert space the inner product is fundamental, the norm being derived from it. However, one may instead take the norm as a starting point (or, even more generally, the metric, as done by Fréchet in 1906). The abstract properties of a norm were first identified by Riesz in 1918 as being satisfied by the supremum norm, and were axiomatized by Banach in his thesis in 1922. A norm on a vector space V over a field K as above is a function $\|\,.\,\|:V\to\mathbb{R}^+$ with the properties:

1. $\|f+g\|\leq\|f\|+\|g\|$ for all $f,g\in V$;
2. $\|tf\|=|t|\,\|f\|$ for all $f\in V$ and $t\in K$;
3. $\|f\|=0\Rightarrow f=0$.

The usual norm on $\mathbb{R}^n$ satisfies these axioms, but there are many other possibilities, such as

$$\|f\|_p=\left(\sum_{k=1}^{n}|f(k)|^p\right)^{1/p} \tag{1.11}$$

for any $p\in\mathbb{R}$ with $1\leq p<\infty$, or

$$\|f\|_\infty=\sup\{|f(k)|,k=1,\ldots,n\}. \tag{1.12}$$

In the finite-dimensional case, these norms (and indeed all other norms) are all equivalent in the sense that they lead to the same criterion of convergence (technically, they generate the same topology): if we say that

$f_n \to f$ when $\|f_n - f\| \to 0$for some norm on $\mathbb{R}^n$, then this implies convergence with respect to any other norm. This is no longer the case in infinite dimension. For example, one may define $l^p(N)$ as the subspace of $\mathbb{R}^\infty$ that consists of all vectors $f \in \mathbb{R}^\infty$ for which

$$\| f \|_p = \left(\sum_{k=1}^{\infty} \mid f(k) \mid^p \right)^{1/p} \tag{1.13}$$

is finite. It can be shown that $\| \centerdot \|_p$ is indeed a norm on $\ell^p(N)$, and that this space is complete in this norm. As with Hilbert spaces, the examples that originally motivated Riesz to give his definition were not ℓ^p spaces but the far more general L^p spaces, which he began to study in 1910. For example, $L^p(a,b)$ consists of all (equivalence classes of Lebesgue) integrable functions f on (a,b) for which

$$\| f \|_p = \left(\int_a^b dx \mid f(x) \mid^p \right)^{1/p} \tag{1.14}$$

is finite, still for $1 \leq p < \infty$, and also $\| f \|_\infty = \sup\{\mid f(x) \mid, x \in (a,b)\}$. Eventually, in 1922 Banach defined what is now called a Banach space as a vector space (over K as before) that is complete in some given norm. Long before the abstract definitions of a Hilbert space and a Banach space were given, people began to study the infinite-dimensional generalization of functions on $\mathbb{R}^n$. In the hands of Volterra, the calculus of variations originally inspired the study of functions $\phi : V \to K$, later called functionals, and led to early ideas about possible continuity of such functions. However, although the calculus of variations involved nonlinear functionals as well, only linear functionals turned out to be tractable at the time (until the emergence of nonlinear functional analysis much later). Indeed, even today (continuous) linear functionals still form the main scalar-valued functions that are studied on infinite-dimensional (topological) vector spaces. For this reason, throughout this text a functional will denote a continuous linear functional. For $H = \mathrm{L}^2(a,b)$, it was independently proved by Riesz and Frćhet in 1907 that any functional on H is of the form $g \mapsto (f,g)$ for some $f \in H$. More generally, in 1910 Riesz showed that any functional on $L^p(a,b)$ is given by an element $L^q(a,b)$, where $1/p + 1/q = 1$, by the same formula. Since $p = 2$ implies $q = 2$, this of course implies the earlier Hilbert space result.

The same result for arbitrary Hilbert spaces H was written down only in $1934-35$, again by Riesz, although it is not very difficult. The second class of "functions" on Hilbert spaces and Banach spaces that could be analyzed in detail were the generalizations of matrices on $\mathbb{R}^n$, that is, linear maps from the given space to itself. Such functions are now called operators. Or linear operators, but for us linearity is part of the definition of an operator. For example, the integral equation (I.6) is then simply of the form $(1+K)f = g$, where $1 : \mathrm{L}^2(a,b) \to \mathrm{L}^2(a,b)$ is the identity operator $1f = f$, and $K : \mathrm{L}^2(a,b) \to \mathrm{L}^2(a,b)$ is the operator given by

$$(Kf)(x) = \int_a^b dy K(x,y) f(y). \tag{1.15}$$

This is easy for us to write down, but in fact it took some time before integral of differential equations were interpreted in terms of operators acting on functions. For example, Hilbert and Schmidt did not really have the operator concept but (from the modern point of view) worked in terms of the associated quadratic form. That is, the operator $a : H \to H$ defines a map $q : H \times H \to K$ by $\langle f, g \rangle \mapsto (f, ag)$. They managed to generalize practically all results of linear algebra to operators, notably the existence of a complete set of eigenvectors for operators of the stated type with symmetric kernel, that is, $K(x,y) = K(y,x)$. The associated quadratic form then satisfies $q(f,g) = q(g,f)$.

The abstract concept of a (bounded) operator (between what we now call Banach spaces) is due to Riesz in 1913. It turned out that Hilbert and Schmidt had studied a special class of operators we now call "compact", whereas an even more famous student of Hilbert's, Weyl, had investigated a singular class of operators now called "unbounded" in the context of ordinary differential equations. Spectral theory and eigenfunctions expansions were studied by Riesz himself for general bounded operators on Hilbert spaces (seen by him as a special case of general normed spaces), and later, more specifically in the Hilbert space case, by Hellinger and Toeplitz (culminating in their pre-von Neumann review article of 1927). In the Hilbert space case, the results of all these authors were generalized almost beyond recognition by von Neumann in his book from 1932 [15], to whose origins we now turn.

John von Neumann (1903−1957) was a Hungarian prodigy; he wrote his first mathematical paper at the age of seventeen. Except for this first paper, his early work was in set theory and the foundations of mathematics. In the Fall of 1926, he moved to Göttingen to work with Hilbert, the most prominent mathematician of his time. Around 1920, Hilbert had initiated his Beweistheory, an approach to the axiomatization of mathematics that was doomed to fail in view of Gödel's later work. However, at the time that von Neumann arrived, Hilbert was mainly interested in quantum mechanics.

1.2.2 Why Hilbert space (level 0)?

Chapter V

Quantum mechanics and Hilbert space I: states and observables

We are now going to apply the previous machinery to quantum mechanics, referring to the Introduction for history and motivation. The mathematical formalism of quantum mechanics is easier to understand if it is compared with classical mechanics, of which it is a modification. We therefore start with a rapid overview of the latter, emphasizing its mathematical structure.

V.1 Classical mechanics

The formalism of classical mechanics is based on the notion of a **phase space** M and **time-evolution**, going back to Descartes and Newton, and brought into its modern form by Hamilton. The phase space of a given physical system is a collection of points, each of which is interpreted as a possible state of the system. At each instance of time, a given state is supposed to completely characterize the 'state of affairs' of the system, in that:

1. The value of any observable (i.e., any question that may possibly be asked about the system, such as the value of its energy, or angular momentum,...) is determined by it.[1]

2. Together with the equations of motion, the state at $t = 0$ is the only required ingredient for the prediction of the future of the system.[2]

Observables are given by functions f on M. The relationship between states (i.e. points of M) and observables is at follows:

The value of the observable f in the state x is $f(x)$.

This may be reformulated in terms of questions and answers. Here an observable f is identified with the question: *what is the value of f?* A state is then a list of answers to all such questions.

A very basic type of observable is defined by a subset $S \subset M$. This observable is the characteristic function χ_S of S, given by $\chi_S(x) = 1$ when $x \in S$ and $\chi_S(x) = 0$ when $x \notin S$. The corresponding question is: *is the system in some state lying in $S \subset M$?* The answer yes is identified with the value $\chi_S = 1$ and the answer no corresponds to $\chi_S = 0$. Sich a question with only two possible answers is called a **yes-no question**.

In these notes we only look at the special case $M = \mathbb{R}^{2n}$, which describes a physical system consisting of a point particles moving in $\mathbb{R}^n$. We use coordinates $(q, p) := (q^i, p_i)$, where $i = 1, \ldots, n$. The q variable ("position") denotes the position of the particle, whereas the meaning of

[1] Philosophers would say that any quantity pertaining to the system *supervenes* on its states; this means that no change in a given quantity is possibly without a change in the state. For example, most scientists would agree that the mind supervenes on the brain (seen as a physical system).

[2] We do not say that such a prediction is always possible in practice. But *if* it is possible at all, it merely requires the state and the equations of motion.

the p variable ("momentum") depends on the time-evolution of the system. For example, for a free particle of mass m one has the relation $\vec{p} = m\vec{v}$, where v is the velocity of the particle (see below). Let us note that one may look at, say, q^i also as an observable: seen as a function on M, one simply has $q^i(q,p) = q^i$, etc.

Given the phase space M, the specification of the system is completed by specifying a function h on M, called the **Hamiltonian** of the system. For $M = \mathbb{R}^{2n}$ we therefore have h as a function of (q,p), informally written as $h = h(q,p)$. The Hamiltonian plays a dual role:

- Regarded as an observable it gives the value of the energy;
- it determines the time-evolution of the system.

Indeed, given h the time-evolution is determined by **Hamilton's equations**

$$\begin{aligned} \dot{q}^i := \frac{dq^i}{dt} &= \frac{\partial h}{\partial p_i}; \\ \dot{p}_i := \frac{dp_i}{dt} &= -\frac{\partial h}{\partial q^i}. \end{aligned} \tag{V.1}$$

For example, a particle with mass m moving in a potential V has Hamiltonian

$$h(q,p) = \frac{p^2}{2m} + V(q), \tag{V.2}$$

where $p^2 := \sum_{i=1}^n (p_i)^2$. The equations (V.1) then read $\dot{q}^i = p_i/m$ and $\dot{p}_i = -\partial V/\partial q^i$. With the force defined by $F^i := -\partial V/\partial q^i$, these are precisely Newton's equations $d^2q^i/dt^2 = F^i/m$, or $\vec{F} = m\vec{a}$. In principle, h may explicitly depend on time as well.

V.2 Quantum mechanics

Quantum mechanics is based on the postulate that the phase space is a Hilbert space H, with the additional stipulations that:

1. Only vectors of norm 1 correspond to physical states;
2. Vectors differing by a "phase", i.e., by a complex number of modulus 1, correspond to the same physical state.

In other word, $\psi \in H$ and $z\psi$ with $z \in \mathbb{C}$ and $|z| = 1$ give the same state.[3] We here stick to the physicists' convention of denoting elements of Hilbert spaces by Greek letters.[4]

The reason for the first point lies in the probability interpretation of quantum mechanics. The simplest example of this interpretation is given by the quantum mechanics of a particle moving in $\mathbb{R}^3$. In that case the Hilbert space may be taken to be $H = L^2(\mathbb{R}^3)$, and Born and Pauli claimed in 1926 that the meaning of the 'wavefunction' $\psi \in L^2(\mathbb{R}^3)$ was as follows: *the probability $P(\psi, x \in \Delta)$ that the particle in state ψ is found to be in a region $\Delta \subseteq \mathbb{R}^3$ is*

$$P(x \in \Delta|\psi) = (\psi, \chi_\Delta \psi) = \int_\Delta d^3x \, \|\psi(x)\|^2. \tag{V.3}$$

Here χ_Δ is the characteristic function of Δ, given by $\chi_\Delta(x) = 1$ when $x \in \Delta$ and $\chi_\Delta(x) = 0$ when $x \notin \Delta$. It follows that

$$P(x \in \mathbb{R}^n|\psi) = \|\psi\|^2 = (\psi,\psi) = 1, \tag{V.4}$$

[3]It follows that the true state space of a quantum-mechanical system is the **projective** Hilbert space $\mathbb{P}H$, which may be defined as the quotient $SH/\sim$, where $SH := \{f \in H \mid \|f\| = 1\}$ and $f \sim g$ iff $f = zg$ for some $z \in \mathbb{C}$ with $|z| = 1$.

[4]This notation was initially used by Schrödinger in order to make his wave mechanics, a precursor of quantum mechanics, look even more mysterious than it already was.

1.2.3 Why Hilbert space (level 1)?

One could go through great pains to learn the profound mathematics of Hilbert spaces and operators on them but what in experiment suggests the specific form of quantum mechanics with its "postulates"? Why should measurable quantities be represented by operators on a Hilbert space? Why should the complete information about a system be represented by a vector from a Hilbert space?

It looks like we make a lot of assumptions for setting up quantum mechanics. The arguments below will show, that we make one less than we make for classical mechanics, and that this intails all the strangeness. It is a bit like in general relativity: by omitting one postulate from geometry, we enter a whole new space of possibilities.

Overview

- Associate "physical quantity" Q with a mathematical object Q
- Key step: Q should be part of an algebra
- Define the "state" of a system, that leads to an expectation value for any measurement on the system.
- Given a "state" and the algebra of observables, a Hilbert space can be constructed and the observables will be represented on it as linear operators in the Hilbert space (GNS representation).
- In turn, any Hilbert space that allows representing the algebra of observables as linear operators on it is equivalent to a direct sum of GNS representations.

Physical quantities and observables Assume we have "physical quantities" Q. The concept is deliberately vague to leave it as general as possible. Itshould be a number or set of numbers that are associated with something we can measure directly or determine indirectly by a set of manipulations on an something we call a physical system. For being able to establish logically consistent relations between physical quantities, we want to map them into mathematical objects.

An algebra for the observables Let Q be the mathematical objects corresponding to the physical quantities. We want to be able to perform basic algebraic operations with them: add them, multiply them, scale them, in brief: they should be members of an "algebra" A:

1. A is a vector space
2. there is a multiplication: $(P,Q) \to O =: PQ \in A$
3. $P(Q_1 + Q_2) = PQ_1 + PQ_2$ for $P, Q_1, Q_2 \in A$
4. $P(\alpha Q) = \alpha(PQ)$ for $\alpha \in C$
5. $\exists \mathbb{1} \in A : \mathbb{1}Q = Q\mathbb{1} = Q$

We would need to come back later and see what "physical quantity" may correspond to $Q+Q$ or QQ. How can we map this back into $Q+Q$ etc.? A few extra mathematical properties we want for our Q:

1. There should be a norm: we should be able to tell how "large" at most the quantity is and it should be compatible with multiplicative structure of the algebra. This norm should have the properties $\|PQ\| \leq \|P\| \, \|Q\}$, $\|\mathbb{1}\| = \mathbb{1}$

2. There should be an adjoint element with all the properties of a hermitian conjugate

3. A should be complete: if a sequence of Q_n "converges" (Cauchy series), the should be an element in A such that the series converges to this element.

It is not obvious, whether these further assumptions are innocent or already imply deep postulates on the physics we want to mirror in our mathematics. Note, however, that the "observables" of classical mechanics are simply functions on the phase space and have all those properties with the norm $\|F\| = \sup_{x,p}$ mod $F(x,p)$ |, if we restrict ourselves to bounded functions: after all, there is no single apparatus to measure infinitely large values. Note, that in this sense momentum p would not fit into the algebra, as it is unbounded. However, momentum restriced to any finite range does fit. An algebra with these properties is called C-algebra.

Spectrum of Q Any physical quantity Q can assume a set of values, which should be identical with the spectrum of Q: Q, so to speak, is a general container for the possible values of the measurable quantity.

Let us remark that the spectrum $\sigma(Q)$ of an element of the algebra can be defined just like for a linear operator by looking at $(Q-z)^{-1}$ for $z \in \mathbb{C}$:

$$\sigma(Q) = \mathbb{C} \mid \{z \in \mathbb{C} \mid \exists (Q-z)^{-1}\} \tag{1.16}$$

The state of a system We call an element of the algebra positive, if its spectrum has strictly non-negative values. A more fundamental definition of positivity is $A > 0 : A = B^* B, B \in A$. Using the definition of "positive", we can introduce a partial ordering in the algebra by The "state of a system" is a positive linear functional f with $f(1) = \mathbb{1}$

- Linear: $f(\alpha A + \beta) = \alpha f(A) + \beta f(B)$, we want that ... (we have it in classical mechanics).

- Positive: $f(Q) \geq 0$ for $Q > 0$. Note the $\geq$ rather than $>$: the observable Q may well have spectral values $= 0$ in certain places. If a state f_0 only tests these places, the result $f_0(Q) = 0$, although $Q > 0$.

A state f is a very general definition of what we expect of something that gives an "expectation value" $f(Q)$ for a physical quantity Q: linear, not negative if there are only positive values available, and $= 1$, if we only measure that the system exists at all, without referring to any physical property ($Q = 1$).

Gelfand isomorphism Can be paraphrased as follows: "Any commuting C-algebra is equivalent to an algebra of continuous functions from the character set of the algebra into the complex numbers. A character of an abelian C-algebra is a homomorphism of the algebra into the complex numbers.

If you are not familiar with the concept of "character", for simplicity, think of subset of the linear operators on the Hilbert space and imagine a single character as a common eigenvector shared by all the operators (bypassing also the question of possibly degenerate eigenvalues). For the algebra formed by the polynomials of a single "normal" operator, a character can be associated with a given spectral value of the operator. The character set is nothing newly introduced, no new condition on our algebra: given a commuting C^* algebra, we can give its character set. For defining "continuity" we need a topology. A weak topology is used: a sequence of characters χ_n converges, if the sequence of real numbers $\chi_n(Q)$ converges for each $Q \in A$.

Illustration on the example of bounded linear operators Gelfand isomorphism is the correspondence of what you know as "any set of commuting operators can be diagonalized simultaneously". The statement can be derived without any reference to a "Hilbert space" or 'states on the Hilbert space". It only uses the precise definition of "character". However, to develop a feeling for its meaning, in can be helpfull to discuss a very simplified version for linear operators on a Hilbert space. Assume a commutative C^* algebra A of bounded linear operators on a Hilbert space. You have learnt that all operators of the algebra can be "diagonalized" simultaneously. Assume for simplicity that all these operators have a strictly discrete spectrum with eigenfunctions from the Hilbert space. Let $\{|i\rangle\}$ denote the set of simultaneous eigenvectors of all operators in the algebra. Assume for simplicity that we can choose the $|i\rangle$ orthonormal: $\langle i|j\rangle = \delta_{ij}$. Then any operator $A \in A$ can be written as

$$A = \sum_{\chi_i} |\chi_i\rangle\, f_{\mathrm{A}}(i)\, \langle\chi_i| = \sum_{\chi_i} |\chi_i\rangle\, \chi_i(A)\, \langle chi_i| \tag{1.17}$$

The set shared eigenvectors $\{|i\rangle\}$ defines the character set $X = \{\chi_i\}$ of this particular C^* algebra:

$$\chi(A) = \langle i|\, A\, |i\rangle \tag{1.18}$$

The $f_{\mathrm{A}}(i)$ can be understood as mapping one particular character χ_i into the complex numbers.

States, measures, and integrals We can identify states with integration measures on the spectrum of an element Q. What is an integral? It is a continuous (in the sense of some topology) linear map from a space of functions f into the complex numbers. We write it as

$$\int_{\chi\in X} d\mu(\chi) f(\chi) \tag{1.19}$$

where $d(\chi)$ is the integration "measure" on the set of χs.

What is a state? It is a (positive, continuous) linear map from the C^* algebra into the complex numbers.

State $\Leftrightarrow$ measure on the character set. To the extent that we can associate the character set with the "spectrum" of the observable any measure on the character set is a measure on the spectrum.

Using again the analogy with a C^* algebra of bounded linear operators: a state can be constructed using a "density matrix" ρ by:

$$f_\rho(A) = \mathrm{Tr}\,\rho A \tag{1.20}$$

In the simplest case of a pure state $\rho = |\Psi\rangle\,\langle\Psi|$ ($\|\Psi\| = 1$)

$$f_\rho(a) = \mathrm{Tr}\,\rho A = \sum_i |\,\langle\Psi|\,|i\rangle\,|_2\ \chi_i(A) =: \sum_{\chi\in X} \mu(\chi) f_{\mathrm{A}}(\chi) \tag{1.21}$$

The integration measure induced by f_ρ with $\rho = |\Psi\rangle\,\langle\Psi|$ is just $\mu(\chi_i) =|\,\langle\Psi|\,|i\rangle\,|_2$. We are back to the simplest quantum mechanical situation.

The structure of classical and quantum mechanics Postulate: Observables of a physical system are described by hermitian $Q = Q^*$ elements of a C^* algebra A and the state of a physical system is mapped into a state on A. The possible measurement results for Q are the spectrum of Q and their probability distribution in a state f is given by the measure df, which is the probability measure induced by it on the spectrum of Q.

In classical mechanics, all Q commute. In quantum mechanics, they do not commute. Here is the fundamen- tal mathematical difference between the two theories.

Where is the Hilbert space? C^* algebras can always be represented as bounded operators on a Hilbert space. "Represented": what matters about the algebra is its addition and multiplication laws, and, as it is C^* also the conjugation operation. Let $\pi : A \to B(H)$ be a mapping that assigns a bounded operator on the Hilbert space to each element of $A : Q \to \pi(Q)$. We call π a *-homomorphism, if it preserves the C^* algebraic properties. If we have a state f, we can use it to define a scalar product

$$\langle Q|Q\rangle = f(Q^*Q) \tag{1.22}$$

and with it turn the algebra into a Hilbert space. We can then use this Hilbert space to represent the algebra on it. Let us call this Hilbert space $H(A, f)$. Note: $f(Q^*Q)$ will not be a legitimate scalar product on the complete algebra as in general there will be $0 = A \in A$ such that $f(A^*A) = 0$. This can be fixed, loosely speaking, by removing those A from the space used for representing A. Using the concepts of quotient algebra and left sided ideal this can be done as follows: first observe that the $A \in \mathbb{N}$ with $f(A^*A) = 0$ are a left-sided ideal of the algebra:

$$f(A^*A) = 0 \Rightarrow f(A^*B^*BA) = 0 \forall B \in A \tag{1.23}$$

To go into the quotient algebra

$$A/N = \{[B] \mid B \in A\}, \quad [B] = \{B + A \mid A \in \mathbb{N}\} \tag{1.24}$$

The scalar product of quotient algebra is defined by

$$\langle [B]|[B]\rangle = \inf_{A\in\mathbb{N}} f((B+A) * (B+A)) > 0 \tag{1.25}$$

for $[B] = [0]$. Note that $[0] = A|A \in \mathbb{N}$.

GNS representation (Gelfand, Naimark, Segal) Having constructed $H(A, f)$ we get a representation of the algebra on that Hilbert space as follows. Let $|q\rangle \in H(A, f)$ be the vector in the Hilbert space that corresponds to an element $Q \in A$. Let P be any element in A. Then $PQ \in A$ with a corresponding $|pq\rangle \in H(A, f)$. We define the linear operator on $H(A, f)$ $\pi_f(P) : |q\rangle \to \pi_f(P)\,|q\rangle = |pq\rangle$.

A vector $|c\rangle$ from the a Hilbert space is called "cyclic" w.r.t. a representation π, if the vectors $\{\pi(Q)\,|c\rangle \mid \pi(Q) \in \pi(A)\}$ are dense in the Hilbert space.

Irreducibility of a representation can be also phrased as: all vectors of the Hilbert space are cyclic. By construction, the vector corresponding to $|\mathbb{1}\rangle_f$ in the GNS representation π_f for state f representation is cyclic.

Pure-state $\Leftrightarrow$ GNS construction is irreducible. States form a convex space, i.e. if f_1 and f_2 are states, then also $f = \alpha f_1 + (1-\alpha) f_2, \alpha \in [0,1]$ is a state. States that cannot be decomposed in this way are called "pure". Without discussing this further, we mention

- The density matrix corresponding to pure states has the form $\rho = |\Psi\rangle\,\langle\Psi|$
- The GNS representation π_ρ for a pure state ρ is irreducible.

Direct sum of representations Suppose there are two representations π_1 and π_2 of an algebra on two Hilbert spaces, H_1 and H_2, respectively. With the direct sum of the Hilbert spaces

$$\Psi \in H_1 \oplus H_2 : \Psi = \psi_1 \oplus \psi_2; \langle\Psi|\Psi\rangle = \langle\psi_1|\psi_1\rangle + \langle\psi_2|\psi_2\rangle \tag{1.26}$$

the direct sum of the two representations is constructed by

$$\Pi(A)\Psi = \pi_1(A) \oplus \pi_2(A) \tag{1.27}$$

Equivalence of any cyclic representation to GNS Clearly, two representations that are related by a unitary transformation ("change of basis") will be considered equivalent. If the transformation is between two different Hilbert spaces, we must replace "unitary transformation" with "isomorphism", i.e. a linear, bijective, norm-conserving transformation:

$$H_1 \underset{\longrightarrow}{U} H_2 : \|U\Psi_1\|_2 = \|\Psi_1\|_1 \tag{1.28}$$

Two representations related by an isomorphism $\pi_2(A) = U\pi_1(A)U^{-1}$ are called equivalent.

Theorem: Any representation of the C^* algebra on a Hilbert space with a cyclic vector is equivalent to a GNS representation. Sketch of the proof: Assume a specific representation π which has a cyclic vector $|c\rangle$. Then we can define a state on the algebra by

$$f_c(A) = \langle c|\, \pi(A)\, |c\rangle\,. \tag{1.29}$$

The GNS representation π_{f_c} is then equivalent to π. The map U between the two representations

$$|a\rangle = \pi(A)\,|c\rangle \underset{\longrightarrow}{U} |[A]\rangle_{f_c} \tag{1.30}$$

is obviously isometric and invertible as $\langle [A]|[A]\rangle_{f_c} = 0 \Leftrightarrow \langle a|a\rangle = 0$.

Equivalence any representation to a sum of GNS From representation theory: "Any representation" of a C^* algebra (with unity) on a Hilbert space is the direct sum of of representations of with a cyclic vector. Therefore: any representation of a C^* algebra is equivalent to a direct sum of GNS representations.

Let the mathematical dust settle an try to see what we have done. Using only the algebra of observables and one or several states, we have constructed one ore several Hilbert spaces. We can map the algebraic structure onto linear operators on each of these Hilbert spaces. These are the GNS representations. If, in turn, we more or less arbitrarily pick a Hilbert space and represent our algebra on it, this Hilbert space can be put into a one-to-one relation to a sum of the GNS representations. It is equivalent to it. It is all in the C^* algebra and the states. These states we introduced in the closest analogy to probability measures on phase space. The Hilbert space representation pops out automatically.

What is new in quantum mechanics it non-commutativity. For handling this, the Hilbert space representation turned out to be a convenient - by many considered the best - mathematical environment. For classical mechanics, working in the Hilbert space would be an overkill: we just need functions on the phase space.

- Bibliography

- F. Strocchi, Mathematical Structure of Quantum Mechanics
- W. Thirring, Mathematical Physics III: Quantum Mechanic
- J. von Neumann: Mathematische Grundlagen der Quantenmechanik
- M. Reed and B. Simon, Methods of Mathematical Physics, vol. I Functional Analysis, (New York: Academic Press, 1972)
- R. D. Richtmyer, Principles of Advanced Mathematical Physics, vol. I, (Berlin: Springer, 1985)
- L. Debnath and P. Mikusinski, Introduction to Hilbert Spaces with Appli- cations, (Boston: Academic Press, 1990)
- Young, Nicholas (1988), An introduction to Hilbert space, Cambridge University Press, ISBN 0-521-33071-8

1.3 Spectral Theory

Chapter VII

Spectral theory for selfadjoint operators

We denote the kernel or null space of a map a by $N(a)$ and its range or image by $R(a)$. As before, $D(a)$ denotes the domain of a. Also, $a - z$ for $z \in \mathbb{C}$ denotes the operator $a - z1$.

VII.1 Resolvent and spectrum

The theory of the spectrum of a closed operator on a Hilbert space (which may be bounded or unbounded) is a generalization of the theory of eigenvalues of a matrix. From linear algebra we recall:

Proposition VII.1 *Let $a : \mathbb{C}^n \to \mathbb{C}^n$ be a linear map. The a is injective iff it is surjective.*

This follows from the fundamental fact that if $a : V \to W$ is a lnear map between vector spaces, one has $R(a) \cong V/N(a)$. If $V = W = \mathbb{C}^n$, one one count dimensions to infer that $\dim(R(a)) = n - \dim(N(a))$. Surjectivity of a yields $\dim(R(a)) = n$, hence $\dim(N(a)) = 0$, hence $N(a) = 0$, and vice versa.[1] ■

Corollary VII.2 *Let $a : \mathbb{C}^n \to \mathbb{C}^n$ be a linear map. Then $a - z$ is invertible (i.e., injective and surjective) iff z is not an eigenvalue of a, i.e., if there exists no $f \in \mathbb{C}^n$ such that $af = zf$.*

Defining the **spectrum** $\sigma(a)$ of $a : \mathbb{C}^n \to \mathbb{C}^n$ as the set of eigenvalues of a and the **resolvent** $\rho(a)$ as the set of all $z \in \mathbb{C}$ for which $a - z$ is invertible, we therefore have

$$\sigma(a) = \mathbb{C}\backslash\rho(a). \tag{VII.1}$$

If $z \in \rho(a)$, the equation $(a - z)f = g$ for the unknown $f \in \mathbb{C}^n$ has a unique solution for any g; existence follows from the surjectivity of $a - z$, whereas uniqueness follows from its injectivity (if $a - z$ fails to be injective then any element of its kernel can be added to a given solution).

Now, if a is an operator on an infinite-dimensional Hilbert space, it may not have any eigenvalues, even when it is bounded and self-adjoint. For example, if $a(x) = \exp(-x^2)$ the associated multiplication operator $a : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is bounded and self-adjoint, but it has no eigenvalues at all: the equation $af = \lambda f$ for eigenvectors is $\exp(-x^2)f(x) = \lambda f(x)$ for (almost) all $x \in \mathbb{R}$, which holds only if f is nonzero at a single point. But in that case $f = 0$ as an element of L^2. However, the situation is not hopeless. More generally, let any $a \in C_b(\mathbb{R})$, interpreted as a multiplication

[1] In general, this proposition yields the very simplest case of the Atiyah–Singer index theorem, for which these authors received the Abel Prize in 2004. We define the **index** of a linear map $a : V \to W$ as $\text{index}(a) := \dim(\ker(a)) - \dim(\text{coker}(a))$, where $\ker(a) = N(a)$ and $\text{coker}(a) := W/R(a)$, provided both quantities are finite. If V and W are finite-dimensional, Proposition VII.1 yields $\text{index}(a) = \dim(V) - \dim(W)$; in particular, if $V = W$ then $\text{index}(a) = 0$ for any linear map a. In general, the index theorem expresses the index of an operator in terms of topological data; in this simple case the only such data are the dimensions of V and W.

operator $a : L^2(\mathbb{R}) \to L^2(\mathbb{R})$. If $x_0 \in \mathbb{R}$ one may find approximate eigenvectors of a in the following sense: take

$$f_n(x) := (n/\pi)^{1/4} e^{-n(x-x_0)^2/2}. \tag{VII.2}$$

Then $\|f_n\| = 1$ and $\lim_{n\to\infty}(a(x) - a(x_0))f_n = 0$, although the sequence f_n itself has no limit in $L^2(\mathbb{R})$. Thus we may call $\lambda = a(x_0)$ something like a generalized eigenvalue of a for any $x_0 \in \mathbb{R}$, and define the spectrum accordingly: let $a : D(a) \to H$ be a (possibly unbounded) operator on a Hilbert space. We say that $\lambda \in \sigma(a)$ when there exists a sequence (f_n) in $D(a)$ for which $\|f_n\| = 1$ and

$$\lim_{n\to\infty} (a - \lambda) f_n = 0. \tag{VII.3}$$

Of course, when λ is an eigenvalue of a with eigenvector f, we may take $f_n = f$ for all n.

However, this is not the official definition of the spectrum, which is as follows.

Definition VII.3 *Let $a : D(a) \to H$ be a (possibly unbounded) operator on a Hilbert space. The* **resolvent** *$\rho(a)$ is the set of all $z \in \mathbb{C}$ for which $a - z : D(a) \to H$ is injective and surjective (i.e., invertible). The* **spectrum** *$\sigma(a)$ of a is defined by $\sigma(a) := \mathbb{C}\backslash\rho(a)$.*

Hence the *property* (VII.1) has been turned into a definition! We will prove the equivalence of this definition of the spectrum with the definition above later on. In the example just given, one has $\sigma(a) = a(\mathbb{R})$ if the right domain of a is used, namely (III.17). Thus the spectrum can be nonempty even if there aren't any eigenvalues. The subsequent theory shows that these are precisely the right definitons for spectral theory.

The following result explains the role of closedness.[2]

Proposition VII.4 *If an operator $a : D(a) \to R(a) = H$ has an inverse, then a^{-1} is bounded iff a is closed.*

The proof consists of two steps. First, one has that $a : D(a) \to R(a)$ is closed iff a^{-1} is closed. To prove "$\Rightarrow$", assume $g_n \to g$ and $a^{-1}g_n \to f$. Call $f_n := a^{-1}g_n$; then $af_n = g_n \to g$, so if a is closed then by definition $f \in D(a)$ and $af_n \to af$, so $af = g$, hence $f = a^{-1}g$, which means $a^{-1}g_n \to a^{-1}g$. In particular, $g \in R(a) = D(a^{-1})$, and it follows that a^{-1} is closed. The proof of "$\Leftarrow$" is the same, with a and a^{-1} interchanged. Geometrically, the graph of a^{-1} is just the image of the graph of a in $H \oplus H$ under the map $(f, g) \mapsto (g, f)$, hence if one is closed then so is the other.

Secondly, if $R(a) = H$, then $D(a) = H$, hence a^{-1} is bounded by the closed graph theorem (Theorem VI.5). ■

Returning to the equation $(a - z)f = g$, it now follows that when $z \in \rho(a)$, the solution f depends continuously on the initial data g iff a is closed. To avoid pathologies, we therefore assume that a is closed in what follows. Furthermore, as we shall see, practically every argument below breaks down when $(a-z)^{-1}$ is unbounded. This also explains why as far as spectral theory is concerned there isn't much difference between bounded operators and closed unbounded operators: in both cases $(a - z)^{-1}$ is bounded for $z \in \rho(a)$.

As an exercise, one easily shows:

Proposition VII.5 *Let a be a closed operator.*

1. *$\rho(a)$ is open (and hence $\sigma(a)$ is closed) in $\mathbb{C}$.*
2. *$\rho(a^*) = \overline{\rho(a)}$; $\quad \sigma(a^*) = \overline{\sigma(a)}$.*

For unbounded operators the spectrum can (literally) be any subset of $\mathbb{C}$, including the empty set.

[2] Some books define the resolvent of a as the set of those $z \in \mathbb{C}$ for which $(a - z)$ is invertible *and* has bounded inverse. In that case, the resolvent is empty when a is not closed.

VII.2 The spectrum of self-adjoint operators

For a general closed operator a, we may decompose the spectrum as

$$\sigma(a) = \sigma_d(a) \cup \sigma_c(a), \tag{VII.4}$$

where the **discrete spectrum** $\sigma_d(a)$ consists of all eigenvalues of a, and the **continuous spectrum** $\sigma_c(a)$ is the remainder of $\sigma(a)$. Recall that eigenvalues lie in $\sigma(a)$, for if $(a-\lambda)f = 0$ for some nonzero f then $a-\lambda$ cannot be injective. The spectrum of self-adjoint operators has a particularly transparent structure.

Theorem VII.6 *Let a be a self-adjoint operator (i.e., $a^* = a$), and let $z \in \mathbb{C}$. Then one of the following possibilities occurs:*

1. *$R(a-z) = H$ iff $z \in \rho(a)$;*
2. *$R(a-z)^- = H$ but $R(a-z) \neq H$ iff $z \in \sigma_c(a)$;*
3. *$R(a-z)^- \neq H$ iff $z \in \sigma_d(a)$.*

The key to the proof is a very simple result.

Lemma VII.7 *If a is closable (equivalently, if $D(a^*)$ is dense), then $R(a-z)^- = N(a^* - \overline{z})^\perp$ and $N(a^* - \overline{z}) = R(a-z)^\perp$.*

Note that the kernel of a closed operator (in this case $a^* - \overline{z}$) is automatically closed. Easy calculations using the definition of a^* yield the inclusions $R(a-z)^\perp \subset N(a^* - \overline{z})$ and $R(a-z) \subset N(a^* - \overline{z})^\perp$. Since $K^{\perp\perp} = K^-$ for any linear subspace K of a Hilbert space, and $K \subset L$ implies $L^\perp \subset K^\perp$, the claim follows. ■

We first prove Theorem VII.6 for $z \in \mathbb{R}$. If $R(a-z)^- \neq H$, then $N(a-z) = R(a-z)^\perp \neq 0$, so $(a-\lambda)f = 0$ has a nonzero solution and $\lambda \in \sigma_d(a)$. The converse implication has the same proof. If $R(a-z)^- = H$, then $N(a-z) = 0$ and $a-z$ is injective. Now if $R(a-z) = H$ then $a-z$ is surjective as well, and $z \in \rho(a)$. The converse is trivial given the definition of the resolvent. If $R(a-z) \neq H$, then $z \in \sigma_c(a)$ by definition of the continuous spectrum. Conversely, if $z \in \sigma_c(a)$ then $z \notin \sigma_d(a)$ and $z \notin \rho(a)$, so that $R(a-z) = H$ and $R(a-z)^- \neq H$ are exlcuded by the previous 'iff' results for $\rho(a)$ and $\sigma_d(a)$. Hence $R(a-z)^- = H$ but $R(a-z) \neq H$.

To prove Theorem VII.6 for $z \in \mathbb{C}\backslash\mathbb{R}$, we first note that eigenvalues of self-adjoint operators must be real; this is immediate since if $a^* = a$ then $\overline{(f, af)} = (af, f) = (f, af)$, so if f is an eigenvector with eigenvector λ it follows that $\overline{\lambda} = \lambda$. In fact, we will prove that if $z \in \mathbb{C}\backslash\mathbb{R}$, then also $z \in \rho_c(a)$ is impossible, so that $z \in \rho(a)$. To see this we need some lemma's.

Lemma VII.8 *Let a be symmetric. Then $\|(a-z)f\| \geq |\mathrm{Im}\,(z)|\|f\|$.*

Reading Cauchy–Schwarz in the wrong order, we obtain

$$\|(a-z)f\|\|f\| \geq |(f, (a-z)f)| = |(r - i\mathrm{Im}\,(z))\|f\|^2| \geq |\mathrm{Im}\,(z)|\|f\|^2.$$

Here we used the fact that $r := (f, af) - \mathrm{Re}\,(z)$ is a real number by virtue of the symmetry of a. ■

Hence $z \in \mathbb{C}\backslash\mathbb{R}$ implies $N(a - \overline{z}) = 0$. Combining this with Lemma VII.7, we infer that $R(a-z)^- = N(a-\overline{z})^\perp = H$. To infer that actually $R(a-z) = H$ we need yet another lemma.

Lemma VII.9 *Let a be any densely defined operator. If $\|af\| \geq C\|f\|$ for some $C > 0$ and all $f \in D(a)$, then a is injective and $a^{-1} : R(a) \to D(a)$ is bounded with bound $\|a^{-1}\| \leq C^{-1}$.*

Injectivity is trivial, for $af = 0$ cannot have any nonzero solutions given the bound; a linear map a is injective when $af = 0$ implies $f = 0$. For the second claim, note that

$$\|a^{-1}\| = \sup\{\|a^{-1}g\|, g \in D(a^{-1}) = R(a), \|g\| = 1\} =$$

$$\sup\left\{\left\|a^{-1}\frac{af}{\|af\|}\right\|, f \in D(a), f \neq 0\right\} = \sup\left\{\left\|\frac{f}{\|af\|}\right\|, f \in D(a), f \neq 0\right\}.$$

This yields the claim. ■

Combining this with Lemma VII.8, we see that $z \in \mathbb{C}\backslash\mathbb{R}$ implies $(a-z)^{-1} : D((a-z)^{-1}) \to D(a-z) = D(a)$ is bounded, where $D((a-z)^{-1}) = R(a-z)$. To infer that in fact $R(a-z) = H$, we use:

Lemma VII.10 *If b is closed and injective, then $b^{-1} : R(b) \to D(b)$ is closed.*

See the proof of Proposition VII.4. ■

Lemma VII.11 *If b is closed and bounded, then $D(b)$ is closed.*

This is immediate from the definition of closedness. ■

Taking $b = (a-z)^{-1}$, we find that $D((a-z)^{-1})$ is closed. Since we know that $R(a-z)^- = H$, we conclude that $R(a-z) = H$. The same is true for $\overline{z}$. Hence by Lemma VII.7, $N(a-z) = R(a-\overline{z})^\perp = H^\perp = 0$ and $a-z$ is injective. With $a-z$ already known to be surjective, $z \in \rho(a)$.

The proof of the converse implications is the same as for $z \in \mathbb{R}$, and we have finished the proof of Theorem VII.6. ■

Using similar arguments, one can prove

Theorem VII.12 *Let a be a symmetric operator. Then the following properties are equivalent:*

1. *$a^* = a$, i.e., a is self-adjoint;*
2. *a is closed and $N(a^* \pm i) = 0$;*
3. *$R(a \pm i) = H$;*
4. *$R(a-z) = H$ for all $z \in \mathbb{C}\backslash\mathbb{R}$;*
5. *$\sigma(a) \subset \mathbb{R}$.*

Similarly, the following properties are equivalent:

1. *$a^* = a^{**}$, i.e., a is essentially self-adjoint;*
2. *$N(a^* \pm i) = 0$;*
3. *$R(a \pm i)^- = H$;*
4. *$R(a-z)^- = H$ for all $z \in \mathbb{C}\backslash\mathbb{R}$;*
5. *$\sigma(a^-) \subset \mathbb{R}$.*

The second half of the theorem easily follows from the first, on which we will therefore concentrate. The implications $1 \Rightarrow 2$, $1 \Rightarrow 4$, $1 \Rightarrow 5$ and $2 \Rightarrow 3$ are immediate either from Theorem VII.6 or from its proof. The implications $4 \Rightarrow 3$ and $5 \Rightarrow 4$ are trivial. Thus it only remains to prove $3 \Rightarrow 1$.

To do so, assume $R(a \pm i) = H$. For given $f \in D(a^*)$ there must then be a $g \in H$ such that $(a^* - i)f = (a-i)g$. Since a is symmetric, we have $D(a) \subset D(a^*)$, so $f - g \in D(a^*)$, and $(a^* - i)(f-g) = 0$. But $N(a^* - i) = R(a+i)^\perp$ by Lemma VII.7, so $N(a^* - i) = 0$. Hence $f = g$, and in particular $f \in D(a)$ and hence $D(a^*) \subset D(a)$. Since we already know the opposite inclusion, we have $D(a^*) = D(a)$. Given symmetry, this implies $a^* = a$. ■

Corollary VII.13 *Let $a^* = a$. The $\sigma(a) \subset \mathbb{R}$. In other words, the spectrum of a self-adjoint operator is real.*

As an illustration of Theorem VII.12, one can directly show:

Proposition VII.14 *Let $a \in C(\Omega)$ define a real-valued multiplication operator on*

$$D(a) = \{f \in L^2(\Omega) \mid af \in L^2(\Omega)\} \subset H = L^2(\Omega),$$

so that $a^* = a$ (cf. Proposition VI.8.) Then the operator a is injective iff $a(x) \neq 0$ for all $x \in \Omega$, and surjective iff there exists $\varepsilon > 0$ so that $|a(x)| \geq \varepsilon$ for all $x \in \Omega$; in that case a is injective and has bounded inverse. Consequently, $\sigma(a) = a(\Omega)^-$, with $a(\Omega) := \{a(x), x \in \Omega\}$.

Finally, we justify our earlier heuristic definition of the spectrum; the thrust of the theorem lies in its characterization of the continuous spectrum, of course.

Theorem VII.15 *Let a be self-adjoint. Then $\lambda \in \sigma(a)$ iff there exists a sequence (f_n) in $D(a)$ with $\|f_n\| = 1$ for all n such that $\lim_n (a - \lambda)f_n = 0$.*

Suppose $\lambda \in \sigma(a)$. If $\lambda \in \sigma_d(a)$ we are ready, taking $f_n = f$ for all n. If $\lambda \in \sigma_c(a)$, then $R(a-\lambda)^- = H$ but $R(a-\lambda) \neq H$ by Theorem VII.6. Now a is self-adjoint, hence a and $a - \lambda$ are closed, so that also $(a-\lambda)^{-1}$ is closed by Lemma VII.10. Hence $(a-\lambda)^{-1} : R(a-\lambda) \to H$ must be a a densely defined unbounded operator by Lemma VII.11, for if it were bounded then its domain would be closed, which $D((a-\lambda)^{-1}) = R(a-\lambda)$ is not, as we have just shown. Thus there is a sequence g_n in $D((a-\lambda)^{-1})$ with norm 1 and $\|(a-\lambda)^{-1}g_n\| \to \infty$. Then $f_n := (a-\lambda)^{-1}g_n/\|(a-\lambda)^{-1}g_n\|$ has the desired property.

Conversely, if $\lambda \in \rho(a)$ then $(a-\lambda)^{-1}$ is bounded, hence $(a-\lambda)f_n \to 0$ implies $f_n \to 0$, so the sequence (f_n) cannot exist, and $\lambda \in \sigma(a)$ by *reductio ad absurdum.* ■

VII.3 Application to quantum mechanics

The theory of self-adjoint operators has many applications, for example to the theory of boundary value problems for linear partial differential equations. In these notes we focus on applications to quantum mechanics.

In Chapter V we initially assumed that observables in quantum mechanics are mathematically represented by bounded self-adjoint operators, i.e. linear maps $a : B(H) \to B(H)$ such that $\|a\| < \infty$ and $a^* = a$. As already mentioned at the end of that chapter, however, this model is too limited. For example, in physics textbooks you will find the position and momentum operators

$$\begin{aligned} \hat{q}^i &= x^i \\ \hat{p}_i &= -i\hbar\frac{\partial}{\partial x^i}. \end{aligned} \tag{VII.5}$$

Here $\hbar \in \mathbb{R}^+$ is a constant of nature, called **Planck's constant**, and $i = 1, 2, 3$. These operators are allegedly defined on $H = L^2(\mathbb{R}^3)$, but we know from the previous chapter that at least $\hat{q}^i$ is unbounded. It is a multiplication operator of the form $a\psi(x) = \hat{a}\psi(x)$ with $\hat{a} \in C(\mathbb{R}^3)$, in this case $\hat{a}(x) = x^i$. As we have seen, a is bounded iff $\|\hat{a}\|_\infty < \infty$, and this clearly not the case for x^i. Hence the position operator is unbounded. It follows from Proposition VI.8 that $\hat{q}^i$ is self-adjoint on the domain

$$D(\hat{q}^i) = \{\psi \in L^2(\mathbb{R}^3) \mid x^i\psi \in L^2(\mathbb{R}^3)\}, \tag{VII.6}$$

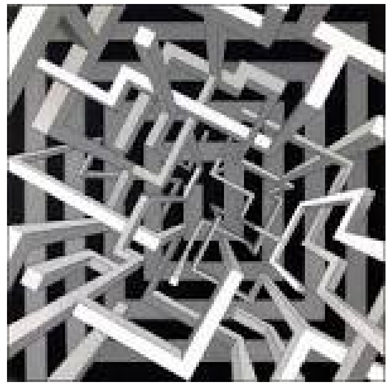
where $x^i\psi$ is shorthand for the function $x \mapsto x^i\psi(x)$.

Although we have not had the opportunity to develop the necessary machinery, the story of the momentum operator $\hat{p}_i$ is similar. If we denote the Fourier transform of $\psi \in L^2(\mathbb{R}^3)$ by $\hat{\psi}$, and call its argument $k = (k_1, k_2, k_3) \in \mathbb{R}^3$, we can write

$$\hat{\psi}(k) = \int_{\mathbb{R}^3} d^3x\, \psi(x)e^{-ikx}, \tag{VII.7}$$

where $kx = x^1k_1 + x^2k_2 + x^3k_3$. This inproper integral is defined as follows.

1.4 More on Hilbert Spaces



Basic Facts About Hilbert Space

The term Euclidean space refers to a finite dimensional linear space with an inner product. A Euclidean space is always complete by virtue of the fact that it is finite dimensional (and we are taking the scalars here to be the reals which have been constructed to be complete). An infinite dimensional inner product space which is complete for the norm induced by the inner product is called a **Hilbert space**. A Hilbert space is in many ways like a Euclidean space (which is why finite dimensional intuituition often works in the infinite dimensional Hilbert space setting) but there are some ways in which the infinite dimensionality leads to subtle differences we need to be aware of.

Subspaces

A subset M of Hilbert space H is a **subspace** of it is closed under the operation of forming linear combinations; i.e., for all x and y in M, $C_1x + C_2y$ belongs to M for all scalars C_1, C_2. The subspace M is said to be **closed** if it contains all its limit points; i.e., every sequence of elements of M that is Cauchy for the H-norm, converges to an element of M. In a Euclidean space every subspace is closed but in a Hilbert space this is not the case.

Examples-

(a) If U is a bounded open set in R^n then H = H^0(U) is a Hilbert space containing M = C(U) as a subspace. It is easy to find a sequence of functions in M that is Cauchy for the H norm but the sequence converges to a function in H that is discontinuous and hence not in M. This proves that M is not closed in H.

(b) Every finite dimensional subspace of a Hilbert space H is closed. For example, if M denotes the span of finitely many elements x_1, ...x_N in H, then the set M of all possible linear combinations of these elements is finite dimensional (of dimension N), hence it is closed in H.

(c) Let M denote a subspace of Hilbert space H and let $M^\perp$ denote the **orthogonal complement of M**.

$$M^\perp = \{x \in H : (x,y)_H = 0, \forall y \in M\}$$

Then $M^\perp$ is easily seen to be a subspace and it is closed, whether or not M itself is closed. To see this, suppose $\{x_n\}$ is a Cauchy sequence in $M^\perp$ converging to limit x in H. For arbitrary y in M,

$$(x,y)_H = (x - x_n, y)_H + (x_n, y)_H = (x - x_n, y)_H + 0 \to 0, \text{ as n tends to infinity.}$$

Then the limit point x is orthogonal to every y in M which is to say, x is in $M^\perp$, and $M^\perp$ is closed.

Lemma 1- Let M denote a subspace of Hilbert space H. Then $(M^\perp)^\perp = \bar{M}$.

If M is a subspace of H that is not closed, then M is contained in a closed subspace $\bar{M}$ of H, consisting of M together with all its limit points. $\bar{M}$ is called the **closure** of M and M is said to be **dense** in $\bar{M}$. This means that for every x in $\bar{M}$ there is a sequence of elements of M that converge to x in the norm of H. Equivalently, to say M is dense in $\bar{M}$ means that for every x in $\bar{M}$ and every $\epsilon > 0$, there is a y in M such that $\|x - y\|_H < \epsilon.$

Lemma 2 A subspace M of Hilbert space H is dense in H if and only if $M^\perp = \{0\}$.

A Hilbert space H is said to be **separable** if H contains a countable dense subset $\{h_n\}$. In this case, for every x in H and every $\epsilon > 0$ there exists an integer N_ϵ and scalars $\{a_n\}$ such that

$$\left\| x - \sum_{n=1}^{N} a_n h_n \right\|_H < \epsilon \quad \text{for} \quad N > N_\epsilon$$

If H is a separable Hilbert space, then the Gram-Schmidt procedure can be used to construct an orthonormal basis for H out of a countable dense subset. An orthonormal basis for H is a set of mutually orthogonal unit vectors, $\{\phi_n\}$ in H with the following property:

1) For $f \in H,\ (\phi_n, f)_H = 0$ for every n if and only if $f = 0$

When the orthonormal set $\{\phi_n\}$ has property 1, then it is said to be **dense** or **complete** in H. Of course, not every orthonormal set in H is complete. Other equivalent ways of characterizing completeness for orthonormal sets can be stated as follows:

2) For all f in H and every $\epsilon > 0$, there exists an integer N_ϵ such that

$$\left\| f - \sum_{n=1}^{N} (f, \phi_n)_H \phi_n \right\|_H < \epsilon \quad \text{for} \quad N > N_\epsilon$$

3) For every f in H, $\sum_{n=1}^{\infty} f_n^2 = \|f\|_H^2$ where $f_n = (f, \phi_n)_H$

In a Euclidean space, E, where all subspaces M are closed, it is a fact that for each y in E there is a unique z in M such that $\|y - z\|$ is minimal. This element z, which is just the orthogonal projection of y onto M, is the "best approximation to y from within M". In an infinite dimensional Hilbert space, a similar result is true for closed subspaces but for subspaces that are not closed there may fail to be a "best" approximation in M.

Hilbert Space Projection Theorem Let M be a closed subspace of Hilbert space H and let y in H be given. Then

(i) there exists a unique x_y in M such that $\|y - x_y\|_H \le \|y - z\|_H$ for all z in M
(x_y is the unique point of M that is closest to y, the best approximation in M to y)

(ii) $(y - x_y, z)_H = 0$ for all z in M; i.e., $y - x_y \perp M$

(iii) every y in H can be uniquely expressed as $y = x_y + z_y$ where

$$Py = x_y \in M, \;\; Qy = z_y \in M^{\perp}$$

and $\|y\|_H^2 = \|Py\|_H^2 + \|Qy\|_H^2$ i.e., $H = M \oplus M^{\perp}$.

The proof of this result will be given later.

Linear Functionals and Bilinear Forms

A real valued function defined on H, is said to be a ***functional*** on H. The functional, L, is said to be:

(a) **Linear** if, for all x and y in H, $L(C_1x + C_2y) = C_1Lx + C_2Ly,$ for all scalars C_1, C_2.

(b) **Bounded** if there exists a constant C such that $|Lx| \leq C\|x\|_H$ for all x in H

(c) **Continuous** if $\|x_n - x\|_H \to 0$ implies that $|Lx_n - Lx| \to 0$

It is not difficult to show that the only example of a linear functional on a Euclidean space E is $Lx = (x,z)_E$ for some z in E, fixed. For example, if F is a linear functional on E, then for arbitrary x in E,

$$F(x) = F\left(\sum_{i=1}^{n} x_i \;\; _i\right) = \sum_{i=1}^{n} x_i F(\;\; _i) = \sum_{i=1}^{n} x_i F_i = (x, z_F)_E = x^{\top} z_F$$

where $\{e_i\}$ denotes the standard basis in E and $\bar{z}_F$ denotes the n-tuple whose i-th component is $F_i = F(\;\; _i)$. This displays the isomorphism between functionals F and elements, z_F, in E. This isomorphism also exists in an abstract Hilbert space.

Riesz Representation Theorem For every continuous linear functional f on Hilbert space H there exists a unique element z_f in H such that $f(x) = (x, z_f)_H$ for all x in H.

Proof- Let $N_f = \{x \in H : f(x) = 0\}$.Then N_f is easily seen to be a closed subspace of H. If $N_f = H$ then $z_f = 0$ and we are done. If $N_f \neq H$ then $H = N_f \oplus N_f^{\perp}$ by the Hilbert space projection theorem. Since N_f is not all of H, $N_f^{\perp}$ must contain nonzero vectors, and we denote by z_0 an element of $N_f^{\perp}$ such that $\|z_0\|_H = 1$. Then for any x in H,

$$w = f(x)z_0 - f(z_0)x$$

belongs to N_f hence $w \perp z_0$. But in that case,

$$(f(x)z_0 - f(z_0)x, z_0)_H = f(x)(z_0, z_0)_H - f(z_0)(x, z_0)_H = 0.$$

This leads to, $f(x) = f(z_0)(x, z_0)_H = (x, f(z_0)z_0)_H$ which is to say $z_f = f(z_0)z_0$.

To see that z_f is unique, suppose that $f(x) = (z_f, x)_H = (w_f, x)_H$ for all x in H. Subtracting leads to the result that $(z_f - w_f, x)_H = 0$ for all x in H. In particular, choosing $x = z_f - w_f$ leads to $\|(z_f - w_f\|_H = 0$. ■

A real valued function $a(x,y)$ defined on $H \times H$ is said to be:

(a) **Bilinear** if, for all $x_1, x_2, y_1, y_2 \in H$ and all scalars C_1, C_2

$$a(C_1x_1 + C_2x_2, y_1) = C_1a(x_1, y_1) + C_2a(x_2, y_1)$$

$$a(x_1, C_1y_1 + C_2y_2) = C_1a(x_1, y_1) + C_2a(x_2, y_1)$$

(b) **Bounded** if there exists a constant $b > 0$ such that,

$|a(x,y)| \leq b\|x\|_H\|y\|_H$ for all x,y in H

(c) **Continuous** if $x_n \to x,$ and $y_n \to y$ in H, implies $a(x_n, y_n) \to a(x,y)$ in R

(d) **Symmetric** if $a(x,y) = a(y,x)$ for all x,y in H

(e) **Positive** or **coercive** if there exists a $C > 0$ such that

$a(x,x) \geq C\|x\|_H^2$ for all x in H

It is not hard to show that for both linear functionals and bilinear forms, boundedness is equivalent to continuity. If a(x,y) is a bilinear form on $H \times H$, and F(x) is a linear functional on H, then $\Phi(x) = a(x,x)/2 - F(x) + Const$ is called a **quadratic functional on H**. In a Euclidean space a quadratic functional has a unique extreme point located at the point where the gradient of the functional vanishes. This result generalizes to the infinite dimensional situation.

Lemma 3 Suppose $a(x,y)$ is a positive, bounded and symmetric bilinear form on Hilbert space H, and F(x) is a bounded linear functional on H. Consider the following problems

(a) minimize $\Phi(x) = a(x,x)/2 - F(x) + Const$ over H

(b) find x in H satisfying $a(x,y) = F(y)$ for all y in H.

Then

i) x in H solves (a) if and only if x solves (b)

ii) there is at most on x in H solving (a) and (b)

iii) there is at least one x in H solving (a) and (b)

Proof- For t in R and x,y fixed in H, let $f(t) = \Phi(x+ty)$. Then f(t) is a real valued function of the real variable t and it follows from the symmetry of $a(x,y)$ that

$$f(t) = t^2/2\ a(y,y) + t[a(x,y) - F(y)] + 1/2\ a(x,x) - F(x) + Const$$

and

$$f'(t) = t\ a(y,y) + [a(x,y) - F(y)]$$

It follows that $\Phi(x)$ has a global minimum at x in H if and only if $f(t)$ has a global minimum at $t = 0$; i.e.,

$$\Phi(x+ty) = \Phi(x) + tf'(0) + t^2/2\ a(x,x) \geq \Phi(x), \quad \forall t \in R \quad and \quad \forall y \in H$$

if and only if

$$f'(0) = a(x,y) - F(y) = 0. \quad \forall y \in H.$$

This establishes the equivalence of (a) and (b).

To show that $\Phi(x)$ has at most one minimum in H, suppose

$a(x_1,y) = F(y)$ and $a(x_2,y) = F(y)$ for all y in H.

Then $a(x_1,y) - a(x_2,y) = a(x_1 - x_2, y) = 0$ for all y in H. In particular, for $y = x_1 - x_2$

$0 = a(x_1 - x_2, x_1 - x_2) \geq C\|x_1 - x_2\|_H^2$; i.e., $x_1 = x_2$

To show that $\Phi(x)$ has at least one minimum in H, let $\alpha = \inf_{x \in H}\Phi(x)$. Now

$$\Phi(x) = 1/2\ a(x,x) - F(x) \geq 1/2\ C\|x\|_H^2 - b\|x\|_H$$

and it is evident that $\Phi(x)$ tends to infinity as $\|x\|_H$ tends to infinity. This means $\alpha > -\infty$ (i.e.," the parabola opens upward rather than downward"). Moreover since α is an infimum, there exists a sequence x_n in H such that $\Phi(x_n) \to \alpha$ as n tends to infinity. Note that

$$2[a(x_n,x_n) + a(x_m,x_m)] = a(x_n - x_m, x_n - x_m) + a(x_m + x_n, x_m + x_n)$$

which leads to the result,

$$\Phi(x_m) + \Phi(x_n) = 1/4\ a(x_m - x_n, x_m - x_n) + 2\ \Phi[(x_m + x_n)/2] \geq 1/4C\|x_m - x_n\|_H^2 + 2\alpha.$$

But $\Phi(x_m) + \Phi(x_n)$ tends to 2α as n tends to infinity and in view of the previous line, the minimizing sequence $\{x_n\}$ must be a Cauchy sequence with limit x in the Hilbert space H. Finally, since $\Phi(x)$ is continuous, $\Phi(x_n) \to \Phi(x) = \alpha$.■

Applications of the lemma-

(i) This lemma can now be used to prove the Hilbert space projection theorem.

For M a closed subspace in H and for y a fixed but arbitrary element in H, note that

$$\|x-y\|_H^2 = (x-y,x-y)_H = \|x\|_H^2 - 2(x,y)_H + \|y\|_H^2 \quad \text{for all x in M.}$$

Since M is closed in H, it follows that M is itself a Hilbert space for the norm and inner product inherited from H.

Define

$$a(z,x) = (z,x)_H \qquad \text{for x and z in M,}$$
$$F(z) = (y,z)_H \qquad \text{for z in M,}$$
and
$$\Phi(z) = 1/2\ a(z,z) - F(z) + 1/2\|y\|_H^2 \qquad \text{for z in M}$$

Clearly $a(z,x)$ is a positive, bounded and symmetric bilinear form on M, F is a bounded linear functional on M. Then it follows from the lemma that there exists a unique element $x_y \in M$ which minimizes $\Phi(z)$ over M. It follows also form the equivalence of problems (a) and (b) that x_y satisfies $a(x_y,z) = F(z)$, for all z in M; i.e., $(x_y,z)_H = (y,z)_H$ for all z in M. But this is just the assertion that $(x_y - y,z)_H = 0$ for all z in M, that is, $x_y - y \perp M$. Finally, for y in H, fixed, let the unique element x_y in M be denoted by $Py = x_y \in M$. Then

$$y - Py \perp M, \text{ and } z = y - Py \in M^\perp.$$

To see that this decomposition of elements of H is unique, suppose

$$y = x_y + z, \quad x_y \in M, \quad z \in M^\perp,$$
and
$$y = X_y + Z, \quad X_y \in M, \quad Z \in M^\perp,$$
Then
$$x_y + z = X_y + Z, \text{ and } x_y - X_y = Z - z.$$

But $\quad x_y - X_y \in M, \quad Z - z \in M^\perp, \quad M \cap M^\perp = \{0\},$

and it follows that $\quad x_y - X_y = Z - z = 0$.■

(ii) Recall that for U open and bounded in R^n, $H_0^1(U) = M$, is a closed subspace of $H^1(U) = H$. Then by the projection theorem, every y in H can be uniquely expressed as a sum, $y = x_y + z$, with $x_y \in M$, and $z \in M^\perp$. To characterize the subspace $M^\perp$, choose arbitrary $\phi \in C_0^\infty(U)$ and $\psi \in C^\infty(U)$ and write

$$(\phi,\psi)_H = \int_U [\phi\psi + \nabla\phi \bullet \nabla\psi]dx = \int_U \phi[\psi - \nabla^2\psi]dx + \int_{\partial U} \phi \partial_N \psi dS$$

$$= (\phi, \psi - \nabla^2\psi)_0 + 0. \text{ (Recall that } (u,v)_0 \text{ denotes the H}^0\text{(U) inner product).}$$

Now suppose $\psi \in C^\infty(U) \cap M^\perp$. Then $(\phi,\psi)_H = 0,$ for all $\phi \in C_0^\infty(U),$ and since $C_0^\infty(U)$ is dense in M, $(u,\psi)_H = 0,$ for all u in M. That is, $(u,\psi - \nabla^2\psi)_0 = 0$ for all u in M. But this implies that $\psi \in C^\infty(U) \cap M^\perp$ satisfies $\psi - \nabla^2\psi = 0,$ in $H^0(U)$. Then, since $C^\infty(U)$ is dense in $H = H^1(U)$ (cf. Theorem 2 pg 250 in the text) it follows that

$$M^\perp = \{z \in H : z - \nabla^2 z \in H^0(U), \text{ and } z - \nabla^2 z = 0\ \}$$

The lemma requires that the bilinear form $a(x,y)$ be symmetric. For application to existence theorems for partial differential equations, this is an unacceptable restriction. Fortunately, the most important part of the result remains true even when the form is not symmetric.

For A an n by n, not necessarily symmetric, but positive definite matrix, consider the problem Ax = f. For any n by n matrix $dimN_A = dimN_{A^\top},$ and for A positive definite, $dimN_A = 0,$ which is to say that the solution of $Ax = f$ is unique if it exists. Since $R^n = R_A \oplus N_{A^\top},$ it follows that $R^n = R,$ which is to say, a solution for $Ax = f$ exists for every f in R^n. The situation in an abstract Hilbert space H is very close to this.

Lax-Milgram Lemma- Suppose a(u,v) is a positive and bounded bilinear form on Hilbert space H; i.e.,

$$|a(u,v)| \leq \alpha \|u\|_H \|v\|_H \quad \forall u,v \in H$$

and

$$a(u,u) \geq \beta \|u\|_H^2 \quad \forall u \in H.$$

Suppose also that $F(v)$ is a bounded linear functional on H. Then there exists a unique U in H such that

$$a(U,v) = F(v) \quad \forall v \in H.$$

Proof- For each fixed u in H, the mapping $v \rightsquigarrow a(u,v)$ is a bounded linear functional on H. It follows that there exists a unique $z_u \in H$ such that

$$a(u,v) = (z_u,v)_H \quad \forall v \in H.$$

Let $Au = z_u$; i.e., $a(u,v) = (Au,v)_H \ \forall u \in H.$ Clearly A is a linear mapping of H into H, and since

$$\|Au\|_H^2 = |(Au,Au)_H| = |a(u,Au)| \leq \alpha \|u\|_H \|Au\|_H$$

it is evident that A is also bounded. Note further, that

$$\beta \|u\|_H^2 \leq a(u,u) = (Au,u)_H \leq \|Au\|_H \|u\|_H$$

i.e., $\beta \|u\|_H \leq \|Au\|_H \quad \forall u \in H.$

This estimate implies that A is one-to one and that R_A, the range of A, is closed in H. Finally, we will show that $R_A = H$. Since the range is closed, we can use the projection theorem to write, $H = R_A \oplus R_A^\perp$. If $u \in R_A^\perp$, then

$$0 = (Au,u)_H = a(u,u) \geq \beta\|u\|_H^2; \quad \text{i.e.,} \ R_A^\perp = \{0\}.$$

Since $F(v)$ is a bounded linear functional on H, it follows from the Riesz theorem that there is a unique $z_F \in H$ such that $F(v) = (z_F,v)_H$ for all v in H. Then the equation $a(u,v) = F(v)$ can be expressed as

$$(Au,v)_H = (z_F,v)_H \ \ \forall v \in H; \ \text{i.e.,} \ Au = z_F.$$

But A has been seen to be one-to-one and onto and it follows that there exists a unique $U \in H$ such that $AU = z_F$.■

Convergence

In R^N convergence of x_n to x means

$$\|x_n - x\|_{R^N} = \left[\sum_{i=1}^N [(x_n - x) \bullet e_i]^2\right]^{1/2} \to 0 \ \text{as} \ n \to \infty.$$

Here e_i denotes the i-th vector in the standard basis.This is equivalent to,

$$(x_n - x) \bullet e_i \to 0 \ \text{as} \ n \to \infty, \text{ for } i = 1,...,N,$$

and to $$(x_n - x) \bullet z \to 0 \ \text{as} \ n \to \infty, \text{ for every } z \in R^N$$

In an infinite dimensional Hilbert space H, convergence of x_n to x means $\|x_n - x\|_H \to 0$ as $n \to \infty$. This is called **strong convergence** in H and it implies that

$$(x_n - x, v)_H \to 0 \ \text{as} \ n \to \infty \ \ \forall v \in H.$$

This last mode of convergence is referred to as **weak convergence** and, in a general Hilbert space, weak convergence does not imply strong convergence. Thus while there is no distinction between weak and strong convergence in a finite dimensional space, the two notions of convergence are not the same in a space of infinite dimensions.

In R^N, a sequence $\{x_n\}$ is said to be bounded if there is a constant M such that $|x_n| \leq M$ for all n. Then the Bolzano-Weierstrass theorem asserts that every bounded sequence $\{x_n\}$ contains a convergent subsequence. To see why this is true, note that $\{x_n \bullet e_1\}$ is a bounded sequence of real numbers and hence contains a subsequence $\{x_{n,1} \bullet e_1\}$ that is convergent. Similarly, $\{x_{n,1} \bullet e_2\}$ is also a bounded sequence of real numbers and thus contains a subsequence $\{x_{n,2} \bullet e_2\}$ that is convergent. Proceeding in this way, we can generate a sequence of subsequences, $\{x_{n,k}\} \subset \{x_n\}$ such that $\{x_{n,k} \bullet e_j\}$ is convergent for $j \leq k$. Then the diagonal sequence $\{x_{n,n}\}$ is such that $\{x_{n,n} \bullet ej\}$ is convergent for

$1 \le j \le N,$ which is to say, $\{x_{n,n}\}$ is convergent.

The analogue of this result in a Hilbert space is the following.

Lemma 4- Every bounded sequence in a separable Hilbert space contains a subsequence that is weakly convergent.

Proof- Suppose that $\|x_n\| \le M$ for all n and let $\{\phi_j\}$ denote a complete orthonormal family in H. Proceeding as we did in R^N, let $\{x_{n,k}\} \subset \{x_n\}$ denote a subsequence such that $\{(x_{n,k},\phi_j)_H\}$ is convergent (in R) for $j \le k$. Then for each j, $(x_{n,j},\phi_j)_H$ converges to a real limit a_j as n tends to infinity. It follows that the diagonal subsequence $\{x_{n,n}\}$ is such that $(x_{n,n},\phi_j)_H$ converges to a_j for $j \ge 1$. Now define

$$F(v) = Lim_n\ (x_{n,n},v)_H \quad \text{for v in H.}$$

Then $\quad |F(v)| \le |\lim_n (x_{n,n},v)_H| \le M\|v\|_H$

from which it follows that F is a continuous linear functional on H. By the Riesz theorem, there exists an element, z_F in H such that $F(v) = (z_F,v)_H$ for all v in H.

But $$F(v) = F\left(\sum_i (v,\phi_i)_H\phi_i\right) = \lim_n (x_{n,n},\sum_i (v,\phi_i)_H\phi_i)_H$$

$$= \sum_i \lim_n (x_{n,n},\phi_i)_H (v,\phi_i)_H = \sum_i a_i(v,\phi_i)_H \ ;$$

That is, $\quad F(v) = (z_F,v)_H = \sum_i a_i(v,\phi_i)_H \quad$ for all v in H. Then by the Parseval-Plancherel identity, it follows that

$$z_F = \sum_i a_i\phi_i$$

and

$$(x_{n,n},v)_H \to (z_F,v)_H \quad \text{for all v in H.} \blacksquare$$

We should point out that a sequence $\{x_n\}$ is H is said to be **strongly bounded** if there is an M such that $\|x_n\|_H \le M$ for all n, and it is said to be **weakly bounded** if $|(x_n,v)_H| \le M$ for all n and all v in H. These two notions of boundedness coincide in a Hilbert space so it is sufficient to use the term bounded for either case.

Lemma 5- A sequence in a Hilbert space is weakly bounded if and only if it is strongly bounded.

1.5 Examples

Definition 37 *Two Hilbert spaces H_1 and H_2 are said to be* **isomorphic** *if there is a unitary linear transformation U from H_1 onto H_2.*

Definition 38 *Let H_1 and H_2 be Hilbert spaces. The* **direct sum** *$H_1 \oplus H_2$ of Hilbert spaces H_1 and H_2 is the set of ordered pairs $z = (x, y)$ with $x \in H_1$ and $y \in H_2$ with inner product*

$$(z_1, z_2)_{H_1 \oplus H_2} = (x_1, x_2)_{H_1} + (y_1, y_2)_{H_2} \tag{5.62}$$

6 Examples of Hilbert Spaces

1. Finite Dimensional Vectors. $\mathbb{C}^N$ is the space of N-tuples $x = (x_1, \dots, x_N)$ of complex numbers. It is a Hilbert space with the inner product

$$(x, y) = \sum_{n=1}^{N} x_n^* y_n \,. \tag{6.63}$$

2. Square Summable Sequences of Complex Numbers. l^2 is the space of sequences of complex numbers $x = \{x_n\}_{n=1}^{\infty}$ such that

$$\sum_{n=1}^{\infty} |x_n|^2 < \infty \,. \tag{6.64}$$

It is a Hilbert space with the inner product

$$(x, y) = \sum_{n=1}^{\infty} x_n^* y_n \,. \tag{6.65}$$

3. Square Integrable Functions on $\mathbb{R}$. $L^2(\mathbb{R})$ is the space of complex valued functions such that

$$\int_{\mathbb{R}} |f(x)|^2 \, dx < \infty \,. \tag{6.66}$$

It is a Hilbert space with the inner product

$$(f, g) = \int_{\mathbb{R}} f^*(x) g(x) \, dx \tag{6.67}$$

4. Square Integrable Functions on $\mathbb{R}^n$. Let Ω be an open set in $\mathbb{R}^n$ (in particular, Ω can be the whole $\mathbb{R}^n$). The space $L^2(\Omega)$ is the set of complex valued functions such that

$$\int_{\Omega} |f(x)|^2 dx < \infty \,, \tag{6.68}$$

where $x = (x_1, \dots, x_n) \in \Omega$ and $dx = dx_1 \cdots dx_n$. It is a Hilbert space with the inner product

$$(f, g) = \int_{\Omega} f^*(x) g(x) \, dx \tag{6.69}$$

5. Square Integrable Vector Valued Functions. Let Ω be an open set in $\mathbb{R}^n$ (in particular, Ω can be the whole $\mathbb{R}^n$) and V be a finite-dimensional vector space. The space $L^2(V,\Omega)$ is the set of vector valued functions $f = (f_1, \dots, f_N)$ on Ω such that

$$\sum_{i=1}^{N} \int_\Omega |f_i(x)|^2 dx < \infty \,. \tag{6.70}$$

It is a Hilbert space with the inner product

$$(f,g) = \sum_{i=1}^{N} \int_\Omega f_i^*(x) g_i(x)\, dx \tag{6.71}$$

6. Sobolev Spaces. Let Ω be an open set in $\mathbb{R}^n$ (in particular, Ω can be the whole $\mathbb{R}^n$) and V a finite-dimensional complex vector space. Let $C^m(V,\Omega)$ be the space of complex vector valued functions that have partial derivatives of all orders less or equal to m. Let $\alpha = (\alpha_1, \dots, \alpha_n)$, $\alpha \in \mathbb{N}$, be a multiindex of nonnegative integers, $\alpha_i \geq 0$, and let $|\alpha| = \alpha_1 + \cdots + \alpha_n$. Define

$$D^\alpha f = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} f \,. \tag{6.72}$$

Then $f \in C^m(V,\Omega)$ iff

$$|D^\alpha f_i(x)| < \infty \qquad \forall \alpha,\ |\alpha| \leq m,\ \forall i = 1, \dots, N,\ \forall x \in \Omega \,. \tag{6.73}$$

The space $H^m(V,\Omega)$ is the space of complex vector valued functions such that $D^\alpha f \in L^2(V,\Omega)$ $\forall \alpha,\ |\alpha| \leq m$, i.e. such that

$$\sum_{i=1}^{N} \int_\Omega |D^\alpha f_i(x)|^2 dx < \infty \qquad \forall \alpha,\ |\alpha| \leq m \,. \tag{6.74}$$

It is a Hilbert space with the inner product

$$(f,g) = \sum_{\alpha,\, |\alpha| \leq m} \sum_{i=1}^{N} \int_\Omega (D^\alpha f_i(x))^* D^\alpha g_i(x)\, dx \tag{6.75}$$

Remark. More precisely, the Sobolev space $H^m(V,\Omega)$ is the completion of the space defined above.

7 Projection Theorem

Definition 39 *Let M be a closed subspace of a Hilbert space H. The set, $M^\perp$, of vectors in H which are orthogonal to M is called the* **othogonal complement** *of M.*

1.6 Basic Definitions

Definition *Isomorphism* Formally, an isomorphism is bijective morphism. Informally, an isomorphism is a map that preserves sets and relations among elements. “A is isomorphic to B” is written $A \cong B$. Unfortunately, this symbol is also used to denote geometric congruence. An isomorphism from a set of elements onto itself is called an automorphism.

Definition *Homeomorphism* A homeomorphism, also called a continuous transformation, is an equivalence relation and one-to-one correspondence between points in two geometric figures or topological spaces that is continuous in both directions. A homeomorphism which also preserves distances is called an isometry. Affine transformations are another type of common geometric homeomorphism. The similarity in meaning and form of the words "homomorphism" and "homeomorphism" is unfortunate and a common source of confusion.

Definition *Cardinality* In formal set theory, a cardinal number (also called “the cardinality”) is a type of number defined in such a way that any method of counting sets using it gives the same result. (This is not true for the ordinal numbers.) In fact, the cardinal numbers are obtained by collecting all ordinal numbers which are obtainable by counting a given set. The cardinality of a set is also frequently referred to as the "power" of a set.

Definition *Image* If $f : D \to Y$ is a map (a.k.a. function, transformation, etc.) over a domain D, then the image of f, also called the range of D under f, is defined as the set of all values that f can take as its argument varies over D, i.e.,

$$\mathrm{Range}(f) = f(D) = \{f(X) : X \in D\}. \tag{1.31}$$

Definition *Surjection* Let f be a function defined on a set A and taking values in a set B. Then f is said to be a surjection (or surjective map) if, for any $b \in B$, there exists an $a \in A$ for which $b = f(a)$. A surjection is sometimes referred to as being "onto."

Let the function be an operator which maps points in the domain to every point in the range and let V be a vector space with $A, B \in V$. Then a transformation T defined on V is a surjection if there is an $A \in V$ such that $T(A) = B \forall B$.

Definition *Injection* Let f be a function defined on a set A and taking values in a set B. Then f is said to be an injection (or injective map, or embedding) if, whenever $f(x) = f(y)$, it must be the case that $x = y$. Equivalently, $x \not= y$ implies $f(x) \neq f(y)$. In other words, f is an injection if it maps distinct objects to distinct objects. An injection is sometimes also called one-to-one.

A linear transformation is injective if the kernel of the function is zero, i.e., a function $f(x)$ is injective iff $\mathrm{Ker}(f) = 0$. A function which is both an injection and a surjection is said to be a bijection.

Definition *Topology* Topology is the mathematical study of the properties that are preserved through deformations, twistings, and stretchings of objects. Tearing, however, is not allowed. A circle is topologically equivalent to an ellipse (into which it can be deformed by stretching) and a sphere is equivalent to an ellipsoid.

There is also a formal definition for a topology defined in terms of set operations. A set X along with a collection T of subsets of it is said to be a topology if the subsets in T obey the following properties:

1. The (trivial) subsets X and the empty set $\emptyset$ are in T.

2. Whenever sets A and B are in T, then so is $A \cap B$.

3. Whenever two or more sets are in T, then so is their union.

This definition can be used to enumerate the topologies on n symbols. For example, the unique topology of order 1 is $\{\emptyset, \{1\}\}$, while the four topologies of order 2 are $\{\emptyset, \{1\}, \{1,2\}\}, \{\emptyset, \{1,2\}\}, \{\emptyset, \{1,2\}, \{2\}\}$, and $\{\emptyset, \{1\}, \{2\}, \{1,2\}\}$. The numbers of topologies on sets of cardinalities $n = 1, 2, \ldots$ are $1, 4, 29, 355, 6942, \ldots$.

A set X for which a topology T has been specified is called a topological space. For example, the set $X = \{1,2,3,4\}$ together with the subsets $T = \{\emptyset, \{1\}, \{2,3,4\}, \{1,2,3,4\}\}$ comprises a topology, and X is a topological space.

Definition *Vector Space Span* The span of subspace generated by vectors v_1 and v_2 in V is $\text{span}(v_1, v_2) = \{rv_1 + sv_2 : r, s \in \mathbb{R}\}$.

Definition *Dense* A set A in a first-countable space is dense in B if $B = A \cup L$, where L is the set of limit points of A. For example, the rational numbers are dense in the reals.

Definition *Supremum*

The supremum is the least upper bound of a set S, defined as a quantity M such that no member of the set exceeds M, but if ε is any positive quantity, however small, there is a member that exceeds $M - \varepsilon$. When it exists (which is not required by this definition, e.g., $\sup \mathbb{R}$ does not exist), it is denoted $\sup_{x \in S} x$ (or sometimes simply $\sup_S$ for short).

More formally, the supremum $sup_{x \in S} x$ for S a (nonempty) subset of the affinely extended real numbers $\mathbb{R}^+ = \mathbb{R} \cup \{\pm\infty\}$ is the smallest value y in $\mathbb{R}^+$ such that for all x in S we have $x \leq y$. Using this definition, $\sup_{x \in S} x$ always exists and, in particular, $\sup \mathbb{R} = \infty$.

Whenever a supremum exists, its value is unique. On the real line, the supremum of a set is the same as the supremum of its set closure.

Consider the real numbers with their usual order. Then for any set $M \subseteq \mathbb{R}$, the supremum $\sup M$ exists (in $\mathbb{R}$) if and only if M is bounded from above and nonempty.

Definition *Gram-Schmidt Orthonormalization* Gram-Schmidt orthogonalization, also called the Gram-Schmidt process, is a procedure which takes a nonorthogonal set of linearly independent functions and constructs an orthogonal basis over an arbitrary interval with respect to an arbitrary weighting function $w(x)$.

Definition *Operator Spectrum* Let T be a linear operator on a separable Hilbert space. The spectrum $\sigma(T)$ of T is the set of λ such that $(T - \lambda \mathbb{1})$ is not invertible on all of the Hilbert space, where the lambdas are complex numbers and $\mathbb{1}$ is the identity operator. The definition can also be stated in terms of the resolvent of an operator $\rho(T) = \{\lambda : (T - \lambda \mathbb{1})\text{is invertible}\}$, and then the spectrum is defined to be the complement of $\rho(T)$ in the complex plane. It is easy to demonstrate that $\rho(T)$ is an open set, which shows that the spectrum $\sigma(T)$ is closed.

Definition *Bounded operators* Let V be a vector space. An operator on V is a linear map $a : V \times V$ (i.e., $a(\lambda v + \mu w) = \lambda a(v) + \mu a(w)$ for all $\lambda, \mu \in K$ and $v, w \in V$). We usually write av for $a(v)$. If V is a normed vector space, we call an operator $a : V \times V$ bounded when

$$\| a \| = \sup\{\| av \|, v \in V, \| v \| = 1\} < \infty. \tag{1.32}$$

It easily follows that if a is bounded, then

$$a = \inf\{C \geq 0 | \| av \| \leq C \| v \| \forall v \in V\}. \tag{1.33}$$

Moreover, if a is bounded, then

$$\|av\| \leq \|a\| \|v\| \tag{1.34}$$

for all $v \in V$. Another useful property, which immediately follows, is

$$\|ab\| \leq \|a\| \|b\|. \tag{1.35}$$

When V is finite-dimensional, any operator is bounded. There are many ways to see this. Indeed, let $V = \mathbb{C}^n$, so that $B(\mathbb{C}^n) = M_n(\mathbb{C})$ is the space of all complex $n \times n$ matrices. The Hilbert space norm on $\mathbb{C}^n$ is the usual one, i.e., $\|z\|^2 = \sum_{k=1}^n \overline{z}_k z_k$, and the induced norm on $M_n(\mathbb{C})$ is (1).

For example, a is not bounded when $a(x) = x$, and bounded when $a(x) = \exp -x^2$.

Definition *Closed Operators* A closed operator $a : D(a) \to H$ is a linear map from a dense subspace $D(a) \subset H$ to H for which either of the following equivalent conditions holds:

1. If $f_n \to f$ in H for a sequence (f_n) in $D(a)$ and (af_n) converges in H, then $f \in D(a)$ and $af_n \to af$ in H.
2. The graph of a is closed.
3. The domain $D(a)$ is closed in the norm $\|f\|_a^2 = \|f\|^2 + \|af\|^2$.

Note that the $\|\,\textbf{.}\,\|_a$ comes from the new inner product $(f,g)_a = (f,g) + (af,ag)$ on $D(a)$. Hence $D(a)$ is a Hilbert space in the new inner product when a is closed. An operator a that is not closed may often be extended into a closed one. The condition for this to be possible is as follows.

Definition *Closable Operator* A closable operator $a : D(a) \to H$ is a linear map from a dense subspace $D(a) \subset H$ to H with the property that the closure $G(a)^-$ of its graph is itself the graph $G(a^-)$ of some operator a^- (called the closure of a). In other words, a^- is a closed extension of a. It is clear that a^-1 is uniquely defined by its graph $G(a^-) = G(a)^-$.

Definition *Graph* For any $a : X \to Y$, we define the graph of a to be the set

$$\{(x,y) \in X \times Y \mid Tx = y\}. \tag{1.36}$$

If X is any topological space and Y is Hausdorff, then it is straightforward to show that the graph of a is closed whenever a is continuous.

If X and Y are Banach spaces, and a is an everywhere-defined (i.e. the domain $D(a)$ of a is X) linear operator, then the converse is true as well. This is the content of the closed graph theorem: if the graph of a is closed in $X \times Y$ (with the product topology), we say that a is a closed operator, and, in this setting, we may conclude that a is continuous.

Definition *Ortogonality* We say that two vectors $f, g \in H$ are orthogonal, written $f \perp g$, when $(f,g) = 0$. Similary, two subspaces10 $K \subset H$ and $L \subset H$ are said to be orthogonal $(K \perp L)$ when $(f,g) = 0$ for all $f \in K$ and all $g \in L$. A vector f is called orthogonal to a subspace K, written $f \perp K$, when $(f,g) = 0$ for all $g \in K$, etc. We define the orthogonal complement $K^\perp$ of a subspace $K \subset H$ as

$$K^\perp = \{f \in H | f \perp K\}. \tag{1.37}$$

This set is automatically linear, so that the map $K \mapsto K^\perp$, called orthocomplementation, is an operation from subspaces of H to subspaces of H. Clearly, $H^\perp = 0$ and $0^\perp = H$.

Lemma 1.6.1 *For any subspace $K \subset H$ one has $\overline{K^{\perp}} = \overline{K}^{\perp} = K^{\perp}$.*

Definition *The adjoint* Let H be a Hilbert space, and let $a : H \to H$ be a bounded operator. The inner product on H gives rise to a map $a \mapsto a^*$, which is familiar from linear algebra: if $H = \mathbb{C}^n$, so that, upon choosing the standard basis (e_i), a is a matrix $a = (a_{ij})$ with $a_{ij} = (e_i, ae_j)$, then the adjoint is given by $a^* = (\overline{a_{ji}})$. In other words, one has

$$(a^* f, g) = (f, ag) \tag{1.38}$$

for all $f, g \in \mathbb{C}^n$. This equation defines the adjoint also in the general case, but to prove existence of a^* a theorem is needed.

Definition *Banach Space* A Banach space is a complete vector space V with a norm $\| \mathbin{.} \|$.

Hilbert spaces with their norm given by the inner product are examples of Banach spaces. While a Hilbert space is always a Banach space, the converse need not hold. Therefore, it is possible for a Banach space not to have a norm given by an inner product. For instance, the supremum norm cannot be given by an inner product.

The supremum norm is the norm defined on V by $\| f \| = \sup_{x \in K} |f(x)|$.

Definition *C^*-algebra* A C^*-algebra is a Banach algebra with an antiautomorphic involution $*$ which satisfies

- $(x^*)^* = x$
- $x^* y^* = (yx)^*$
- $x^* + y^* = (x + y)^*$
- $(cx)^* = \overline{c} x^*$

where $\overline{c}$ is the complex conjugate of c, and whose norm satisfies $\|xx^*\| = \|x\|^2$.

A Banach algebra is an algebra B over a field K endowed with a norm $\| \mathbin{.} \|$ such that B is a Banach space under the norm and $\|xy\| \leq \|x\| \|y\|$. K is frequently taken to be the complex numbers in order to ensure that the operator spectrum fully characterizes an operator (i.e., the spectral theorems for normal or compact normal operators do not, in general, hold in the operator spectrum over the real numbers). If B is commutative and has a unit, then x in B is invertible.

Definition *Quotient Space* The quotient space $X/\sim$? of a topological space X and an equivalence relation $\sim$ on X is the set of equivalence classes of points in X (under the equivalence relation $\sim$) together with the following topology given to subsets of $X/\sim$: a subset U of $X/\sim$ is called open iff $\cup_{[a] \in U} a$ is open in X. Quotient spaces are also called factor spaces.

This can be stated in terms of maps as follows: if $q : X \to X/\sim$ denotes the map that sends each point to its equivalence class in $X/\sim$, the topology on $X/\sim$ can be specified by prescribing that a subset of $X/\sim$ is open iff $q^(- 1)$ [the set] is open.

Let D^n be the closed n-dimensional disk and $S^(n-1)$ its boundary, the $(n-1)$-dimensional sphere. Then $D^n/S^(n-1)$ (which is homeomorphic to S^n), provides an example of a quotient space. Here, $D^n/S^(n-1)$ is interpreted as the space obtained when the boundary of the n-disk is collapsed to a point, and is formally the "quotient space by the equivalence relation generated by the relations that all points in $S^(n-1)$ are equivalent."

Definition Let V be a vector space over a field K (where $K = \mathbb{R}$ or $K = \mathbb{C}$). An inner product on V is a map $V \times V \to K$, written as $\langle f, g \rangle \mapsto (f, g)$, satisfying, for all $f, g, h \in V$ and $t \in K$:

1. $(f, f) \in \mathbb{R}^+ = [0, \infty)$ (positivity);
2. $(g, f) = \overline{(f, g)}$ (symmetry);
3. $(f, tg) = t(f, g)$ (linearity 1);
4. $(f, g+h) = (f, g) + (f, h)$ (linearity 2);
5. $(f, f) = 0 \Rightarrow f = 0$ (positive definiteness).

A **norm** on V is a function $\| \cdot \| : V \to \mathbb{R}^+$ satisfying, for all $f, g, h \in V$ and $t \in K$:

1. $\|f + g\} \leq \| f \| + \| g \|$ (triangle inequality);
2. $\|tf\| = | t | \, \| f \|$ (homogeneity);
3. $\| f \| = 0 \Rightarrow f = 0$ (positive definiteness).

A **metric** on V is a function $d : V \times V \to \mathbb{R}^+$ satisfying, for all $f, g, h \in V$:

1. $d(f, g) \leq d(f, h) + d(h, g)$ (triangle inequality);
2. $d(f, g) = d(g, f)$ for all $f, g \in V$ (symmetry);
3. $d(f, g) = 0 \Leftrightarrow f = g$ (definiteness).

The notion of a metric applies to any set, not necessarily to a vector space, like an inner product and a norm. Apart from a norm, an inner product defines another structure called a transition probability, which is of great importance to quantum mechanics. Abstractly, a transition probability on a set S is a function $p : S \times S \to [0, 1]$ satisfying $p(x, y) = 1 \Leftrightarrow x = y$ (cf. Property 3 of a metric) and $p(x, y) = p(y, x)$.

Now take the set S of all vectors in a complex inner product space that have norm 1, and define an equivalence relation on S by $f \sim g$ iff $f = zg$ for some $z \in \mathbb{C}$ with $| z | = 1$. (Without taking equivalence classes the first axiom would not be satisfied). The set $S = S/\sim$ is then equipped with a transition probability defined by $p([f], [g]) = | (f, g) |^2$. Here $[f]$ is the equivalence class of f with $f = 1$, etc. In quantum mechanics vectors of norm 1 are (pure) states, so that the transition probability between two states is determined by their angle θ. (Recall the elementary formula from Euclidean geometry $(x, y) = \|x\| \|y\| \cos \theta$, where θ is the angle between x and y in $\mathbb{R}^n$.)

These structures are related in the following way:

Proposition 1.6.2 *1. An inner product on V defines a norm on V by means of $\| f \| = \sqrt{(f, f)}$.*

2. A norm on V defines a metric on V through $d(f, g) = \| f - g \|$.

The proof of this claim is an easy exercise; part 1 is based on the Cauchy-Schwarz inequality

$$| (f, g) | \leq \| f \| \| g \|, \tag{1.39}$$

whose proof in itself is an exercise, and part 2 is really trivial: the three axioms on a norm immediately imply the corresponding properties of the metric. The question arises when a norm comes from an inner product in the stated way: this question is answered by the Jordan-vonNeumann theorem:

Theorem 1.6.3 *A norm $\|\,.\,\|$ on a vector space comes from an inner product through $\|f\| = \sqrt{(f,f)}$ if and only if*

$$\|f+g\|^2 + \|f-g\|^2 = 2\left(\|f\|^2 + \|g\|^2\right) \tag{1.40}$$

Applied to the l^p and L^p spaces this yields the result that the norm in these spaces comes from an inner product if and only if p = 2. There is no (known) counterpart of this result for the transition from a norm to a metric. It is very easy to find examples of metrics that do not come from a norm: on any vector space (or indeed any set) V the formula $d(f,g) = \delta_{fg}$ defines a metric not derived from a norm. Also, if d is any metric on V, then $d' = d/(1+d)$ is a metric, too: since cleary $d'(f,g) \leq 1$ for all f, g, this metric can never come from a norm.

Theorem 1.6.4 Riesz-Fischer *A function is mbls integrable iff its Fourier series is* L^2*-convergent. The application of this theorem requires use of the Lebesgue integral.*

Theorem 1.6.5 Parseval *If a function has a Fourier series given by*

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx), \tag{1.41}$$

then

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2 dx = \sum_{n=-\infty}^{\infty} |a_n|^2. \tag{1.42}$$

Any separable Hilbert space has an orthonormal basis. The proof is an exercise, based on the Gram-Schmidt procedure. Let (e_i) by an orthonormal basis of a separable Hilbert space H. By definition, any $f \in H$ can be written as $f = \sum_i c_i e_i$. Taking the inner product with a fixed e_j, one has

$$(e_j, f) = (e_j, \lim_{\mathrm{N}\to\infty} \sum_{i=1}^{\mathrm{N}} c_i e_i) = \lim_{\mathrm{N}\to\infty} c_i\,(e_j, e_i) = c_j. \tag{1.43}$$

Here we have assumed $N > k$, and the limit may be taken outside the inner product since if $f_n \to f$ in H then $(g, f_n) \to (g, f)$ for fixed $g \in H$, as follows from Cauchy-Schwarz. It follows that

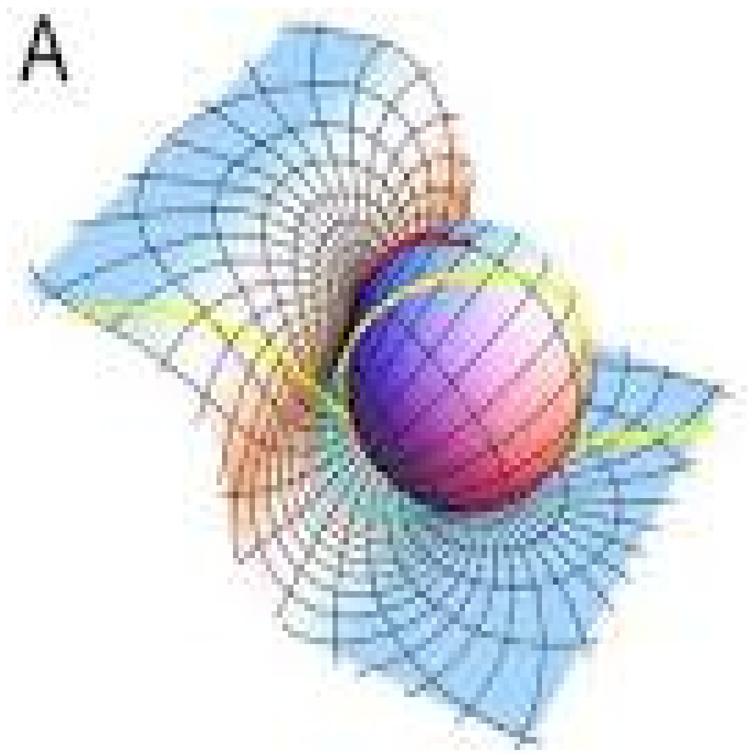
$$f = \sum_i (e_i, f)\, e_i \tag{1.44}$$

from which one obtains Parseval's equality

$$\sum_i |(e_i, f)|^2 = \|f\|^2. \tag{1.45}$$

Chapter 2

Special Functions

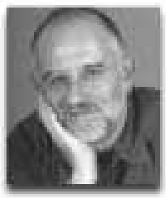


2.1 Introduction

Reference Frame

Why are special functions special?

Michael Berry



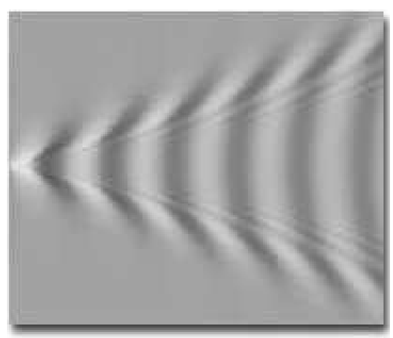
According to legend, Leo Szilard's baths were ruined by his conversion to biology. H had enjoyed soaking for hours while thinking about physics. But as a convert he found this pleasure punctuated by the frequent need to leap out and search for a fac In physics--particularly theoretical physics--we can get by with a few basic principles without knowing many facts; that is why the subject attracts those of us cursed with poor memory.

But there is a corpus of mathematical information that we do need. Much of this consists of formulas for the "special" functions. How many of us remember the expansion of cos $5x$ in terms of cos x and sin x, or whether an integral obtained in the course of a calculation can be identified as one of the many representations of a Bessel function, or whether the asymptotic expression for the gamma function involves (n + 1/2) or (n - 1/2)? For such knowledge, we theorists have traditionally relied on compilations of formulas. When I started research, my peers were using Jahnke and Emde's *Tables of Functions with Formulae and Curves* (J&E)[1] or Erdélyi and coauthors' *Higher Transcendental Functions*.[2]

Then in 1964 came Abramowitz and Stegun's *Handbook of Mathematical Functions* (A&S),[3] perhaps the most successful work of mathematical reference eve published. It has been on the desk of every theoretical physicist. Over the years, I have worn out three copies. Several years ago, I was invited to contemplate being marooned on the proverbial desert island. What book would I most wish to have there, in addition to the Bible and the complete works of Shakespeare? My immedia answer was: A&S. If I could substitute for the Bible, I would choose Gradsteyn and Ryzhik's *Table of Integrals, Series and Products*.[4] Compounding the impiety, I would give up Shakespeare in favor of Prudnikov, Brychkov and Marichev's of *Integrals and Series*.[5] On the island, there would be much time to think about physics and much physics to think about: waves on the water that carve ridges on th sand beneath and focus sunlight there; shapes of clouds; subtle tints in the sky. . . . With the arrogance that keeps us theorists going, I harbor the delusion that it would be not too difficult to guess the underlying physics and formulate the governing equations. It is when contemplating how to solve these equations--to convert formulations into explanations--that humility sets in. Then, compendia of formulas become indispensable.

Nowadays the emphasis is shifting away from books towards computers. With a few keystrokes, the expansion of cos 5*x*, the numerical values of Bessel functions, and many analytical integrals can all be obtained easily using software such as Mathematic and Maple. (In the spirit of the times, I must be even handed and refer to both the competing religions.) A variety of resources is available online. The most ambitious initiative in this direction is being prepared by NIST, the descendant of the US National Bureau of Standards, which published A&S. NIST's forthcoming Digital Library of Mathematical Functions (DLMF) will be a free Web-based collection of formulas (http://dlmf. nist.gov), cross-linked and with live graphics that can be magnified and rotated. (Stripped-down versions of the project will be issued as a book and a CD-ROM for people who prefer those media.)

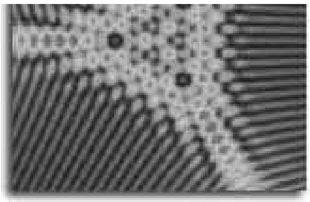
The DLMF will reflect a substantial increase in our knowledge of special functions since 1964, and will also include new families of functions. Some of these functions were (with one class of exceptions) known to mathematicians in 1964, but they were not well known to scientists, and had rarely been applied in physics. They are new in the sense that, in the years since 1964, they have been found useful in several branches of physics. For example, string theory and quantum chaology now make us of automorphic functions and zeta functions; in the theory of solitons and integrable dynamical systems, Painlevé transcendents are widely employed; and in optics and quantum mechanics, a central role is played by "diffraction catastrophe" integrals, generated by the polynomials of singularity theory--my own favorite, and the subject of a chapter I am writing with Christopher Howls for the DLMF.



KELVIN'S SHIP-WAVE pattern, calculated with the Airy function, the simplest special function in the hierarchy of diffraction catastrophes.

This continuing and indeed increasing reliance on special functions is a surprising development in the sociology of our profession. One of the principal applications of these functions was in the compact expression of approximations to physical problems for which explicit analytical solutions could not be found. But since the 1960s, when scientific computing became widespread, direct and "exact" numerical solution of the equations of physics has become available in many cases. It was often claimed that this would make the special functions redundant. Similar skepticism came from some pure mathematicians, whose ignorance about special functions, and lack of interest in them, was almost total. I remember that when singularity theory was being applied to optics in the 1970s, and I was seeking a graduate student to pursue these investigations, a mathematician recommended somebody as being very bright, very knowledgeable, and interested in applications. But this student had never heard of Bessel functions (nor could he carry out the simplest integrations, bu that is another story).

The persistence of special functions is puzzling as well as surprising. What are they, other than just names for mathematical objects that are useful only in situations of contrived simplicity? Why are we so pleased when a



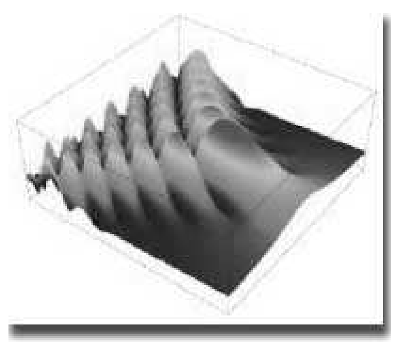
A CROSS SECTION of the elliptic umbilic, a member of the hierarchy of diffraction catastrophes.

complicated calculation "comes out" as a Bessel function, or a Laguerre polynomial? What determines which functions are "special"? These are slippery and subtle questions to which I do not have clear answers. Instead, I offer the following observations.

There are mathematical theories in which some classes of special functions appear naturally. A familiar classification is by increasing complexity, starting with polynomials and algebraic functions and progressing through the "elementary" or "lower" transcendental functions (logarithms, exponentials, sines and cosines, and so on) to the "higher" transcendental functions (Bessel, parabolic cylinder, and so on). Functions of hypergeometric type can be ordered by the behavior of singular points of the differential equations representing them, or by a group-theoretical analysis of their symmetries. But all these classifications are incomplete, in the sense of omitting whc classes that we find useful. For example, Mathieu functions fall outside the hypergeometric class, and gamma and zeta functions are not the solutions of simple differential equations. Moreover, even when the classifications do apply, the connections they provide often appear remote and unhelpful in our applications.

One reason for the continuing popularity of special functions could be that they enshrine sets of recognizable and communicable patterns and so constitute a common currency. Compilations like A&S and the DLMF assist the process of standardization, much as a dictionary enshrines the words in common use at a given time. Formal grammar, while interesting for its own sake, is rarely useful to those wh use natural language to communicate. Arguing by analogy, I wonder if that is why th formal classifications of special functions have not proved very useful in applications

Sometimes the patterns embodying special functions are conjured up in the form of pictures. I wonder how useful sines and cosines would be without the images, which we all share, of how they oscillate. In 1960, the publication in J&E of a three-dimensional graph showing the poles of the gamma function in the complex plane acquired an almost iconic status. With the more sophisticated graphics available now, the far more complicated behavior of functions of several variables can be explored in a variety of two-dimensional sections and three-dimensional plots, generating a large class of new and shared insights.



THE CUSP, a member of the hierarchy of diffraction catastrophes.

"New" is important here. Just as new words come into the language, so the set of special functions increases. The increase is driven by more sophisticated applications and by new technology that enables more functions to be depicted in forms that car be readily assimilated.

Sometimes the patterns are associated with the asymptotic behavior of the functions or of their singularities. Of the two Airy functions, Ai is the one that decays towards infinity, while Bi grows; the J Bessel functions are regular at the origin, the Y Bessel functions have a pole or a branch point.

Perhaps standardization is simply a matter of establishing uniformity of definition ar notation. Although simple, this is far from trivial. To emphasize the importance of notation, Robert Dingle in his graduate lectures in theoretical physics at the University of St. Andrews in Scotland would occasionally replace the letters representing variables by nameless invented squiggles, thereby inducing instant incomprehensibility. Extending this one level higher, to the names of functions, just imagine how much confusion the physicist John Doe would cause if he insisted on replacing sin x by doe(x), even with a definition helpfully provided at the start of eac paper.

To paraphrase an aphorism attributed to the biochemist Albert Szent-Györgyi, perhaps special functions provide an economical and shared culture analogous to books: places to keep our knowledge in, so that we can use our heads for better things.

References

1. E. Jahnke, F. Emde, *Tables of Functions with Formulae and Curves*, Dover Publications, New York (1945).
2. A Erdélyi, W. Magnus, F. Oberhettinger, F. G. Tricomi, *Higher Transcendenta Functions*, 5 vols., Krieger Publishing, Melbourne, Fla. (1981) [first published in 1953].
3. M. Abramowitz, I. A. Stegun, eds., *Handbook of Mathematical Functions wit Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards Applied Mathematics Series, vol. 55, US Government Printing Office, Washington, DC (1964).
4. I. S. Gradsteyn, I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th e (translated from Russian by Scripta Technika), Academic Press, New York (2000) [first published in 1965].
5. A. P. Prudnikov, Yu. A. Brychkov, O. I. Marichev, *Integrals and Series*, 5 vols. (translated from Russian by N. M. Queen), Gordon and Breach, New York (1986-1992).

Michael Berry *is Royal Society Research Professor in the physics department of Bristol University, UK, and studies physical asymptotics.*

2.2 Elementary Transcendental Functions

LECTURE 8

N^{th} Roots; Univalent Functions

With Euler's formula in hand we are now able to define radicals of complex numbers.

DEFINITION 8.1. *Let $w \neq 0$, $n \in \mathbb{N}$, a number z is called an* <u>n^{th} *root*</u> *of w if $z^n = w$.*

PROPOSITION 8.2 (Formula for the n^{th} roots). *Let $w = |w|e^{i\theta} \neq 0$ and $n \in \mathbb{N}$, then w has n distinct n^{th} roots $z_1, z_2, ..., z_n$ given by*

$$z_k = \sqrt[n]{|w|} \exp i \frac{\theta + 2\pi(k-1)}{n}, \ k = 1, 2, ..., n \tag{8.1}$$

PROOF. Consider

$$z_1 = \sqrt[n]{|w|} e^{i\frac{\theta}{n}}.$$

By de Moivre's formula

$$z_1^n = |w|e^{i\theta} = w$$

and hence z_1 is a n^{th} root of w.

Define now

$$z_k := z_1 e^{i\frac{2\pi(k-1)}{n}}, \ k \in \mathbb{N} \tag{8.2}$$

One has

$$z_k^n = \underbrace{z_1^n}_{=w} \underbrace{e^{2\pi i(k-1)}}_{=1} = w$$

Thus so-defined sequence $\{z_k\}$ gives n^{th} roots. It follows from (8.2) that

$$z_k = z_{k-1} e^{i\frac{2\pi}{n}}$$

which implies (the details should be verified in Exercise 8.1) that all $z_1, z_2, ..., z_n$ are distinct but $z_{n+1} = z_1$, $z_{n+2} = z_2$, ... , $z_{2n} = z_n$. $\square$

EXAMPLE 8.3. *Find the sixth roots of unity, i.e., find all solutions to $z^6 = 1$.*

By (8.1)

$$z_k = \exp i \frac{2\pi(k-1)}{6}, \ k = 1, 2, ..., 6$$

which gives $z_1 = 1, \ z_2 = 1/2 + i\sqrt{3}/2, \ z_3 = -1/2 + i\sqrt{3}/2, \ z_4 = -1.$

To find z_5 and z_6 notice that equation $z^6 = 1$ is equivalent to $\overline{z}^6 = 1$ and hence $\overline{z_2}, \overline{z_3}$ are also solutions to $z^6 = 1$. Thus $z_5 = \overline{z_3} = -1/2 - i\sqrt{3}/2$, and $z_6 = \overline{z_2} = 1/2 - i\sqrt{3}/2$.

Figure 1 shows the six sixth roots of unity on the complex plane.

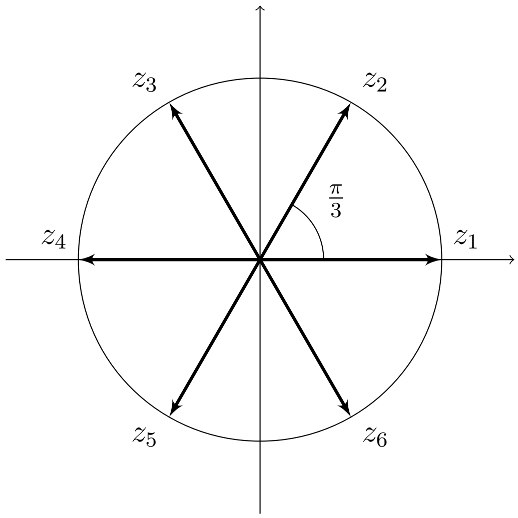


FIGURE 1. The sixth roots of the unity

REMARK 8.4. *Proposition 8.2 says that the equation*

$$z^{2008} = 1$$

has 2008 distinct solutions and if you mistakenly think that $z = \pm 1$ *then you lose 2006 solutions.*

What we have discussed in this lecture so far suggests that defining the function $\sqrt[n]{z}$ will require some effort.

DEFINITION 8.5. *An analytic function* $f : E \to \mathbb{C}$ *is called* *univalent* *on* E *if*

$$f(z_1) = f(z_2) \Rightarrow z_1 = z_2.$$

Note that generic complex valued functions don't like to be univalent. E.g. $f(z) = |z|$ maps any circle $|z| = r$ onto one point $\{r\} \in \mathbb{R}$, and hence $|z|$ is very non-univalent.

EXAMPLE 8.6. *The linear function* $f(z) = az + b$ *is univalent (if* $a \neq 0$*) on* $\mathbb{C}$*. The function* $\phi(z) = z/a - b/a$ *satisfies*

$$f \circ \phi = \phi \circ f = \mathbb{I}$$

and hence ϕ *can be viewed as the inverse of* f*.*

EXAMPLE 8.7. *The function* $f(z) = 1/z$ *is univalent on* $\mathbb{C} \setminus \{0\}$ *and its inverse is* $\phi(z) = 1/z$*.*

EXAMPLE 8.8. *The function* $f(z) = z^2$ *is not univalent on* $\mathbb{C}$*. But as in the real valued case we can properly restrict* z^2 *to make it univalent. To do so we consider* z^2 *as a* *mapping* *on* $\mathbb{C}$*.*

We introduce a *sector*:

$$S(\alpha, \beta) := \{z | \ \alpha < \arg z < \beta\}.$$

Let $z \in S(\alpha, \beta)$. We have $z = |z|e^{i\theta}, \alpha < \theta < \beta$, and hence

$$z^2 = |z|^2 e^{2i\theta} \in S(2\alpha, 2\beta).$$

In particular, if $\alpha = 0$ and $\beta = 2\pi$ then

$$z^2 : S(0, 2\pi) \to S(0, 4\pi).$$

That is, loosely speaking, z^2 maps $\mathbb{C}$ onto two copies of $\mathbb{C}$.
However if we take $\alpha = 0$ and $\beta = \pi$ then

$$z^2 : S(0, \pi) \to S(0, 2\pi)$$

which is a univalent function. See Figure 2 for illustration.

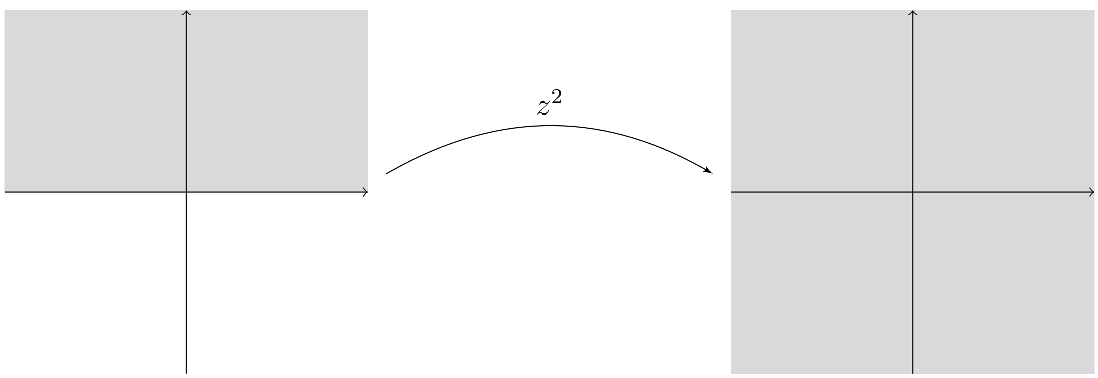


FIGURE 2. Transformation of the upper half plane under z^2

Exercises

Exercise 8.1 Let $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ be the n^{th} roots of unity. Pick a fixed $\omega \in \Omega$. Show

(1) $\omega\omega_k \in \Omega$.
(2) $\omega\omega_k \neq \omega\omega_m$ if $k \neq m$, i.e., $\{\omega\omega_k\}_{k=1}^n = \Omega$.
(3) $\displaystyle\sum_{k=1}^{n} \omega_k = 0$.
(4) $\displaystyle\sum_{k=0}^{n-1} \omega^k = 0$ for any $\omega \neq 1$.

Exercise 8.2 Describe a linear function as a mapping, i.e., describe what geometric transformation $f(z) = az + b$ represents on $\mathbb{C}$. Find the images $f(\Omega)$ of Ω for the following sets:

(1) $\Omega = \{z| \ \arg z = \theta_0\}$, where θ_0 is fixed. (A ray)
(2) $\Omega = \{z| \ |z| = r_0\}$, where r_0 is fixed. (A circle)

Exercise 8.3 Let $f(z) = 1/z$. Find the images $f(\Omega)$ of the following sets:

(1) Ω is a straight line passing through 0.
(2) $\Omega = \{z| \ |z| = r_0\}$, where r_0 is fixed.

Describe the geometrical transformation the function $f(z)$ does to point z in Figure 3.

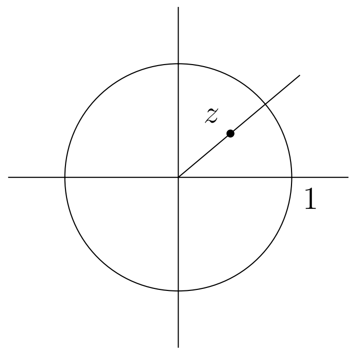


FIGURE 3. Transformation of z under $1/z$?

LECTURE 9

z^n, $\sqrt[n]{z}$, e^z, $\log z$ and all that

In this lecture we continue to study elementary functions of a single complex variable as mappings.

1. The Function z^n

Let $z = re^{i\theta}$. Then

$$w := z^n = r^n e^{in\theta}$$

and hence z^n maps a sector $S(\alpha, \beta)$ onto the sector $S(n\alpha, n\beta)$. In particular, if $\alpha = 0$ and $\beta = 2\pi$, then z^n maps $\mathbb{C} = S[0, 2\pi)$ onto n copies of $\mathbb{C}$. In other words, z^n is an n-valent function.

Now restrict z^n to $S\left(0, \frac{2\pi}{n}\right)$. Then $w \in S(0, 2\pi)$. Moreover, every ray

$$R_\theta := \left\{ z \,|\, z = re^{i\theta}, r \in \mathbb{R}_+ \right\}, \quad \text{where } \mathbb{R}_+ = [0, \infty),$$

in $S\left(0, \frac{2\pi}{n}\right)$ is mapped onto the ray $R_{n\theta}$.

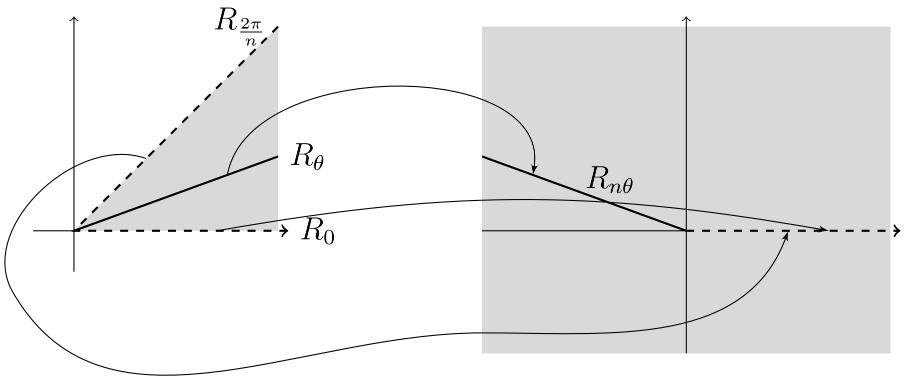


Figure 1. Mapping of a sector under z^n

Observe that $z^n : S\left(0, \frac{2\pi}{n}\right) \to S(0, 2\pi)$ is univalent and we may now introduce the inverse of z^n.

2. The Function $\sqrt[n]{z}$

As we have seen in Lecture 8, the equation $z^n = w$ has n solutions $\{z_1, z_2, \ldots, z_n\}$. In order to define $\sqrt[n]{w}$ we have to decide which of these solutions we want to pick up. It is a personal choice, but common sense suggests that we choose z_1. Doing so allows $\sqrt[n]{1} = 1$, and has the nice property that our complex n^{th} root function will correspond to the usual positive n^{th} root function on the real number line (or on $\mathbb{R}_+$ if n is even).

Definition 9.1. *Let $z = |z|e^{i\arg z} \neq 0$ where $\arg z \in [0, 2\pi)$. The principal branch of $\sqrt[n]{z}$ is defined by*

$$\sqrt[n]{z} = \sqrt[n]{|z|}e^{\frac{i\arg z}{n}}. \tag{9.1}$$

Remark 9.2. *$\sqrt[n]{z}$ can also be denoted as $z^{\frac{1}{n}}$. In some books, they distinguish between these two representations, using one to denote the multi-valued function and the other to denote the principal branch. We choose not to do so in this course. Also, note that the principal n^{th} root of z defined by (9.1) is z_1 in the n^{th} root formula (given in Lecture 8). If occasionally we choose a different branch of $\sqrt[n]{z}$ we must specify how we define it.*

Let us now look at (9.1) as a function of z. Formula (9.1) defines a one-to-one function $\sqrt[n]{z}$,

$$\sqrt[n]{z} : S(0, 2\pi) \to S(0, 2\pi/n). \tag{9.2}$$

Proposition 9.3. *The function $\sqrt[n]{z}$ defined by (9.1) and (9.2) is analytic on $\mathbb{C}\setminus\mathbb{R}_+$ but not analytic on $\mathbb{C}$.*

The proof is left as an exercise.

3. The Function e^z

Consider a strip

$$K_0 = \{z : 0 < \operatorname{Im} z < 2\pi\}.$$

Let $z = x + iy \in K_0$. Then $e^z = e^x e^{iy}$ maps K_0 onto $\mathbb{C} \setminus \mathbb{R}_+$. Moreover, every vertical segment

$$k_x = \{z = x + iy \, : \, y \in (0, 2\pi)\}$$

is mapped onto the circle $C_{e^x}(0) \setminus \mathbb{R}_+$, where

$$C_r(z_0) := \{z \, : \, |z - z_0| = r\}.$$

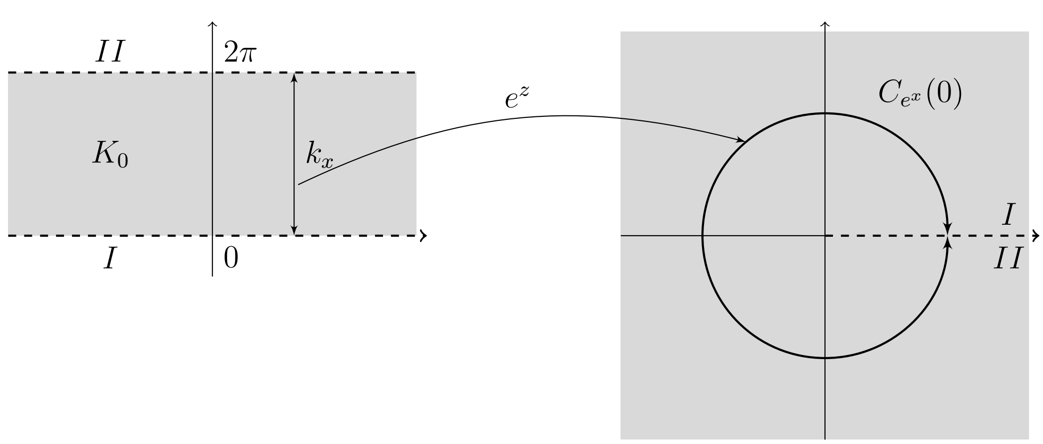


Figure 2. Mapping of a strip under e^z

Observe that e^z is univalent on K_0. Let

$$K_{2\pi} := \{z \,:\, 2\pi < \operatorname{Im} z < 4\pi\}.$$

It is clear that e^z maps $K_{2\pi}$ onto $\mathbb{C} \setminus \mathbb{R}_+$. Similarly, for all $n \in \mathbb{Z}$,

$$\exp K_{2\pi n} = \mathbb{C} \setminus \mathbb{R}_+,$$

which means that e^z is an infinitely-valent (infinitely valued) function.

REMARK 9.4. *We keep our strips open for a purpose which will be clear in the next section. Of course, e^z maps $\mathbb{R}$ onto $\mathbb{R}_+ \setminus \{0\}$ and maps*

$$\mathbb{R} + 2i\pi := \{z \,|\, z = x + iy, x \in \mathbb{R}, y = 2\pi\}$$

onto $\mathbb{R}_+ \setminus \{0\}$.

4. The Function log z

Since the mapping

$$e^z : K_0 \to \mathbb{C} \setminus \mathbb{R}_+$$

is one-to-one, we can formally define the complex logarithm.

DEFINITION 9.5. *Let $z = |z|e^{i \arg z} \neq 0$ where $\arg z \in [0, 2\pi)$. The principal branch of $\log z$ is defined by*

$$\log z = \log |z| + i \arg z \tag{9.3}$$

where $\log |z|$ is the usual natural logarithm of the positive real number $|z|$.

REMARK 9.6. *In the literature, the complex natural logarithm is sometimes denoted* ln *or* Log. *In complex analysis* log, ln, *and* Log *all refer to the logarithm base e, as opposed to real analysis where* log *often denotes the logarithm base* 10.

Let us now look at (9.3) as a function of z:

$$\log z : \mathbb{C} \setminus \mathbb{R}_+ \to K_0. \tag{9.4}$$

PROPOSITION 9.7. *The function $\log z$ defined by* (9.3) *and* (9.4) *is the inverse of*

$$e^z : K_0 \to \mathbb{C} \setminus \mathbb{R}_+.$$

The proof is left as an exercise.

PROPOSITION 9.8. *The complex logarithm function $\log z$ is analytic on $\mathbb{C} \setminus \mathbb{R}_+$, but is not analytic on $\mathbb{C}$. Furthermore*

$$(\log z)' = \frac{1}{z} \quad , \quad z \in \mathbb{C} \setminus \mathbb{R}_+.$$

The proof is left as an exercise.

Exercises

Exercise 9.1 Prove Proposition 9.3.

Exercise 9.2 Prove Proposition 9.7.

Exercise 9.3 Prove Proposition 9.8.

LECTURE 10

$\log z$ continued; $\sin z$, and $\cos z$

1. $\log z$ continued

In this lecture we continue studying elementary functions. We start with an important remark.

REMARK 10.1. *As we have seen, e^z retains all of its original properties and on top of that, e^z also gains some new ones. (eg.* $\mathrm{Ran}(e^z) = \mathbb{C} \setminus 0$ *not just* $\mathbb{R}_+$*, and e^z is now periodic). On the contrary* $\log z$ *loses some of its common properties. E.g. the property* $\log(z_1 z_2) = \log z_1 + \log z_2$*, alas, no longer holds in general. So watch out!*

REMARK 10.2. *Our definition of the principal branch of* $\log z$ *(Definition 9.5) is not the only one. All depends on a particular situation. Another common way to choose the principal branch is to let $z = |z|e^{i \arg z}$ where* $\arg z \in [-\pi, \pi)$*, (not* $[0, 2\pi)$*). Then*

$$\log z := \log |z| + i \arg z. \tag{10.1}$$

I.e. this $\log$ *is defined on* $\mathbb{C} \setminus \mathbb{R}_-$.

The log defined by (10.1) has the nice property $\log \bar{z} = \overline{\log z}$ where as the one we defined before did not.

Before we are done with logarithms, let us state and prove an important proposition.

PROPOSITION 10.3. *Suppose $f : E \to \mathbb{C}$ and $g : f(E) \to \mathbb{C}$ are continuous and*

$$g(f(z)) = z. \tag{10.2}$$

Then,

(1) *if g is differentiable and $g'(z) \neq 0$, then f is differentiable and*

$$f'(z) = (g'(f(z)))^{-1} \tag{10.3}$$

(2) *if g is analytic and $g'(f(z)) \neq 0$, then f is analytic.*

PROOF. Assume that f is continuous and let $z, z + \Delta z \in E$. Then from (10.2) we see that:

$$g(f(z)) = z, \ g(f(z + \Delta z)) = z + \Delta z \tag{10.4}$$

$$\begin{aligned}
&\Rightarrow f(z) \neq f(z + \Delta z) \text{ if } \Delta z \neq 0 \\
&\Rightarrow \Delta f(z) := f(z + \Delta z) - f(z) \neq 0 \\
&\Rightarrow f(z + \Delta z) = f(z) + \Delta f(z).
\end{aligned}$$

Consider $\dfrac{g(f(z+\Delta z))-g(f(z))}{\Delta z}$. We have by equation (10.4)

$$1=\frac{g(f(z+\Delta z))-g(f(z))}{\Delta z}=\underbrace{\frac{g(f(z)+\Delta f(z))-g(f(z))}{\Delta f(z)}}_{\to g'(f(z)) \text{ as } \Delta f(z)\to 0}\cdot\frac{\Delta f(z)}{\Delta z}.$$

Then by taking the limit as Δz goes to zero we see that

$$\begin{aligned} 1 &= \lim_{\Delta z\to 0}\frac{g(f(z)+\Delta f(z))-g(f(z))}{\Delta f(z)}\cdot\frac{\Delta f(z)}{\Delta z} \\ \Rightarrow \quad 1 &= g'(f(z))\cdot\lim_{\Delta z\to 0}\frac{\Delta f(z)}{\Delta z}. \end{aligned} \tag{10.5}$$

(We have used $\Delta f(z)\to 0$ as $\Delta z\to 0$ due to continuity of f.) It now follows from (10.5) that the limit on the right hand side exists and (10.3) follows. Statement (1) is proven.

If g is analytic then g' is continuous and so f' is continuous. Then statement (2) is also proven because f' is continuous which implies that f is analytic.

$\square$

Corollary 10.4. *The function* $\log z$ *is analytic on* $\mathbb{C}\setminus\mathbb{R}_+$ *and*

$$(\log z)'=\frac{1}{z}. \tag{10.6}$$

Proof. Using Proposition 10.3 $f(z)=\log z$, $E=\mathbb{C}\setminus\mathbb{R}_+$ and $g(z)=e^z$. Since

$$\exp\log z=z \text{ for all } z\in\mathbb{C}\setminus\mathbb{R}_+$$

(10.2) holds and the conditions of Proposition 10.3 are satisfied since e^z is analytic. Thus, $\log z$ is also analytic and by (10.3)

$$\begin{aligned}(\log z)' &= \left(\frac{d}{dw}e^w\,|_{w=\log z}\right)^{-1} \\ &= (e^{\log z})^{-1}=\frac{1}{z}.\end{aligned}$$

$\square$

Remark 10.5. *Did you enjoy proving it by definition? Would Proposition 10.3 have helped prove Exercise 9.3?*

2. The Functions $\sin z$, $\cos z$, and $\tan z$

We restrict ourselves to $\sin z$ only. Then by definition,

$$\begin{aligned}\sin z &= \frac{e^{iz}-e^{-iz}}{2i} = \frac{e^{i(x+iy)}-e^{-i(x+iy)}}{2i} = \frac{e^{-y}e^{ix}-e^{y}e^{-ix}}{2i}\\ &= \frac{1}{2i}(e^{-y}(\cos x + i\sin x) - e^{y}(\cos x - i\sin x))\\ &= \frac{1}{2i}(e^{-y}\cos x + ie^{-y}\sin x - e^{y}\cos x - ie^{y}\sin x)\\ &= \frac{e^{y}\sin x + e^{-y}\sin x}{2} + i\frac{e^{y}\cos x - e^{-y}\cos x}{2}\\ &= \sin x\cosh y + i\cos x\sinh y.\end{aligned}$$

I told you once that it is almost never a good idea to separate the real and imaginary parts of the function. This is a rare case of when it is needed. So we get

$$\sin(x+iy) = \underbrace{\sin x\cosh y}_{u} + i\underbrace{\cos x\sinh y}_{v}.$$

We have

$$\sin x = \frac{u}{\cosh y}, \qquad \cos x = \frac{v}{\sinh y}$$

and hence

$$1 = \sin^2 x + \cos^2 x = \left(\frac{u}{\cosh y}\right)^2 + \left(\frac{v}{\sinh y}\right)^2. \tag{10.7}$$

It follows from (10.7) that $\sin z$ maps a segment $\{z|z = x+iy_0, x \in [-\frac{\pi}{2}, \frac{\pi}{2}]\}$, where $y_0 \geq 0$ is fixed, onto the upper semiellipse

$$\left(\frac{u}{\cosh y_0}\right)^2 + \left(\frac{v}{\sinh y_0}\right)^2 = 1.$$

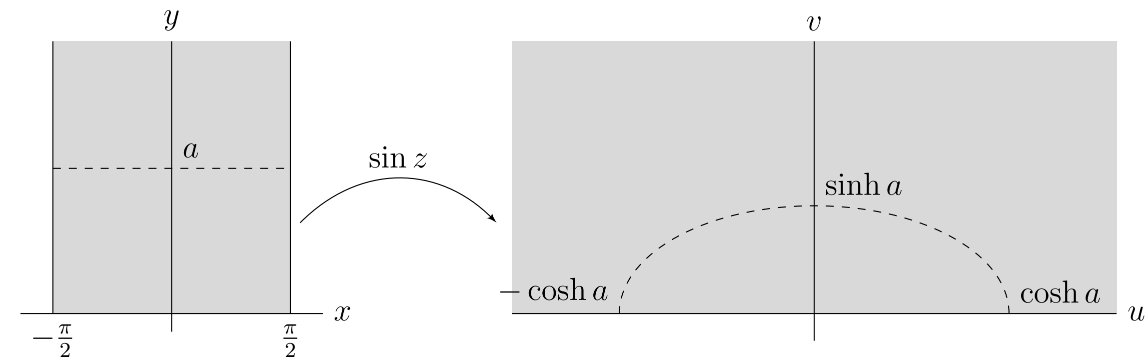


FIGURE 1. Mapping of $\sin z$ for $z = x + ia$ where $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $a \geq 0$

Similarly, one can see

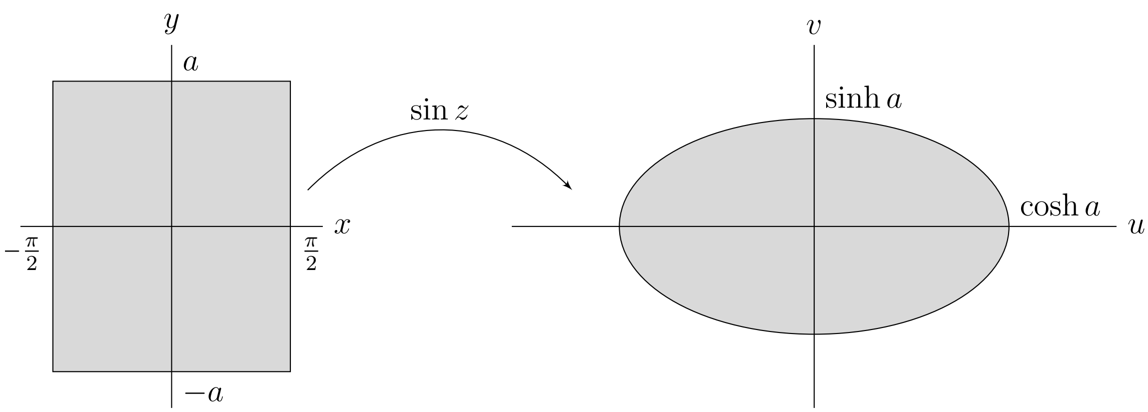


FIGURE 2. Mapping of $\sin z$ for $z = x + iy$ where $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $y \in [-a, a]$

Exercises

Exercise 10.1 Let $z_1, z_2, ..., z_n \in \mathbb{C}^+ := \{z | \operatorname{Im} z > 0\}$. Show that if $z_1 \cdot z_2 \cdot \ldots \cdot z_k \in \mathbb{C}^+$ for all $k \leq n$ then,

$$\log \prod_{k=1}^{n} z_k = \sum_{k=1}^{n} \log z_k, \tag{10.8}$$

where the log is defined on $\mathbb{C} \setminus \mathbb{R}_+$. Give a counterexample to equation (10.8) if we remove the restriction on $z_1, z_2, ..., z_n$. Note, in this respect, that in general

$$\log z^n \neq n \log z.$$

Exercise 10.2 Let $\log z$ be defined on $\mathbb{C} \setminus \mathbb{R}_+$. Come up with reasonable conditional statements regarding the basic properties of $\log z$.
E.g. $\log z_1 z_2 = \log z_1 + \log z_2$, if blah, blah, blah.

Exercise 10.3 Treat $\cos z$ in a way similar to what we did with $\sin z$.

2.3 More on Riemann Surfaces

Lecture 1

What are Riemann surfaces?

1.1 Problem: Natural algebraic expressions have 'ambiguities' in their solutions; that is, they define multi-valued rather than single-valued functions.

In the real case, there is usually an obvious way to fix this ambiguity, by selecting one *branch* of the function. For example, consider $f(x) = \sqrt{x}$. For real x, this is only defined for $x \geq 0$, where we conventionally select the positive square root (Fig.1.1).

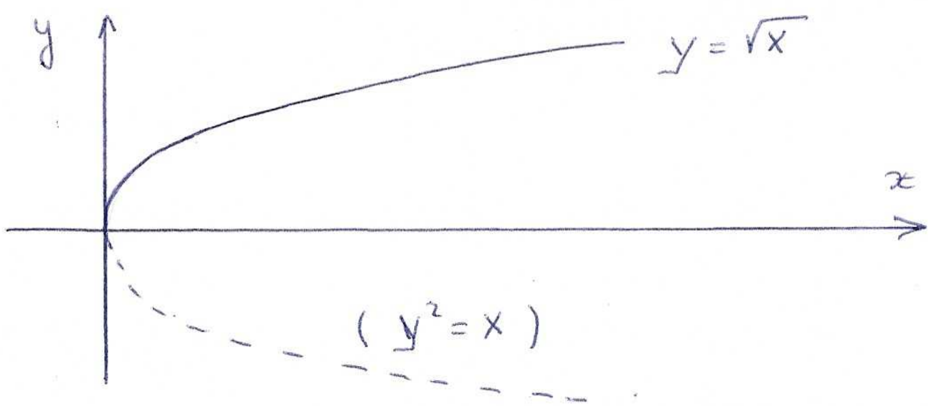


Fig.1.1

We get a continuous function $[0, \infty) \to \mathbb{R}$, analytic everywhere except at 0. Clearly, there is a problem at 0, where the function is not differentiable; so this is the best we can do.

In the complex story, we take '$w = \sqrt{z}$' to mean $w^2 = z$; but, to get a single-valued function of z, we must make a choice, and a *continuous* choice requires a *cut* in the domain.

A standard way to do that is to define '$\sqrt{z}$' : $\mathbb{C} \setminus \mathbb{R}^- \to \mathbb{C}$ to be the square root with positive real part. There is a unique such, for z away from the negative real axis. This function is continuous and in fact *complex-analytic*, or *holomorphic*, away from the *negative* real axis.

A different choice for $\sqrt{z}$ is the square root with positive imaginary part. This is uniquely defined away from the *positive* real axis, and determines a complex-analytic function on $\mathbb{C} \setminus \mathbb{R}^+$.

In formulae: $z = re^{i\theta} \implies \sqrt{z} = \sqrt{r}e^{i\theta/2}$, but in the first case we take $-\pi < \theta < \pi$, and, in the second, $0 < \theta < 2\pi$.

Either way, there is no continuous extension of the function over the missing half-line: when z approaches a point on the half-line from opposite sides, the limits of the chosen values of $\sqrt{z}$ differ by a sign. A restatement of this familiar problem is: starting at a point $z_0 \neq 0$ in the plane, any choice of $\sqrt{z_0}$, followed continuously around the origin once, will lead to the opposite choice of $\sqrt{z_0}$ upon return; z_0 needs to travel around the origin twice, before $\sqrt{z_0}$ travels once.

Clearly, there is a problem at 0, but the problem along the real axis is our own — there is no discontinuity in the function, only in the choice of value. We could avoid this problem by allowing multi-valued functions; but another point of view has proved more profitable.

The idea is to replace the complex plane, as domain of the multi-valued function, by the graph of the function. In this picture, the function becomes projection to the w-axis, which is well-defined single-valued! (Fig. 1.2)

In the case of $w = \sqrt{z}$, the graph of the function is a closed subset in $\mathbb{C}^2$,

$$S = \{(z, w) \in \mathbb{C}^2 \mid w^2 = z\}.$$

In this case, it is easy to see that the function $w = w(z)$,

$$S \to \mathbb{C}, \qquad (z, w) \mapsto w$$

defines a homeomorphism (*diffeomorphism*, in fact) of the graph S with the w-plane. This is exceptional; it will not happen with more complicated functions.

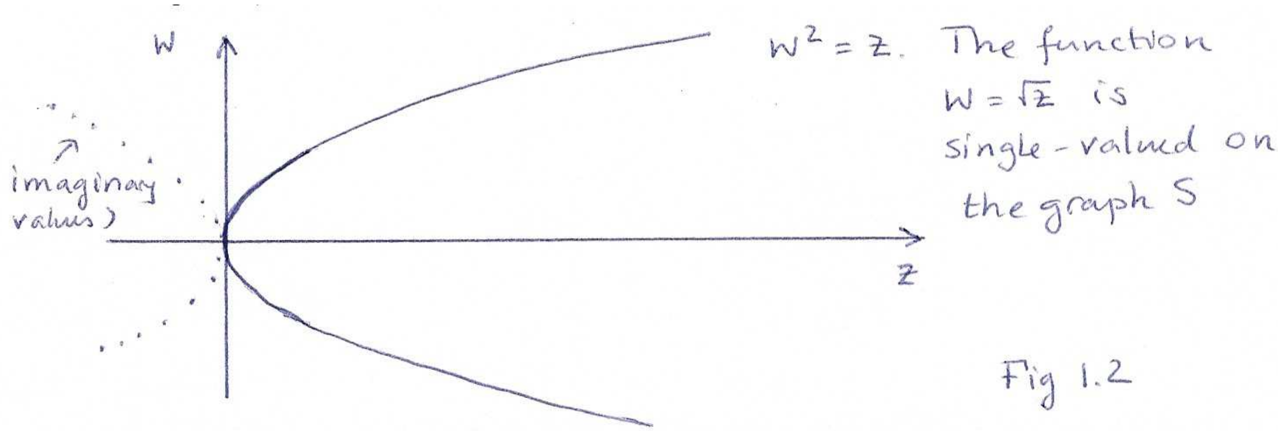


Fig 1.2

The graph S is a very simple example of a (concrete, non-singular) Riemann surface. Thus, the basic idea of Riemann surface theory is to replace the domain of a multi-valued function, e.g. a function defined by a polynomial equation

$$P(z, w) = w^n + p_{n-1}(z)w^{n-1} + \cdots + p_1(z)w + p_0(z)$$

by its graph

$$S = \{(z, w) \in \mathbb{C}^2 \mid P(z, w) = 0\},$$

and to study the function w as a function on the 'Riemann surface' S, rather than as a multi-valued function of z.

This is all well, provided we understand

- what kind of objects Riemann surfaces are;
- how to do complex analysis on them (what are the analytic functions?)

The two questions are closely related, as will become clear when we start answering them properly, in the next lecture; for now, we just note the moral definitions.

1.2 Moral definition. An *abstract Riemann surface* is a surface (a real, 2-dimensional manifold) with a 'good' notion of complex-analytic functions.

The most important examples, and the first to arise, historically, were the graphs of multi-valued analytic functions:

1.3 Moral definition: A *(concrete) Riemann surface* in $\mathbb{C}^2$ is a locally closed subset which is locally — near each of its points (z_0, w_0) — the graph of a multi-valued complex-analytic function.

1.4 Remarks:

(i) *locally closed* means *closed in some open set.* The reason for 'locally closed' and not 'closed' is that the domain of an analytic function is often an open set in $\mathbb{C}$, and not all of $\mathbb{C}$. For instance, there is no sensible way to extend the definition of the function $z \mapsto exp(1/z)$ to $z = 0$; and its graph is not closed in $\mathbb{C}^2$.

(ii) Some of the literature uses a more restrictive definition of the term *multi-valued function*, not including things such as $\sqrt{z}$. But this need not concern us, as we shall not really be using multi-valued functions in the course.

The Riemann surface $S = \{(z, w) \in \mathbb{C}^2 \mid z = w^2\}$ is identified with the complex w-plane by projection. It is then clear what a holomorphic function on S should be: an analytic function of w, regarded as a function on S. We won't be so lucky in general, in the sense that Riemann surfaces will not be identifiable with their w- or z-projections. However, a class of greatest importance for us, that of *non-singular* Riemann surfaces, is defined by the following property:

1.5 Moral definition: A Riemann surface S in $\mathbb{C}^2$ is *non-singular* if each point (z_0, w_0) has the property that

- either the projection to the z-plane
- or the projection to the w-plane
- or both

can be used to identify a neighbourhood of (z_0, w_0) on S homeomorphically with a disc in the z-plane around z_0, or with a disc in the w-plane around w_0.

We can then use this identification to define what it means for a function on S to be holomorphic near (z_0, w_0).

1.6 Remark. We allowed concrete Riemann surfaces to be singular. In the literature, that is usually disallowed (and our singular Riemann surfaces are called *analytic sets*). We are mostly concerned with non-sigular surfaces, so this will not cause trouble.

An interesting example

Let us conclude the lecture with an example of a Riemann surface with an interesting shape, which cannot be identified by projection (or in any other way) with the z-plane or the w-plane.

Start with the function $w = \sqrt{(z^2 - 1)(z^2 - k^2)}$ where $k \in \mathbb{C}$, $k \neq \pm 1$, whose graph is the Riemann surface

$$T = \{(z, w) \in \mathbb{C}^2 \mid w^2 = (z^2 - 1)(z^2 - k^2)\}.$$

There are *two* values for w for every value of z, other than $z = \pm 1$ and $z = \pm k$, in which cases $w = 0$. A real snapshot of the graph (when $k \in \mathbb{R}$) is indicated in Fig. (1.3), where the dotted lines indicate that the values are imaginary.

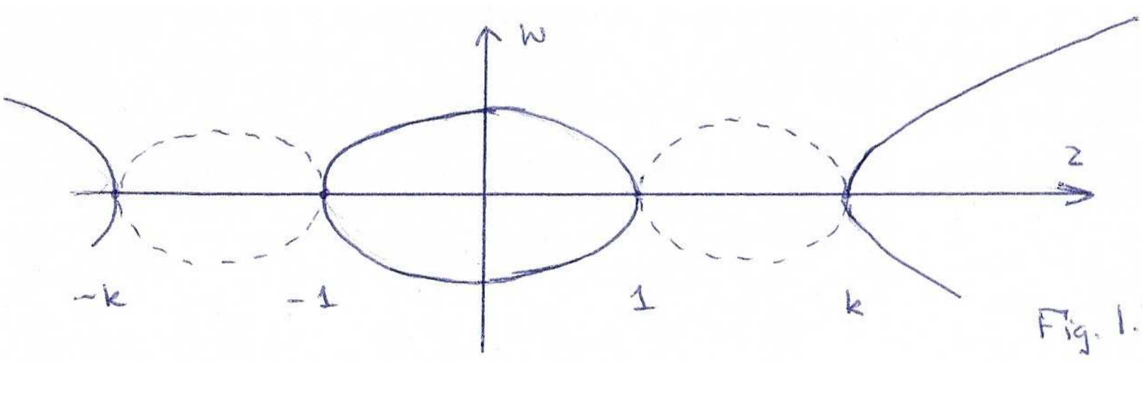


Fig. 1.3

Near $z = 1$, $z = 1 + \epsilon$ and the function is expressible as

$$w = \sqrt{\epsilon(2+\epsilon)(1+\epsilon+k)(1+\epsilon-k)} = \sqrt{\epsilon}\sqrt{2+\epsilon}\sqrt{(1+k)+\epsilon}\sqrt{(1-k)+\epsilon}.$$

A choice of sign for $\sqrt{2(1+k)(1-k)}$ leads to a holomorphic function $\sqrt{2+\epsilon}\sqrt{(1+k)+\epsilon}\sqrt{(1-k)+\epsilon}$ for small ϵ, so $w = \sqrt{\epsilon}\ \times$ (a holomorphic function of ϵ), and the qualitative behaviour of the function near $w = 1$ is like that of $\sqrt{\epsilon} = \sqrt{z-1}$.

Similarly, w behaves like the square root near -1, $\pm k$. The important thing is that there is no continuous single-valued choice of w near these points: any choice of w, followed continuously round any of the four points, leads to the opposite choice upon return.

Defining a continuous branch for the function necessitates some cuts. The simplest way is to remove the open line segments joining 1 with k and -1 with $-k$. On the complement of these segments, we can make a continuous choice of w, which gives an analytic function (for $z \neq \pm 1, \pm k$). The other 'branch' of the graph is obtained by a global change of sign.

Thus, ignoring the cut intervals for a moment, the graph of w breaks up into two pieces, each of which can be identified, via projection, with the z-plane minus two intervals (Fig. 1.4).

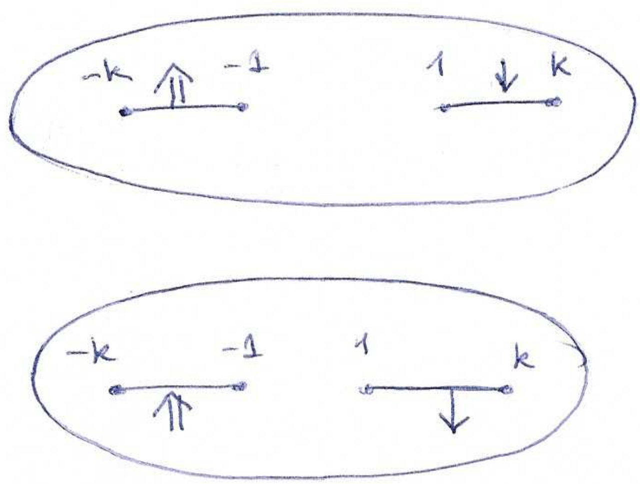


Fig. 1.4

Now over the said intervals, the function also takes two values, except at the endpoints where those coincide. To understand how to assemble the two branches of the graph, recall that the value of w jumps to its negative as we cross the cuts. Thus, if we start on the upper sheet and travel that route, we find ourselves exiting on the lower sheet. Thus,

- the far edges of the cuts on the top sheet must be identified with the near edges of the cuts on the lower sheet;
- the near edges of the cuts on the top sheet must be identified with the far edges on the lower sheet;
- matching endpoints are identified;
- there are no other identifications.

A moment's thought will convince us that we cannot do all this in $\mathbb{R}^3$, with the sheets positioned as depicted, without introducing spurious crossings. To rescue something, we flip the bottom sheet about the real axis. The matching edges of the cuts are now aligned, and we can perform the identifications by stretching each of the surfaces around the cut to pull out a tube. We obtain the following picture, representing two planes (ignore the boundaries) joined by two tubes (Fig. 1.5.a).

For another look at this surface, recall that the function

$$z \mapsto R^2/z$$

identifies the exterior of the circle $|z| \leq R$ with the punctured disc $\{|z| < R \mid z \neq 0\}$. (This identification is even bi-holomorphic, but we don't care about this yet.) Using that, we can pull the exteriors of the discs, missing from the picture above, into the picture as punctured discs, and obtain a torus with two missing points as the definitive form of our Riemann surface (Fig. 1.5.b).

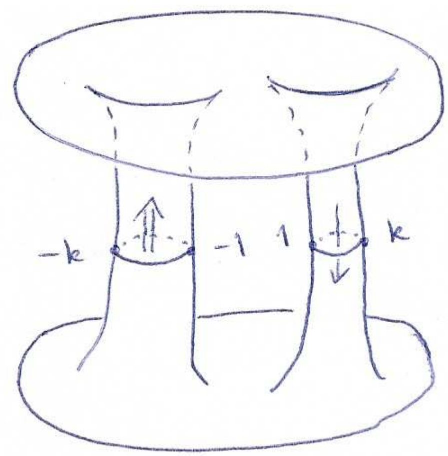


Fig 1.5.a

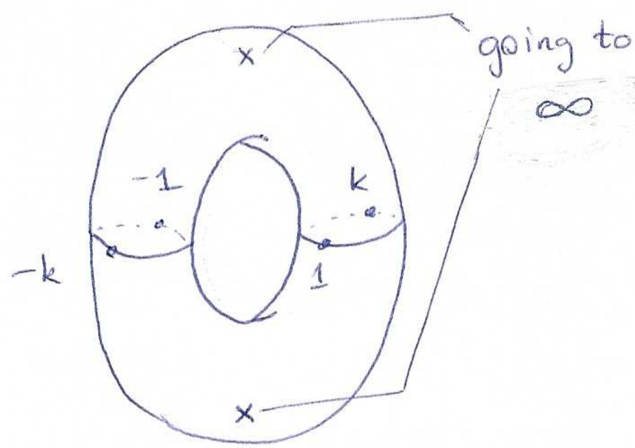


Fig 1.5.b

Lecture 2

The example considered at the end of the Lecture 1 raises the first serious questions for the course, which we plan to address once we define things properly: What shape can a Riemann surface have? And, how can we tell the topological shape of a Riemann surface, other than by creative cutting and pasting?

The answer to the first question (which will need some qualification) is that any orientable surface can be given the structure of a Riemann surface. One answer to the second question, at least for a large class of surfaces, will be the *Riemann-Hurwitz theorem* (Lecture 6).

2.1 Remark. Recall that a surface is *orientable* if there is a continuous choice of clockwise rotations on it. (A non-orientable surface is the Möbius strip; a compact example without boundary is the Klein bottle.) Orientability of Riemann surfaces will follow from our desire to do complex analysis on them; notice that the complex plane carries a natural orientation, in which multiplication by i is counter-clockwise rotation.

Concrete Riemann Surfaces

Historically, Riemann surfaces arose as graphs of analytic functions, with multiple values, defined over domains in $\mathbb{C}$. Inspired by this, we now give a precise definition of a concrete Riemann surface; but we need a preliminary notion.

2.4 The Gamma Function

1

THE GAMMA FUNCTION

1.1. Definition of the Gamma Function

One of the simplest and most important special functions is the *gamma function*, knowledge of whose properties is a prerequisite for the study of many other special functions, notably the cylinder functions and the hypergeometric function. Since the gamma function is usually studied in courses on complex variable theory, and even in advanced calculus,[1] the treatment given here will be deliberately brief.

The gamma function is defined by the formula

$$\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}\,dt, \qquad \operatorname{Re} z > 0, \tag{1.1.1}$$

whenever the complex variable z has a positive real part $\operatorname{Re} z$. We can write (1.1.1) as a sum of two integrals, i.e.,

$$\Gamma(z) = \int_0^1 e^{-t}t^{z-1}\,dt + \int_1^\infty e^{-t}t^{z-1}\,dt, \tag{1.1.2}$$

where it can easily be shown[2] that the first integral defines a function $P(z)$

[1] See D. V. Widder, *Advanced Calculus*, second edition, Prentice-Hall, Inc., Englewood Cliffs, N.J. (1961), Chap. 11.

[2] See E. C. Titchmarsh, *The Theory of Functions*, second edition, Oxford University Press, London (1939), p. 100, noting that the integrand $e^{-t}t^{z-1}$ is analytic in z and continuous in z and t for $\operatorname{Re} z > 0$, $0 < t < \infty$, while the first integral is uniformly convergent for $\operatorname{Re} z \geqslant \delta > 0$ and the second integral is uniformly convergent for $\operatorname{Re} z \leqslant A < \infty$, since then

$$\left|\int_0^1 e^{-t}t^{z-1}\,dt\right| \leqslant \int_0^1 e^{-t}t^{\delta-1}\,dt < \infty, \qquad \left|\int_1^\infty e^{-t}t^{z-1}\,dt\right| \leqslant \int_1^\infty e^{-t}t^{A-1}\,dt < \infty.$$

which is analytic in the half-plane $\operatorname{Re} z > 0$, while the second integral defines an entire function. It follows that the function $\Gamma(z) = P(z) + Q(z)$ is analytic in the half-plane $\operatorname{Re} z > 0$.

The values of $\Gamma(z)$ in the rest of the complex plane can be found by analytic continuation of the function defined by (1.1.1). First we replace the exponential in the integral for $P(z)$ by its power series expansion, and then we integrate term by term, obtaining

$$\begin{aligned} P(z) &= \int_0^1 t^{z-1}\,dt \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} t^k = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_0^1 t^{k+z-1}\,dt \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{1}{z+k}, \end{aligned} \tag{1.1.3}$$

where it is permissible to reverse the order of integration and summation since[3]

$$\int_0^1 |t^{z-1}|\,dt \sum_{k=0}^{\infty} \left| \frac{(-1)^k}{k!} t^k \right| = \int_0^1 t^{x-1}\,dt \sum_{k=0}^{\infty} \frac{t^k}{k!} = \int_0^1 e^t t^{x-1}\,dt < \infty$$

(the last integral converges for $x = \operatorname{Re} z > 0$). The terms of the series (1.1.3) are analytic functions of z, if $z \neq 0, -1, -2, \ldots$ Moreover, in the region[4]

$$|z + k| \geqslant \delta > 0, \qquad k = 0, 1, 2, \ldots,$$

(1.1.3) is majorized by the convergent series

$$\sum_{k=0}^{\infty} \frac{1}{k!\,\delta},$$

and hence is uniformly convergent in this region. Using Weierstrass' theorem[5] and the arbitrariness of δ, we conclude that the sum of the series (1.1.3) is a meromorphic function with simple poles at the points $z = 0, -1, -2, \ldots$ For $\operatorname{Re} z > 0$ this function coincides with the integral $P(z)$, and hence is the analytic continuation of $P(z)$.

The function $\Gamma(z)$ differs from $P(z)$ by the term $Q(z)$, which, as just shown, is an entire function. Therefore $\Gamma(z)$ is a meromorphic function of the complex variable z, with simple poles at the points $z = 0, -1, -2, \ldots$ An

---

[3] E. C. Titchmarsh, *op. cit.*, p. 45.

[4] By a *region* we mean an open connected point set (of two or more dimensions) together with some, all, or possibly none of its boundary points. In the latter case, we often speak of an *open region* or *domain*, in the former case, of a *closed region* or *closed domain*.

[5] See A. I. Markushevich, *Theory of Functions of a Complex Variable, Vol. I* (translated by R. A. Silverman), Prentice-Hall, Inc., Englewood Cliffs, N.J. (1965), Theorem 15.6, p. 326.

analytic expression for $\Gamma(z)$, suitable for defining $\Gamma(z)$ in the whole complex plane, is given by

$$\Gamma(z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{1}{z+k} + \int_1^{\infty} e^{-t} t^{z-1}\, dt, \qquad z \neq 0, -1, -2, \ldots \quad (1.1.4)$$

It follows from (1.1.4) that $\Gamma(z)$ has the representation

$$\Gamma(z) = \frac{(-1)^n}{n!} \frac{1}{z+n} + \Omega(z+n) \qquad (1.1.5)$$

in a neighborhood of the pole $z = -n$ $(n = 0, 1, 2, \ldots)$, with regular part $\Omega(z+n)$.

1.2. Some Relations Satisfied by the Gamma Function

We now prove three basic relations satisfied by the gamma function:

$$\Gamma(z+1) = z\Gamma(z), \qquad (1.2.1)$$

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z}, \qquad (1.2.2)$$

$$2^{2z-1}\Gamma(z)\Gamma(z+\tfrac{1}{2}) = \sqrt{\pi}\,\Gamma(2z). \qquad (1.2.3)$$

These formulas play an important role in various transformations and calculations involving $\Gamma(z)$.

To prove (1.2.1), we assume that $\operatorname{Re} z > 0$ and use the integral representation (1.1.1). An integration by parts gives

$$\Gamma(z+1) = \int_0^{\infty} e^{-t} t^z\, dt = -e^{-t} t^z \Big|_0^{\infty} + z \int_0^{\infty} e^{-t} t^{z-1}\, dt = z\Gamma(z)$$

The validity of this result for arbitrary complex $z \neq 0, -1, -2, \ldots$ is an immediate consequence of the principle of analytic continuation,[6] since both sides of the formula are analytic everywhere except at the points $z = 0, -1, -2, \ldots$

To derive (1.2.2), we temporarily assume that $0 < \operatorname{Re} z < 1$ and again use (1.1.1), obtaining

$$\Gamma(z)\Gamma(1-z) = \int_0^{\infty} \int_0^{\infty} e^{-(s+t)} s^{-z} t^{z-1}\, ds\, dt.$$

[6] According to this principle, which we will use repeatedly, if $f(z)$ and $\varphi(z)$ are analytic in a domain D and if $f(z) = \varphi(z)$ for all z in a smaller domain D^* contained in D, then $f(z) = \varphi(z)$ for all z in D. The same is true if $f(z) = \varphi(z)$ for all z in any set of points of D with a limit point in D, say, a line segment. See A. I. Markushevich, *op. cit.*, Theorem 17.1, p. 369.

Introducing the new variables

$$u = s + t, \qquad v = \frac{t}{s},$$

we find that[7]

$$\Gamma(z)\Gamma(1 - z) = \int_0^\infty \int_0^\infty e^{-u}v^{z-1}\frac{du\,dv}{1 + v} = \int_0^\infty \frac{v^{z-1}}{1 + v}\,dv = \frac{\pi}{\sin \pi z}.$$

Using the principle of analytic continuation, we see that this formula remains valid everywhere in the complex plane except at the points $z = 0, \pm 1, \pm 2, \ldots$

To prove (1.2.3), known as the *duplication formula*, we assume that $\operatorname{Re} z > 0$ and then use (1.1.1) again, obtaining

$$\begin{aligned} 2^{2z-1}\Gamma(z)\Gamma(z + \tfrac{1}{2}) &= \int_0^\infty \int_0^\infty e^{-(s+t)}(2\sqrt{st})^{2z-1}t^{-1/2}\,ds\,dt \\ &= 4\int_0^\infty \int_0^\infty e^{-(\alpha^2+\beta^2)}(2\alpha\beta)^{2z-1}\alpha\,d\alpha\,d\beta, \end{aligned}$$

where we have introduced new variables $\alpha = \sqrt{s}$, $\beta = \sqrt{t}$. To this formula we add the similar formula obtained by permuting α and β. This gives the more symmetric representation

$$\begin{aligned} 2^{2z-1}\Gamma(z)\Gamma(z + \tfrac{1}{2}) &= 2\int_0^\infty \int_0^\infty e^{-(\alpha^2+\beta^2)}(2\alpha\beta)^{2z-1}(\alpha + \beta)\,d\alpha\,d\beta \\ &= 4\iint_\sigma e^{-(\alpha^2+\beta^2)}(2\alpha\beta)^{2z-1}(\alpha + \beta)\,d\alpha\,d\beta, \end{aligned}$$

where the last integral is over the sector $\sigma: 0 \leqslant \alpha < \infty, 0 \leqslant \beta \leqslant \alpha$. Introducing new variables

$$u = \alpha^2 + \beta^2, \qquad v = 2\alpha\beta,$$

we find that

$$\begin{aligned} 2^{2z-1}\Gamma(z)\Gamma(z + \tfrac{1}{2}) &= \int_0^\infty v^{2z-1}\,dv \int_0^\infty \frac{e^{-u}}{\sqrt{u - v}}\,du \\ &= 2\int_0^\infty e^{-v}v^{2z-1}\,dv \int_0^\infty e^{-w^2}\,dw = \sqrt{\pi}\,\Gamma(2z). \end{aligned}$$

As before, this result can be extended to arbitrary complex values $z \neq 0, -\frac{1}{2}, -1, -\frac{3}{2}, \ldots$, by using the principle of analytic continuation.

We now use formula (1.2.1) to calculate $\Gamma(z)$ for some special values of the variable z. Applying (1.2.1) and noting that $\Gamma(1) = 1$, we find by mathematical induction that

$$\Gamma(n + 1) = n!, \qquad n = 0, 1, 2, \ldots \tag{1.2.4}$$

---

[7] For the evaluation of the integral in the last step, see E. C. Titchmarsh, *op. cit.*, p. 105.

Moreover, setting $z = \frac{1}{2}$ in (1.1.1), we obtain

$$\Gamma(\tfrac{1}{2}) = \int_0^\infty e^{-t}t^{-1/2}\,dt = 2\int_0^\infty e^{-u^2}\,du = \sqrt{\pi}, \tag{1.2.5}$$

and then (1.2.1) implies

$$\Gamma(n + \tfrac{1}{2}) = \frac{1\cdot3\cdot5\cdots(2n-1)}{2^n}\sqrt{\pi}, \qquad n = 1, 2, \ldots \tag{1.2.6}$$

Finally we use (1.2.2) to prove that the function $\Gamma(z)$ has no zeros in the complex plane. First we note that the points $z = n$ $(n = 0, \pm1, \pm2, \ldots)$ cannot be zeros of $\Gamma(z)$, since $\Gamma(n) = (n-1)!$ if $n = 1, 2, \ldots,$ while $\Gamma(n) = \infty$ if $n = 0, -1, -2, \ldots$ The fact that no other value of z can be a zero of $\Gamma(z)$ is an immediate consequence of (1.2.2), since if a nonintegral value of z were a zero of $\Gamma(z)$ it would have to be a pole of $\Gamma(1-z)$, which is impossible. It follows at once that $[\Gamma(z)]^{-1}$ *is an entire function.*

1.3. The Logarithmic Derivative of the Gamma Function

The theory of the gamma function is intimately related to the theory of another special function, i.e., the logarithmic derivative of $\Gamma(z)$:

$$\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}. \tag{1.3.1}$$

Since $\Gamma(z)$ is a meromorphic function with no zeros, $\psi(z)$ can have no singular points other than the poles $z = -n$ $(n = 0, 1, 2, \ldots)$ of $\Gamma(z)$. It follows from (1.1.5) that $\psi(z)$ has the representation[8]

$$\psi(z) = -\frac{1}{z+n} + \Omega(z+n) \tag{1.3.2}$$

in a neighborhood of the point $z = -n$, and hence $\psi(z)$, like $\Gamma(z)$, is a meromorphic function with simple poles at the points $z = 0, -1, -2, \ldots$

The function $\psi(z)$ satisfies relations obtained from formulas (1.2.1–3)[9] by taking logarithmic derivatives. In this way, we find that

$$\psi(z+1) = \frac{1}{z} + \psi(z), \tag{1.3.3}$$

$$\psi(1-z) - \psi(z) = \pi\cot\pi z, \tag{1.3.4}$$

$$\psi(z) + \psi(z+\tfrac{1}{2}) + 2\log 2 = 2\psi(2z). \tag{1.3.5}$$

---

[8] Of course, the regular part $\Omega(z+n)$ in (1.3.2) is not the same as in (1.1.5).

[9] By (1.2.1–3) we mean formulas (1.2.1) through (1.2.3). Similarly, (1.2.1, 4, 6) means formulas (1.2.1), (1.2.4) and (1.2.6), etc.

These formulas can be used to calculate $\psi(z)$ for special values of z. For example, writing

$$\psi(1) = \Gamma'(1) = -\gamma, \tag{1.3.6}$$

where $\gamma = 0.57721566\ldots$ is *Euler's constant*, and using (1.3.3), we obtain

$$\psi(n+1) = -\gamma + \sum_{k=1}^{n} \frac{1}{k}, \qquad n = 1, 2, \ldots \tag{1.3.7}$$

Moreover, substituting $z = \frac{1}{2}$ into (1.3.5), we find that

$$\psi(\tfrac{1}{2}) = -\gamma - 2\log 2, \tag{1.3.8}$$

and then (1.3.3) gives

$$\psi(n+\tfrac{1}{2}) = -\gamma - 2\log 2 + 2\sum_{k=1}^{n} \frac{1}{2k-1}, \qquad n = 1, 2, \ldots \tag{1.3.9}$$

The function $\psi(z)$ has simple representations in the form of definite integrals involving the variable z as a parameter. To derive these representations, we first note that (1.1.1) implies[10]

$$\Gamma'(z) = \int_0^\infty e^{-t} t^{z-1} \log t \, dt, \qquad \operatorname{Re} z > 0. \tag{1.3.10}$$

If we replace the logarithm in the integrand by its expression in terms of the *Frullani integral*[11]

$$\log t = \int_0^\infty \frac{e^{-x} - e^{-xt}}{x} \, dx, \qquad \operatorname{Re} t > 0, \tag{1.3.11}$$

we find that[12]

$$\begin{aligned}\Gamma'(z) &= \int_0^\infty \frac{dx}{x} \int_0^\infty (e^{-x} - e^{-xt}) e^{-t} t^{z-1} \, dt \\ &= \int_0^\infty \frac{dx}{x} \left[e^{-x}\Gamma(z) - \int_0^\infty e^{-t(x+1)} t^{z-1} \, dt \right].\end{aligned}$$

Introducing the new variable of integration $u = t(x+1)$, we find that the integral in brackets equals $(x+1)^{-z}\Gamma(z)$. This leads to the following integral representation of $\psi(z)$:

$$\psi(z) = \int_0^\infty \left[e^{-x} - \frac{1}{(x+1)^z} \right] \frac{dx}{x}, \qquad \operatorname{Re} z > 0. \tag{1.3.12}$$

---

[10] To justify differentiating behind the integral sign, see E. C. Titchmarsh, *op. cit.*, pp. 99–100.

[11] See H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, third edition, Cambridge University Press, London (1956), p. 406, and D. V. Widder, *op. cit.*, p. 357.

[12] Here, as elsewhere in this chapter, we omit detailed justification of the reversal of order of integration. An appropriate argument can always be supplied, usually by proving the absolute convergence of the double integral and then using *Fubini's theorem*. See H. Kestelman, *Modern Theories of Integration*, second revised edition, Dover Publications, Inc., New York (1960), Chap. 8, esp. Theorems 279 and 280.

To obtain another integral representation of $\psi(z)$, we write (1.3.12) in the form

$$\psi(z) = \lim_{\delta\to 0}\int_\delta^\infty \left[e^{-x} - \frac{1}{(x+1)^z}\right]\frac{dx}{x} = \lim_{\delta\to 0}\left[\int_\delta^\infty \frac{e^{-x}}{x}\,dx - \int_\delta^\infty \frac{dx}{(x+1)^z x}\right],$$

and change the variable of integration in the second integral, by setting $x + 1 = e^t$. This gives

$$\begin{aligned}\psi(z) &= \lim_{\delta\to 0}\left[\int_\delta^\infty \frac{e^{-t}}{t}\,dt - \int_{\log(1+\delta)}^\infty \frac{e^{-tz}}{1-e^{-t}}\,dt\right]\\ &= \lim_{\delta\to 0}\left[\int_{\log(1+\delta)}^\infty \left(\frac{e^{-t}}{t} - \frac{e^{-tz}}{1-e^{-t}}\right)dt - \int_{\log(1+\delta)}^\delta \frac{e^{-t}}{t}\,dt\right],\end{aligned}$$

and therefore, since the second integral approaches zero as $\delta \to 0$,

$$\psi(z) = \int_0^\infty \left(\frac{e^{-t}}{t} - \frac{e^{-tz}}{1-e^{-t}}\right)dt, \qquad \operatorname{Re} z > 0. \tag{1.3.13}$$

Setting $z = 1$ and subtracting the result from (1.3.13), we find that

$$\psi(z) = -\gamma + \int_0^\infty \frac{e^{-t} - e^{-tz}}{1-e^{-t}}\,dt, \qquad \operatorname{Re} z > 0, \tag{1.3.14}$$

or

$$\psi(z) = -\gamma + \int_0^1 \frac{1-x^{z-1}}{1-x}\,dx, \qquad \operatorname{Re} z > 0, \tag{1.3.15}$$

where we have introduced the variable of integration $x = e^{-t}$.

From formula (1.3.15) we can deduce an important representation of $\psi(z)$ as an analytic expression valid for all $z \neq 0, -1, -2, \ldots$, i.e., in the whole domain of definition of $\psi(z)$. To obtain this representation, we substitute the power series expansion

$$(1-x)^{-1} = 1 + x + x^2 + \cdots + x^n + \cdots, \qquad 0 \leqslant x < 1$$

into (1.3.15) and integrate term by term (this operation is easily justified). The result is

$$\psi(z) = -\gamma + \sum_{n=0}^\infty \left(\frac{1}{n+1} - \frac{1}{n+z}\right). \tag{1.3.16}$$

The series (1.3.16), whose terms are analytic functions for $z \neq 0, -1, -2, \ldots$, is uniformly convergent in the region defined by the inequalities

$$|z+n| \geqslant \delta > 0, \quad n = 0, 1, 2, \ldots \quad \text{and} \quad |z| < a,$$

since

$$\left|\frac{1}{n+1} - \frac{1}{n+z}\right| < \frac{a+1}{(n+1)(n-a)}$$

for $n \geqslant N > a$, and the series

$$\sum_{n=N}^{\infty} \frac{a+1}{(n+1)(n-a)}$$

converges. Therefore, since δ is arbitrarily small and a arbitrarily large, both sides of (1.3.16) are analytic functions except at the poles $z = 0, -1, -2, \ldots$, and hence, according to the principle of analytic continuation, the original restriction $\operatorname{Re} z > 0$ used to prove this formula can be dropped. If we replace z by $z + 1$ in (1.3.16), integrate the resulting series between the limits 0 and z, and then take exponentials of both sides, we find the following infinite product representation of the gamma function:

$$\frac{1}{\Gamma(z+1)} = e^{\gamma z} \prod_{n=1}^{\infty} e^{-z/n}\left(1 + \frac{z}{n}\right). \tag{1.3.17}$$

This formula can be made the starting point for the theory of the gamma function, instead of the integral representation (1.1.1).

Finally we derive some formulas for Euler's constant γ. Setting $z = 1$ in (1.3.12–13), we obtain

$$\gamma = -\psi(1) = \int_0^{\infty} \left(\frac{1}{1+x} - e^{-x}\right) \frac{dx}{x} = \int_0^{\infty} \left(\frac{1}{1-e^{-t}} - \frac{1}{t}\right) e^{-t}\, dt. \tag{1.3.18}$$

Moreover, (1.3.10) implies

$$\gamma = -\int_0^{\infty} e^{-t} \log t\, dt, \tag{1.3.19}$$

which, when integrated by parts, gives

$$\gamma = \int_0^1 \log t\, d(e^{-t} - 1) + \int_1^{\infty} \log t\, d(e^{-t}) = \int_0^1 \frac{1 - e^{-t}}{t}\, dt - \int_1^{\infty} \frac{e^{-t}}{t}\, dt.$$

Replacing t by $1/t$ in the last integral on the right, we find that

$$\gamma = \int_0^1 \frac{1 - e^{-t} - e^{-1/t}}{t}\, dt. \tag{1.3.20}$$

1.4. Asymptotic Representation of the Gamma Function for Large $|z|$

To describe the behavior of a given function $f(z)$ as $|z| \to \infty$ within a sector $\alpha \leqslant \arg z \leqslant \beta$, it is in many cases sufficient to derive an expression of the form

$$f(z) = \varphi(z)[1 + r(z)], \tag{1.4.1}$$

where $\varphi(z)$ is a function of a simpler structure than $f(z)$, and $r(z)$ converges uniformly to zero as $|z| \to \infty$ within the given sector. Formulas of this type are called *asymptotic representations* of $f(z)$ for large $|z|$. It follows from

(1.4.1) that the ratio $f(z)/\varphi(z)$ converges to unity as $|z| \to \infty$, i.e., the two functions $f(z)$ and $\varphi(z)$ are "asymptotically equal," a fact we indicate by writing

$$f(z) \approx \varphi(z), \qquad |z| \to \infty, \quad \alpha \leqslant \arg z \leqslant \beta. \tag{1.4.2}$$

An estimate of $|r(z)|$ gives the size of the error committed when $f(z)$ is replaced by $\varphi(z)$ for large but finite $|z|$.

We now look for a description of the behavior of the function $f(z)$ as $|z| \to \infty$ which is more exact than that given by (1.4.1). Suppose we succeed in deriving the formula

$$f(z) = \varphi(z)\left[\sum_{n=0}^{N} a_n z^{-n} + r_N(z)\right], \qquad a_0 = 1, \quad N = 1, 2, \ldots, \tag{1.4.3}$$

where $z^N r_N(z)$ converges uniformly to zero as $|z| \to \infty$, $\alpha \leqslant \arg z \leqslant \beta$. [Note that (1.4.3) reduces to (1.4.1) for $N = 0$.] Then we write

$$f(z) \approx \varphi(z) \sum_{n=0}^{\infty} a_n z^{-n}, \qquad |z| \to \infty, \quad \alpha \leqslant \arg z \leqslant \beta, \tag{1.4.4}$$

and the right-hand side is called an *asymptotic series* or *asymptotic expansion* of $f(z)$ for large $|z|$. It should be noted that this definition does not stipulate that the given series converge in the ordinary sense, and on the contrary, the series will usually diverge. Nevertheless, asymptotic series are very useful, since, by taking a finite number of terms, we can obtain an arbitrarily good approximation to the function $f(z)$ for sufficiently large $|z|$. In this book, the reader will find many examples of asymptotic representations and asymptotic series (see Secs. 1.4, 2.2, 3.2, 4.6, 4.14, 4.22, 5.11, etc.). For the general theory of asymptotic series, we refer to the references cited in the Bibliography on p. 300.

To obtain an asymptotic representation of the gamma function $\Gamma(z)$, it is convenient to first derive an asymptotic representation of $\log \Gamma(z)$. To this end, let $\operatorname{Re} z > 0$, and consider the integral representation (1.3.13), with z replaced by $z + 1$, i.e.,

$$\begin{aligned}\psi(z+1) = \frac{\Gamma'(z+1)}{\Gamma(z+1)} &= \int_0^\infty \left(\frac{e^{-t}}{t} - \frac{e^{-tz}}{e^t - 1}\right) dt \\ &= \int_0^\infty \frac{e^{-t} - e^{-tz}}{t}\, dt + \frac{1}{2}\int_0^\infty e^{-tz}\, dt - \int_0^\infty \left(\frac{1}{2} - \frac{1}{t} + \frac{1}{e^t - 1}\right) e^{-tz}\, dt,\end{aligned}$$

or

$$\frac{\Gamma'(z+1)}{\Gamma(z+1)} = \log z + \frac{1}{2z} - \int_0^\infty \left(\frac{1}{2} - \frac{1}{t} + \frac{1}{e^t - 1}\right) e^{-tz}\, dt,$$

where we have used (1.3.11). Integrating the last equation between the limits 1 and z, and bearing in mind that

$$\log \Gamma(z+1) = \log \Gamma(z) + \log z,$$

we find that[13]

$$\log \Gamma(z) = \left(z - \frac{1}{2}\right) \log z - z + 1 + \int_0^\infty \left(\frac{1}{2} - \frac{1}{t} + \frac{1}{e^t - 1}\right) \frac{e^{-tz} - e^{-t}}{t}\, dt, \tag{1.4.5}$$

where $\operatorname{Re} z > 0$. It should be noted that the function

$$f(t) = \left(\frac{1}{2} - \frac{1}{t} + \frac{1}{e^t - 1}\right) \frac{1}{t}, \tag{1.4.6}$$

appearing in the integrand in (1.4.5), is continuous for $t \geqslant 0$, with $f(0) = \frac{1}{12}$, as can easily be verified by expanding $f(t)$ in a power series in a neighborhood of the point $t = 0$.

To simplify (1.4.5), we evaluate the integral

$$\mathscr{I} = \int_0^\infty f(t) e^{-t}\, dt. \tag{1.4.7}$$

This can be done by using the following trick: If

$$\mathscr{J} = \int_0^\infty f(t) e^{-t/2}\, dt, \tag{1.4.8}$$

then

$$\mathscr{J} - \mathscr{I} = \int_0^\infty e^{-t/2}\left[f(t) - \frac{1}{2} f\left(\frac{t}{2}\right)\right] dt = \int_0^\infty \left(\frac{e^{-t/2}}{t} - \frac{1}{e^t - 1}\right) \frac{dt}{t}.$$

It follows that

$$\begin{aligned} \mathscr{J} = (\mathscr{J} - \mathscr{I}) + \mathscr{I} &= \int_0^\infty \left(\frac{e^{-t/2} - e^{-t}}{t} - \frac{e^{-t}}{2}\right) \frac{dt}{t} \\ &= \int_0^\infty \left[-\frac{d}{dt}\left(\frac{e^{-t/2} - e^{-t}}{t}\right) + \frac{e^{-t} - e^{-t/2}}{2t}\right] dt \\ &= -\left.\frac{e^{-t/2} - e^{-t}}{t}\right|_0^\infty + \frac{1}{2}\int_0^\infty \frac{e^{-t} - e^{-t/2}}{t}\, dt = \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}. \end{aligned} \tag{1.4.9}$$

On the other hand, substituting $z = \frac{1}{2}$ into (1.4.5), we find that

$$\mathscr{J} - \mathscr{I} = \tfrac{1}{2} \log \pi - \tfrac{1}{2}, \tag{1.4.10}$$

and hence

$$\mathscr{I} = 1 - \tfrac{1}{2} \log 2\pi. \tag{1.4.11}$$

Using this result, we can write (1.4.5) in the form

$$\log \Gamma(z) = (z - \tfrac{1}{2}) \log z - z + \tfrac{1}{2} \log 2\pi + \omega(z), \tag{1.4.12}$$

---

[13] The choice of the path of integration is unimportant. To justify integration behind the integral sign, we use an absolute convergence argument (cf. footnote 12, p. 6).

where

$$\omega(z) = \int_0^\infty f(t)e^{-tz}\,dt, \qquad \operatorname{Re} z > 0. \tag{1.4.13}$$

Since $f(t)$ decreases monotonically as t increases,[14] the integral (1.4.13) also converges for $\operatorname{Re} z = 0$, $\operatorname{Im} z \neq 0$.[15]

Using (1.4.12) and (1.4.13), we can easily derive an asymptotic representation of $\Gamma(z)$. First let $|\arg z| \leqslant \pi/2$, and integrate (1.4.13) by parts, obtaining

$$\omega(z) = \frac{1}{z}\left[f(0) + \int_0^\infty f'(t)e^{-tz}\,dt\right]. \tag{1.4.14}$$

Since $f'(t) \leqslant 0$, $|f'(t)| = -f'(t)$, we have

$$|\omega(z)| \leqslant \frac{1}{|z|}\left[f(0) - \int_0^\infty f'(t)\,dt\right] = \frac{2f(0)}{|z|},$$

i.e.,

$$|\omega(z)| \leqslant \frac{1}{6|z|}, \qquad |\arg z| \leqslant \frac{\pi}{2}. \tag{1.4.15}$$

Then, taking exponentials of both sides of (1.4.12), we find that

$$\Gamma(z) = e^{(z-\frac{1}{2})\log z - z + \frac{1}{2}\log 2\pi}[1 + r(z)], \qquad |\arg z| \leqslant \frac{\pi}{2}, \tag{1.4.16}$$

where

$$r(z) = e^{\omega(z)} - 1.$$

According to (1.4.15),

$$|r(z)| \leqslant \frac{C}{|z|}, \tag{1.4.17}$$

where C is an absolute constant (we assume that z is bounded away from zero, i.e., $|z| \geqslant a > 0$). Thus $r(z)$ is of order $|z|^{-1}$ as $|z| \to \infty$, a fact indicated by writing[16]

$$r(z) = O(|z|^{-1}), \tag{1.4.18}$$

and hence (1.4.16) is an asymptotic representation of $\Gamma(z)$ in the indicated sector.

To derive an asymptotic representation of $\Gamma(z)$ which is valid in other

[14] This follows at once from the expansion

$$f(t) = 2\sum_{k=1}^\infty \frac{1}{t^2 + 4\pi^2k^2}.$$

See K. Knopp, *Theory and Applications of Infinite Series* (translated by R. C. H. Young), Blackie and Son, Ltd., London (1963), p. 378.

[15] E. C. Titchmarsh, *op. cit.*, p. 21.

[16] We say that $f(z)$ *is of order* $\varphi(z)$ as $z \to z_0$, and write $f(z) = O(\varphi(z))$ as $z \to z_0$ if the inequality $|f(z)| \leqslant A|\varphi(z)|$ holds in a neighborhood of z_0, where A is some constant. If z_0 is not explicitly mentioned, then $z_0 = \infty$.

sectors of the complex plane, we proceed as follows: Let δ be an arbitrarily small fixed positive number, and let

$$\frac{\pi}{2} \leqslant \arg z \leqslant \pi - \delta. \tag{1.4.19}$$

Since $\arg(-z) = \arg z - \pi$, this implies

$$-\frac{\pi}{2} \leqslant \arg(-z) \leqslant -\delta.$$

It follows from (1.2.1–2) that

$$\Gamma(z) = \frac{\pi}{-z\Gamma(-z)\sin \pi z}, \tag{1.4.20}$$

where, according to (1.4.16) and (1.4.18),

$$\Gamma(-z) = e^{-(z+\frac{1}{2})(\log z - \pi i) + z + \frac{1}{2}\log 2\pi}[1 + O(|z|^{-1})]. \tag{1.4.21}$$

On the other hand, in the sector (1.4.19),

$$\begin{aligned} \sin \pi z &= \frac{e^{\pi i z} - e^{-\pi i z}}{2i} = -\frac{e^{-\pi i z}}{2i}(1 - e^{2\pi i z}) \\ &= -\frac{e^{-\pi i z}}{2i}\left(1 - \frac{1}{z} z e^{2\pi i z}\right) = -\frac{e^{-\pi i z}}{2i}[1 + O(|z|^{-1})], \end{aligned} \tag{1.4.22}$$

since $ze^{2\pi i z}$ is bounded in this sector. Substituting (1.4.21–22) into (1.4.20), we again arrive at formula (1.4.16). A similar result is obtained for the sector

$$-(\pi - \delta) \leqslant \arg z \leqslant -\frac{\pi}{2}.$$

Finally, therefore, in any sector

$$|\arg z| \leqslant \pi - \delta,$$

we have the asymptotic representation

$$\Gamma(z) = e^{(z-\frac{1}{2})\log z - z + \frac{1}{2}\log 2\pi}[1 + O(|z|^{-1})]. \tag{1.4.23}$$

Considerations resembling those just given, but much more complicated,[17] lead to the more exact formula

$$\Gamma(z) = e^{(z-\frac{1}{2})\log z - z + \frac{1}{2}\log 2\pi}\left[1 + \frac{1}{12z} + \frac{1}{288z^2} - \frac{139}{51840z^3} + O(|z|^{-4})\right]. \tag{1.4.24}$$

If $z = x$ is a positive real number, then (1.4.16) becomes *Stirling's formula*

$$\Gamma(x) = \sqrt{2\pi}\, x^{x-\frac{1}{2}} e^{-x}[1 + r(x)], \tag{1.4.25}$$

[17] See G. N. Watson, *An expansion related to Stirling's formula, derived by the method of steepest descents*, Quart. J. Pure and Appl. Math., **48**, 1 (1920).

where for $r(x)$ we have a sharper estimate than that given by (1.4.17). In fact, if $z = x > 0$, then

$$|\omega(x)| \leqslant f(0) \int_0^\infty e^{-xt}\, dt = \frac{1}{12x}, \tag{1.4.26}$$

so that

$$|r(x)| \leqslant e^{1/12x} - 1. \tag{1.4.27}$$

Finally, we note that (1.2.4) and (1.4.25) imply the following asymptotic representation of the factorial:

$$n! \approx \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}, \qquad n \to \infty. \tag{1.4.28}$$

1.5. Definite Integrals Related to the Gamma Function

The class of integrals which can be expressed in terms of the gamma function is very large. Here we consider only a few examples, mainly with the intent of deriving some formulas that will be needed later.

Our first result is the formula

$$\int_0^\infty e^{-pt} t^{z-1}\, dt = \frac{\Gamma(z)}{p^z}, \qquad \operatorname{Re} p > 0, \quad \operatorname{Re} z > 0, \tag{1.5.1}$$

which is easily proved for positive real p by making the change of variables $s = pt$, and then using the integral representation (1.1.1). The extension of (1.5.1) to arbitrary complex p with $\operatorname{Re} p > 0$ is accomplished by using the principle of analytic continuation.

Next consider the integral

$$B(x, y) = \int_0^1 t^{x-1}(1 - t)^{y-1}\, dt, \qquad \operatorname{Re} x > 0, \quad \operatorname{Re} y > 0, \tag{1.5.2}$$

known as the *beta function*. It is easy to see that (1.5.2) represents an analytic function in each of the complex variables x and y. If we introduce the new variable of integration $u = t/(1 - t)$, then (1.5.2) becomes

$$B(x, y) = \int_0^\infty \frac{u^{x-1}}{(1 + u)^{x+y}}\, du, \qquad \operatorname{Re} x > 0, \quad \operatorname{Re} y > 0. \tag{1.5.3}$$

Setting $p = 1 + u$, $z = x + y$ in (1.5.1), we find that

$$\frac{1}{(1 + u)^{x+y}} = \frac{1}{\Gamma(x + y)} \int_0^\infty e^{-(1+u)t} t^{x+y-1}\, dt, \tag{1.5.4}$$

and substituting the result into (1.5.3), we obtain

$$\begin{aligned} B(x, y) &= \frac{1}{\Gamma(x + y)} \int_0^\infty e^{-t} t^{x+y-1}\, dt \int_0^\infty e^{-ut} u^{x-1}\, du \\ &= \frac{\Gamma(x)}{\Gamma(x + y)} \int_0^\infty e^{-t} t^{y-1}\, dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)}. \end{aligned} \tag{1.5.5}$$

Thus we have derived the formula

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)}, \tag{1.5.6}$$

relating the beta function to the gamma function, which can be used to derive all the properties of the beta function.

PROBLEMS

1. Prove that

$$|\Gamma(iy)|^2 = \frac{\pi}{y \sinh \pi y}, \qquad |\Gamma(\tfrac{1}{2} + iy)|^2 = \frac{\pi}{\cosh \pi y}$$

for real y.

2. Using (1.5.6), verify the identity

$$\int_0^\infty \frac{\cosh 2yt}{(\cosh t)^{2x}}\, dt = 2^{2x-2} \frac{\Gamma(x + y)\Gamma(x - y)}{\Gamma(2x)}, \qquad \operatorname{Re} x > 0, \qquad \operatorname{Re} x > |\operatorname{Re} y|.$$

3. Prove that

$$\int_0^{\pi/2} \cos^\nu \theta \, d\theta = \int_0^{\pi/2} \sin^\nu \theta \, d\theta = \frac{\sqrt{\pi}}{2} \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\Gamma\left(\frac{\nu}{2} + 1\right)}, \qquad \operatorname{Re} \nu > -1,$$

$$\int_0^{\pi/2} \cos^\mu \theta \sin^\nu \theta \, d\theta = \frac{1}{2} \frac{\Gamma\left(\frac{\mu + 1}{2}\right)\Gamma\left(\frac{\nu + 1}{2}\right)}{\Gamma\left(\frac{\mu + \nu}{2} + 1\right)}, \qquad \operatorname{Re} \mu > -1, \qquad \operatorname{Re} \nu > -1.$$

4. Verify the formula

$$\Gamma(3z) = \frac{3^{3z - 1/2}}{2\pi} \Gamma(z)\Gamma(z + \tfrac{1}{3})\Gamma(z + \tfrac{2}{3}). \tag{i}$$

5. Derive the formula

$$3\psi(3z) = \psi(z) + \psi(z + \tfrac{1}{3}) + \psi(z + \tfrac{2}{3}) + 3 \log 3.$$

Hint. Calculate the logarithmic derivatives of both sides of (i).

6. Derive the following integral representation of the square of the gamma function, where $K_0(t)$ is Macdonald's function (defined in Sec. 5.7):

$$\Gamma^2(z) = 2^{2-2z} \int_0^\infty t^{2z-1} K_0(t)\, dt, \qquad \operatorname{Re} z > 0.$$

Hint. Use formulas (5.10.23), (1.5.1) and the integral in Problem 2.

7. Derive the asymptotic formulas

$$\Gamma(z + \alpha) = e^{(z+\alpha-\frac{1}{2})\log z - z + \frac{1}{2}\log 2\pi}[1 + O(|z|^{-1})],$$

$$\frac{\Gamma(z + \alpha)}{\Gamma(z + \beta)} = z^{\alpha-\beta}\left[1 + \frac{(\alpha - \beta)(\alpha + \beta - 1)}{2z} + O(|z|^{-2})\right],$$

where α and β are arbitrary constants, and $|\arg z| \leqslant \pi - \delta$.

Hint. Use the results of Sec. 1.4.

8. Derive the asymptotic formula

$$|\Gamma(x + iy)| = \sqrt{2\pi}\, e^{-\frac{1}{2}\pi|y|}|y|^{x-\frac{1}{2}}[1 + r(x, y)],$$

where as $|t| \to \infty$, $r(x, y) \to 0$ uniformly in the strip $|x| \leqslant \alpha$ (α is a constant).

9. Show that the integral representation

$$\frac{1}{\Gamma(z)} = \frac{1}{2\pi i}\int_C e^t t^{-z}\, dt$$

holds for arbitrary complex z, where $t^{-z} = e^{-z\log t}$, $|\arg t| < \pi$, and C is the contour shown in Figure 13, p. 117.

10. The *incomplete gamma function* $\gamma(z, \alpha)$ and its *complement* $\Gamma(z, \alpha)$ are defined by the formulas

$$\gamma(z, \alpha) = \int_0^\alpha e^{-t}t^{z-1}\, dt, \qquad \operatorname{Re} z > 0, \quad |\arg \alpha| < \pi,$$

$$\Gamma(z, \alpha) = \int_\alpha^\infty e^{-t}t^{z-1}\, dt, \qquad |\arg \alpha| < \pi,$$

so that

$$\gamma(z, \alpha) + \Gamma(z, \alpha) = \Gamma(z).$$

Prove that for fixed α, $\Gamma(z, \alpha)$ is an entire function of z, while $\gamma(z, \alpha)$ is a meromorphic function of z, with poles at the points $z = 0, -1, -2, \ldots$

11. Derive the formulas

$$\gamma(z + 1, \alpha) = z\gamma(z, \alpha) - e^{-\alpha}\alpha^z,$$

$$\Gamma(z + 1, \alpha) = z\Gamma(z, \alpha) + e^{-\alpha}\alpha^z.$$

12. Derive the following representation of $\gamma(z, \alpha)$:

$$\gamma(z, \alpha) = \sum_{k=0}^\infty \frac{(-1)^k\alpha^{k+z}}{k!(k + 2)}, \qquad z \neq 0, -1, -2, \ldots$$

2

THE PROBABILITY INTEGRAL AND RELATED FUNCTIONS

2.1. The Probability Integral and Its Basic Properties

By the *probability integral* is meant the function defined for any complex z by the integral

$$\Phi(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt, \tag{2.1.1}$$

evaluated along an arbitrary path joining the origin to the point $t = z$. The form of this path does not matter, since the integrand is an entire function of the complex variable t, and in fact we can assume that the integration is along the line segment joining the points $t = 0$ and $t = z$. According to a familiar theorem of complex variable theory,[1] $\Phi(z)$ is an entire function and hence can be expanded in a convergent power series for any value of z. To find this expansion, we need only replace e^{-t^2} by its power series in (2.1.1), and then integrate term by term (this is always permissible for power series[2]), obtaining

$$\Phi(z) = \frac{2}{\sqrt{\pi}} \int_0^z \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k}}{k!}\, dt = \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k+1}}{k!(2k+1)}, \qquad |z| < \infty. \tag{2.1.2}$$

---

[1] If $f(t)$ is analytic in a simply connected domain D, then the integral

$$\varphi(z) = \int_a^z f(t)\, dt,$$

evaluated along any rectifiable path contained in D, defines an analytic function in D. See A. I. Markushevich, *op. cit.*, Theorem 13.5, p. 282. The theorem remains true if $f(a) = \infty$ or $a = \infty$, provided that the improper integral exists.

[2] *Ibid.*, Theorems 16.3 and 15.4, pp. 348 and 325.

2.5 Gamma Function: Applications

ON THE GAMMA FUNCTION AND ITS APPLICATIONS

JOEL AZOSE

1. Introduction

The common method for determining the value of $n!$ is naturally recursive, found by multiplying $1*2*3*...*(n-2)*(n-1)*n$, though this is terribly inefficient for large n. So, in the early 18th century, the question was posed: As the definition for the nth triangle number can be explicitly found, is there an explicit way to determine the value of $n!$ which uses elementary algebraic operations? In 1729, Euler proved no such way exists, though he posited an integral formula for $n!$. Later, Legendre would change the notation of Euler's original formula into that of the gamma function that we use today [1].

While the gamma function's original intent was to model and interpolate the factorial function, mathematicians and geometers have discovered and developed many other interesting applications. In this paper, I plan to examine two of those applications. The first involves a formula for the n-dimensional ball with radius r. A consequence of this formula is that it drastically simplifies the discussion of which fits better: the n-ball in the n-cube or the n-cube in the n-ball. The second application is creating the psi and polygamma functions, which will be described in more depth later, and allow for an alternate method of computing infinite sums of rational functions.

Let us begin with a few definitions: The **gamma function** is defined for $\{z \in \mathbb{C}, z \neq 0, -1, -2, \dots\}$ to be:

$$\Gamma(z) = \int_0^\infty s^{z-1}e^{-s}ds \tag{1.1}$$

Remember some important characteristics of the gamma function:
1) For $z \in \{\mathbb{N} \setminus 0\}, \Gamma(z) = z!$
2) $\Gamma(z+1) = z\Gamma(z)$
3) $\ln(\Gamma(z))$ is convex.

The **beta function** is defined for $\{x, y \in \mathbb{C}, Re(x) > 0, Re(y) > 0\}$ to be:

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt. \tag{1.2}$$

Another identity yields:

$$B(x,y) = 2\int_0^{\pi/2} sin^{2x-1}\theta\, cos^{2y-1}\theta d\theta \tag{1.3}$$

Additionally,

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{1.4}$$

A thorough proof of this last identity appears in Folland's *Advanced Calculus* [2] on pages 345 and 346. To summarize, the argument relies primarily on manipulation of $\Gamma(x)$ and $\Gamma(y)$ in their integral forms (1.1), converting to polar coordinates, and separating the double integral. This identity will be particularly important in our derivation for the formula for the volume of the n-dimensional ball later in the paper.

With these identities in our toolkit, let us begin.

2. Balls And The Gamma Function

2.1. **Volume Of The N-Dimensional Ball.** In his article, *The Largest Unit Ball in Any Euclidean Space*, Jeffrey Nunemacher lays down the basis for one interesting application of the gamma function, though he never explicitly uses the gamma function [3]. He first defines the open ball of radius r of dimenision n, $B_n(r)$, to be the set of points such that, for $1 \leq j \leq n$,

$$\sum x_j^2 < r^2. \tag{2.1}$$

Its volume will be referred to as $\boldsymbol{V_n(r)}$. In an argument that he describes as being "accessible to a multivariable calculus class", Nunemacher uses iterated integrals to derive his formula. He notes that, by definition:

$$V_n(r) = \iint\limits_{B_n(r)} \dots \int 1\; dx_1\, dx_2\, \dots\, dx_n \tag{2.2}$$

By applying (2.1) to the limits of the iterated integral in (2.2) and performing trigonometric substitutions, he gets the following - more relevant - identity, specific to the unit ball, where $r = 1$:

$$V_n = 2V_{n-1} \int_0^{\pi/2} cos^n\theta \; d\theta \tag{2.3}$$

In the rest of *The Largest Unit Ball in Any Euclidean Space*, Nunemacher goes on to determine which unit ball in Euclidean space is the largest. (He ultimately shows that the unit ball of dimension $n = 5$ has the greatest volume, and that the unit ball of dimension $n = 7$ has the greatest surface area, as well as - curiously - noting that V_n goes to 0 as n gets large. While a surprising result, it is not immediately relevant to the topics which I aim to pursue here. If interested, I would refer the reader to Nunemacher's article directly.) Notice, however, that this formula does not use the gamma function. We begin the derivation from here of the Gamma function form.

2.2. **Derivation.** In his 1964 article, *On Round Pegs In Square Holes And Square Pegs In Round Holes* [4], David Singmaster uses the following formula for the volume of an n-dimensional ball:

$$V_n(r) = \frac{\pi^{n/2} r^n}{\Gamma(n/2 + 1)} \tag{2.4}$$

However, he never shows the derivation of this formula, and other references to Singmaster's article claim that the derivation appears explicitly in Nunemacher's article. I feel this to be an important omission, and I have endeavored here to recreate the derivation for the sake of completeness. We shall begin where Nunemacher left off with equation (2.3).

Recall (1.3) and notice its similarity to (2.3). It quickly becomes apparent that (2.3) may be rewritten as:

$$V_n(1) = V_{n-1}(1)\, B(\frac{1}{2}, \frac{n}{2} + \frac{1}{2}) \tag{2.5}$$

Continuing the recursion, we note:

$$V_{n-1}(1) = V_{n-2}(1)\, B(\frac{1}{2}, \frac{n}{2}) \tag{2.6}$$

Consequently,

$$V_n(1) = V_1(1)\, B(\frac{1}{2}, \frac{3}{2}) \dots\, B(\frac{1}{2}, \frac{n}{2})\, B(\frac{1}{2}, \frac{n}{2} + \frac{1}{2}) \tag{2.7}$$

where $V_1(1) = 2 = B(\frac{1}{2}, 1)$. Substituting Gamma for Beta using (1.4) gives:

$$V_n(1) = \left[\frac{\Gamma(\frac{1}{2})\Gamma(1)}{\Gamma(\frac{3}{2})} \frac{\Gamma(\frac{1}{2})\Gamma(\frac{3}{2})}{\Gamma(2)} \cdots \frac{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})} \frac{\Gamma(\frac{1}{2})\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2}+1)} \right], \tag{2.8}$$

which telescopes to:

$$V_n(1) = \left[\frac{(\Gamma(\frac{1}{2}))^n \Gamma(1)}{\Gamma(n/2+1)} \right] \tag{2.9}$$

Since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\Gamma(1) = 1$,

$$V_n(1) = \frac{\pi^{n/2}}{\Gamma(n/2+1)} \tag{2.10}$$

Now the heavy lifting is done. Consider again the recursion relation that we used in (2.3). This recursion relation holds true for the unit ball - that is, when $r = 1$. However, when $r = 1$, we do not see the r in this equation. Instead, when we take the more general form, we get the modified recursion relation:

$$V_n = 2rV_{n-1} \int_0^{\pi/2} cos^n\theta \; d\theta \tag{2.11}$$

Going through the derivation will be virtually identical, except we have dilated the ball's size by a factor of r, and its volume by a factor of r^n. This finally yields:

$$V_n(r) = \frac{\pi^{n/2} r^n}{\Gamma(n/2+1)}, \tag{2.12}$$

which is consistent with with our original statement of (2.4). Now the derivation of the n-ball's volume using the gamma function is complete, and we may proceed to an interesting application.

2.3. **The Packing Problem.** In the motivation for his article, Singmaster explains the purpose of his article: "Some time ago, the following problem occurred to me: which fits better, a round peg in a square hole or a square peg in a round hole? This can easily be solved once one arrives at the following mathematical formulation of the problem. Which is larger: the ratio of the area of a circle to the area of the circumscribed square or the ratio of the area of a square to the area of the circumscribed circle?" [4]

The formula that we derived in the last section will prove invaluable in finding this. Since he is focusing on ratios, Singmaster uses the unit ball in both cases, though it would work similarly with any paired radius.

For the unit ball, the edge of the circumscribed cube is necessarily length 2, since it is equal in length to a diameter of the unit ball. The edge of the n-cube inscribed in the unit n-ball has length $2/\sqrt{n}$, since the diagonal of an n-cube is $\sqrt{n}$ times its edge. (Remember that the diagonal of the n-cube inscribed in the unit n-ball is the diameter of the n-ball.)

So, we construct formulas for the volume of the relevant balls and cubes using (2.4) and the facts which we have just stated:

$$V(n) = \frac{\pi^{n/2}}{\Gamma(n/2+1)}, \tag{2.13}$$

$$V_c(n) = 2^n, \tag{2.14}$$

$$V_i(n) = \frac{2^n}{n^{n/2}}, \tag{2.15}$$

where $V(n)$ represents the volume of the unit n-ball (as derived), $V_c(n)$ the volume of the circumscribed cube, and $V_i(n)$ the volume of the inscribed cube. We consider now the ratios of (2.13) to (2.14) - that is, a round peg in a square hole - and that of (2.15) to (2.13) - a square peg in a round hole.

$$R_1(n) = \frac{V(n)}{V_c(n)} = \frac{\pi^{n/2}}{2^n\Gamma\left(\frac{n+2}{2}\right)} \tag{2.16}$$

$$R_2(n) = \frac{V_i(n)}{V(n)} = \frac{2^n\Gamma\left(\frac{n+2}{2}\right)}{n^{n/2}\pi^{n/2}} \tag{2.17}$$

He then takes $\dfrac{R_1(n)}{R_2(n)}$ and applies Stirling's approximation for the gamma function: For z large,

$$\Gamma(z) \sim z^{z-1/2}\, e^{-z}\, \sqrt{2\pi} \tag{2.18}$$

Singmaster shows that as n goes to infinity, this ratio goes to zero. So, for large enough n, $R_2(n)$ is greater. By simple numerical evaluation, he determines the tipping point to be when $n = 9$. The most important result of this article is the following theorem:

Theorem. *The n-ball fits better in the n-cube better than the n-cube fits in the n-ball if and only if* $n \leq 8$.

3. Psi And Polygamma Functions

In addition to the earlier, more frequently used definitions for the gamma function, Weierstrass proposed the following:

$$\frac{1}{\Gamma(z)} = z\, e^{\gamma z} \prod_{n=1}^{\infty} (1 + z/n)\, e^{-z/n}, \tag{3.1}$$

where γ is the Euler-Mascheroni constant. Van der Laan and Temme reference another proof of this by Hochstadt [1]. This will be useful in developing the new gamma-related functions in the subsections to follow, as well as important identities. Ultimately, we will provide definitions for the **psi function** - also known as the **digamma function** - as well as the **polygamma functions**. We will then examine how the psi function proves to be useful in the computation of infinite rational sums.

3.1. **Definitions.** Traditionally, $\psi(z)$ is defined to be the derivative of $\ln(\Gamma(z))$ with respect to z, also denoted as $\dfrac{\Gamma'(z)}{\Gamma(z)}$. Just as with the gamma function, $\psi(z)$ is defined for $\{z \in \mathbb{C}, z \neq 0, -1, -2, \dots\}$. Van der Laan and Temme provide several very useful definitions for the psi function. The most well-known representation, derived from (3.1) and the definition of $\psi(z)$, is as follows:

$$\psi(z) = -\gamma - \frac{1}{z} + \sum_{n=1}^{\infty} \frac{z}{n(z+n)}, \tag{3.2}$$

though the one that we will ultimately use in the following subsection to compute sums is defined thusly:

$$\psi(z) = -\gamma - \int_0^1 \frac{t^{z-1} - 1}{1 - t}\, dt \tag{3.3}$$

This integral holds true for $Re(z) > -1$, and can be verified by expanding the denominator of the integrand and comparing to (3.2). These two are the most important definitions for the psi function, and they are the two that we will primarily use.

We will now define the **polygamma functions**, $\psi^{(k)}$.This is a family of functions stemming from the gamma and digamma functions. They are useful because they lead to better- and better-converging series. As you might imagine from the notation, the polygamma functions are the higher-order derivatives of $\psi(z)$. Consider these examples from repeated differentiation of (3.2):

$$\psi'(z) = \sum_{n=0}^{\infty} (z+n)^{-2}, \quad \psi^{(k)}(z) = (-1)^{k+1} k! \sum_{n=0}^{\infty} (z+n)^{-k-1} \tag{3.4}$$

Again, we note that, as k increases, $\psi^{(k)}(z)$ becomes more and more convergent. Now, though, we will set aside the polygamma functions and turn our focus back to the psi function and its utility in determining infinite sums.

3.2. **Use In The Computation Of Infinite Sums.** Late in their chapter on some analytical applications of the gamma, digamma, and polygamma functions, van der Laan and Temme state: "An infinite series whose general term is a rational function in the index may always be reduced to a finite series of psi and polygamma functions" [1].

Let us consider the following specific problem to motivate more general results given at the end of this section.

$$\text{Evaluate } \sum_{n=1}^{\infty} \frac{1}{(n+1)(3n+1)}. \tag{3.5}$$

We begin by expressing the summand as u_n, noting that $u_n = \frac{1}{3}\left(\frac{1}{(n+1)(n+1/3)}\right)$. Then we perform partial fraction decomposition to yield that $\frac{1}{(n+1)(n+1/3)} =$

$\frac{3/2}{n+1} - \frac{3/2}{n+1/3}$, so $u_n = \frac{1}{2}\left(\frac{1}{n+1} - \frac{1}{n+1/3}\right)$. Remember the identity that, for all $A > 0$,

$$\frac{1}{A} = \int_0^\infty e^{-Ax}dx \tag{3.6}$$

This identity can be applied, since both denominators of both fractions are necessarily greater than 0. So the sum in (3.5) can be rewritten as:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{(n+1)(3n+1)} &= \frac{1}{2}\sum_{n=1}^{\infty}\left[\int_0^\infty e^{-(n+1)x}dx - \int_0^\infty e^{-(n+1/3)x}dx\right] \\
&= \frac{1}{2}\sum_{n=1}^{\infty}\left[\int_0^\infty e^{-nx}e^{-x}dx - \int_0^\infty e^{-nx}e^{-(1/3)x}dx\right] \\
&= \frac{1}{2}\sum_{n=1}^{\infty}\left[\int_0^\infty e^{-nx}(e^{-x} - e^{-(1/3)x})dx\right] \\
&= \frac{1}{2}\lim_{N\to\infty}\left(\sum_{n=1}^{N}\left[\int_0^\infty e^{-nx}(e^{-x} - e^{-(1/3)x})dx\right]\right)
\end{aligned}$$

Remember from the study of infinite series that $\sum_{n=0}^{N} x^n = \frac{1-x^{N+1}}{1-x}$. When we subtract the first term of the series, $x^0 = 1$, we get the following result:

$$\sum_{n=1}^{N} x^n = \frac{x(1-x^N)}{1-x}. \tag{3.7}$$

Plugging in e^{-x} for x, we see:

$$\sum_{n=1}^{N} e^{-nx} = \frac{e^{-x}(1-e^{-Nx})}{1-e^{-x}} \tag{3.8}$$

Consider the relevant summation. Due to appropriate convergences following from the monotone convergence theorem, we can interchange the summation and integration and continue our manipulations of the sum.

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{1}{(n+1)(3n+1)} &= \frac{1}{2}\lim_{N\to\infty}\left[\int_0^{\infty}\frac{e^{-x}(1-e^{-Nx})}{1-e^{-x}}(e^{-x}-e^{-(1/3)x})dx\right]\\ &= \frac{1}{2}\int_0^{\infty}\frac{e^{-x}(e^{-x}-e^{-(1/3)x})}{1-e^{-x}}dx\end{aligned}$$

Now we make use of a change of variables. Let $t = e^{-x}$. Consequently, $-e^{-x}dx = dt$. We will make this substitution. The negative sign due to this change of variable cancels with the one created by switching the limits of the integral, to yield the following:

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{1}{(n+1)(3n+1)} &= \frac{1}{2}\int_0^{1}\frac{t-t^{1/3}}{1-t}dt\\ &= \frac{1}{2}\int_0^{1}\frac{(t-1)-(t^{1/3}-1)}{1-t}dt\\ &= \frac{1}{2}\int_0^{1}\frac{t-1}{1-t}dt-\frac{1}{2}\int_0^{1}\frac{t^{1/3}-1}{1-t}dt\end{aligned}$$

Compare the two integrals on the right hand side of the above equation to the formula for $\psi(z)$ in (3.3). It becomes obvious that the substitution can be made with the psi function to yield our final result:

$$\sum_{n=1}^{\infty}\frac{1}{(n+1)(3n+1)} = \frac{1}{2}\psi(4/3)-\frac{1}{2}\psi(2). \tag{3.9}$$

Professor Efthimiou of Tel Aviv University puts forth a theorem regarding series of the form

$$S(a,b) = \sum_{n=1}^{\infty}\frac{1}{(n+a)(n+b)}, \tag{3.10}$$

where $a \neq b$, and $\{a, b \in \mathbb{C}; Re(a), Re(b) > 0\}$ that generalizes the result which we have shown for a specific example above:

Theorem. $S(a,b) = \dfrac{\psi(b+1) - \psi(a+1)}{b-a}$. [5]

Let it be noted that, at present, our utility of psi functions in the calculation of infinite sums is relegated to strictly positive fractions. (Admittedly, even this is handy in a pinch, though it is hardly ideal.) However, I hope that the thorough calculation of this example is proof enough for the reader that this derivation can be made, and that the same argument could be made for a similar - that is, strictly positive - function with a denominator of degree 2. If a doubt persists, I urge the reader to create a rational function of this form and follow the same steps as my proof to derive an equivalence with a sum of psi and/or polygamma functions.

4. Future Works

Van der Laan and Temme propose that *every* infinite series of rational functions may be reduced to a finite series of psi and polygamma functions. This seems plausible, but the statement requires more rigorous examination to be taken as sound. The subjects that I would like to delve the most deeply into are what I touched on at the very end with Prof. Efthimiou's theorem and the limits on the utility of the psi function in the calculation of infinite sums. I think that it would be a worthwhile endeavor to try to formulate an analogue of Efthimiou's theorem for a function with denominator of degree n. Finally, I would like to work on examining what could be done with infinite sums of fractions that are not strictly positive. I would like to determine if there is a similar formula for these series, as well.

5. Conclusion

In the first section of this paper, we provided definitions for the gamma function. We then went through a gamma derivation for the formula of the volume of an n-ball and used that in working with ratios involving inscribed and circumscribed cubes to determine the following:

Theorem. *The n-ball fits better in the n-cube better than the n-cube fits in the n-ball if and only if* $n \leq 8$.

In the second section, we presented the psi - also known as the digamma - function and the family of polygamma functions. We expressed a specific infinite sum as the finite sum of psi functions as motivation for the following more general result:

Theorem. *For* $a \neq b$, *and* $\{a, b \in \mathbb{C}; Re(a), Re(b) > 0\}$, $\sum_{n=1}^{\infty} \dfrac{1}{(n+a)(n+b)} = \dfrac{\psi(b+1) - \psi(a+1)}{b-a}$.

References

[1] C.G. van der Laan and N.M. Temme. Calculation of special functions: the gamma function, the exponential integrals and error-like functions. *CWI Tract*, 10, 1984.
[2] Gerald Folland. *Advanced Calculus*. Prentice Hall, Upper Saddle River, NJ, 2002.
[3] Jeffrey Nunemacher. The largest unit ball in any euclidean space. *Mathematics Magazine*, 59(3):170–171, 1986.
[4] David Singmaster. On round pegs in square holes and square pegs in round holes. *Mathematics Magazine*, 37(5):335–337, 1964.
[5] Costas Efthimiou. Finding exact values for infinite sums, 1998.

2.6 Zeta Function

Lecture 8

The Zeta Function of Riemann

1 Elementary Properties of $\zeta(s)$ [16, pp.1-9]

We define $\zeta(s)$, for s complex, by the relation **66**

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \ s = \sigma + it, \ \sigma > 1. \tag{1.1}$$

We define x^s, for $x > 0$, as $e^{s \log x}$, where $\log x$ has its real determination. Then $|n^s| = n^\sigma$, and the series converges for $\sigma > 1$, and uniformly in any finite region in which $\sigma \geq 1 + \delta > 1$. Hence $\zeta(s)$ is regular for $\sigma \geq 1 + \delta > 1$. Its derivatives may be calculated by termwise differentiation of the series.

We could express $\zeta(s)$ also as an infinite product called the *Euler Product*:

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}, \sigma > 1, \tag{1.2}$$

where p runs through all the primes (known to be infinite in number!). It is the Euler product which connects the properties of ζ with the properties of primes. *The infinite product is absolutely convergent for* $\sigma > 1$, since the corresponding series

$$\sum_p \left|\frac{1}{p^s}\right| = \sum_p \frac{1}{p^\sigma} < \infty,\ \sigma > 1$$

Expanding each factor of the product, we can write it as

$$\prod_p \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\right)$$

67 Multiplying out formally we get the expression (1.1) by the unique factorization theorem. This expansion can be justified as follows:

$$\prod_{p\leq p}\left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\right) = 1 + \frac{1}{n_1^s} + \frac{1}{n_2^s} + \cdots,$$

where $n_1, n_2, \ldots$ are those integers none of whose prime factors exceed P. Since all integers $\leq P$ are of this form, we get

$$\begin{aligned}\left|\sum_{n=1}^{\infty}\frac{1}{n^s} - \prod_{p\leq P}\left(1 - \frac{1}{p^2}\right)^{-1}\right| &= \left|\sum_{n=1}^{\infty}\frac{1}{n^s} - 1 - \frac{1}{n_1^s} - \frac{1}{n_2^s} - \cdots\right| \\ &\leq \frac{1}{(P+1)^\sigma} + \frac{1}{(P+2)^\sigma} + \cdots\end{aligned}$$

$\to 0$, as $P \to \infty$, if $\sigma > 1$.

Hence (1.2) has been established for $\sigma > 1$, on the basis of (1.1). As an immediate consequence of (1.2) we observe that $\zeta(s)$ *has no zeros for* $\sigma > 1$, since a convergent infinite product of non-zero factors is non-zero.

We shall now obtain a few formulae involving $\zeta(s)$ which will be of use in the sequel. We first have

$$\log\zeta(s) = -\sum_p \log\left(1 - \frac{1}{p^2}\right),\ \sigma > 1. \tag{1.3}$$

If we write $\pi(x) = \sum\limits_{p \leq x} 1$, then

$$\begin{aligned}
\log\zeta(s) &= -\sum_{n=2}^{\infty}\{\pi(n) - \pi(n-1)\}\log\left(1 - \frac{1}{n^s}\right)\\
&= -\sum_{n=2}^{\infty}\pi(n)\left[\log\left(1 - \frac{1}{n^s}\right) - \log\left(1 - \frac{1}{(n+1)^s}\right)\right]\\
&= \sum_{n=2}^{\infty}\pi(n)\int\limits_{n}^{n+1}\frac{s}{x(x^s - 1)}\,dx
\end{aligned}$$

68

Hence

$$\log\zeta(s) = s\int\limits_{2}^{\infty}\frac{\pi(x)}{x(x^s - 1)}dx, \ \sigma > 1. \tag{1.4}$$

It should be noted that the rearrangement of the series preceding the above formula is permitted because

$$\pi(n) \leq n \text{ and } \log\left(1 - \frac{1}{n^s}\right) = O(n^{-\sigma}).$$

A slightly different version of (1.3) would be

$$\log\zeta(s) = \sum_{p}\sum_{m}\frac{1}{mp^{ms}}, \quad \sigma > 1 \tag{1.3$'$}$$

where p runs through all primes, and m through all positive integers. Differentiating (1.3), we get

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{p}\frac{p^{-s}\log p}{1 - p^{-s}} = \sum_{p,m}\sum\frac{\log\rho}{p^{ms}},$$

or

$$\boxed{\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty}\frac{\wedge(n)}{n^s}, \quad \sigma > 1} \quad , \tag{1.5}$$

where $\wedge(n) = \begin{cases} \log p, & \text{if } n \text{ is a + ve power of a prime } p \\ 0, & \text{otherwise.} \end{cases}$ 69

Again

$$\frac{1}{\zeta(s)} = \prod_p \left(1 - \frac{1}{p^s}\right)$$

$$= \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}, \quad \sigma > 1 \tag{1.6}$$

(1.6) where $\mu(1) = 1$, $\mu(n) = (-1)^k$ if n is the product of k different primes, $\mu(n) = 0$ if n contains a factor to a power higher than the first.

We also have

$$\zeta^2(s) = \sum_{n=1}^{\infty} \frac{d(n)}{n^s}, \quad \sigma > 1$$

where $d(n)$ denotes the number of divisors of n, including 1 and n. For

$$\zeta^2(s) = \sum_{m=1}^{\infty} \frac{1}{m^s} \cdot \sum_{n=1}^{\infty} \frac{1}{n^s} = \sum_{\mu=1}^{\infty} \frac{1}{\mu^s} \sum_{mn=\mu} 1$$

More generally

$$\boxed{\zeta^k(s) = \sum_{n=1}^{\infty} \frac{d_k(n)}{n^s}, \quad \sigma > 1} \tag{1.7}$$

70 where $k = 2, 3, 4, \ldots$, and $d_k(n)$ is the number of ways of expressing n as a product of k factors.

Further

$$\zeta(s) \cdot \zeta(s-a) = \sum_{m=1}^{\infty} \frac{1}{m^s} \sum_{n=1}^{\infty} \frac{n^a}{n^s},$$

$$= \sum_{\mu=1}^{\infty} \frac{1}{\mu^s} \cdot \sum_{mn=\mu} n^a,$$

so that

$$\zeta(s) \cdot \zeta(s-a) = \sum_{\mu=1}^{\infty} \frac{\sigma_2(\mu)}{\mu^\delta} \tag{1.8}$$

where $\sigma_a(\mu)$ denotes the sum of the a^{th} powers of the divisors of μ.

If $a \neq 0$, we get, from expanding the respective terms of the Euler products,

$$\begin{aligned}
\zeta(a)\zeta(s-a) &= \prod_p \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\right)\left(1 + \frac{p^a}{p^s} + \frac{p^{2a}}{p^{2a}} + \cdots\right) \\
&= \prod_p \left(1 + \frac{1+p^a}{p^s} + \frac{1+p^a+p^{2a}}{p^{2s}} + \cdots\right) \\
&= \prod_p \left(1 + \frac{1-p^{2a}}{1-p^a} \cdot \frac{1}{p^s} + \cdots\right)
\end{aligned}$$

Using (1.8) we thus get

$$\sigma_2(n) = \frac{1-p_1^{(m_1+1)a}}{1-p_1^a} \ldots \frac{1-p_r^{(m_r+1)a}}{1-p_r^a} \tag{1.9}$$

if $n = p_1^{m_1}, p_2^{m_2} \ldots p_r^{m_r}$ by comparison of the coefficients of $\dfrac{1}{n^s}$. **71**

More generally, we have

$$\frac{\zeta(s)\cdot\zeta(s-a)\cdot\zeta(s-b)\cdot\zeta(s-a-b)}{\zeta(2s-a-b)} = \prod_p \frac{1-p^{-2s+a+b}}{(1-p^{-s})(1-p^{-s+a})(1-p^{-s+b}) \atop (1-p^{-s+a+b})}$$

for $\sigma > \max\{1, \operatorname{Re} a + 1,\ \operatorname{Re} b + 1, \operatorname{Re}(a+b)+1\}$. Putting $p^{-s} = z$, we get the general term in the right hand side equal to

$$\begin{aligned}
&\frac{1-p^{a+b}z^2}{(1-z)\cdot(1-p^az)\cdot(1-p^bz)(1-p^{a+b}z)} \\
&= \frac{1}{(1-p^a)(1-p^b)}\left\{\frac{1}{1-z} - \frac{p^a}{1-p^az} - \frac{p^b}{1-p^bz} + \frac{p^{a+b}}{1-p^{a+b}z}\right\} \\
&= \frac{1}{(1-p^a)(1-p^b)} \sum_{m=0}^{\infty} \left\{1 - p^{(m+1)a} - p^{(m+1)b} + p^{(m+1)(a+b)}\right\} z^m \\
&= \frac{1}{(1-p^a)(1-p^b)} \sum_{m=o}^{\infty} \left\{1 - p^{(m+1)a}\right\}\left\{1 - p^{(m+1)b}\right\} z^m
\end{aligned}$$

Hence 72

$$\frac{\zeta(s)\cdot\zeta(s-a)\zeta(s-b)\cdot\zeta(s-a-h)}{\zeta(2s-a-b)}$$
$$=\prod_p\sum_{m=0}^{\infty}\frac{1-p^{(m+1)a}}{1-p^a}\cdot\frac{1-p^{(m+1)b}}{1-p^b}\cdot\frac{1}{p^{ms}}$$

Now using (1.9) we get

$$\frac{\zeta(s)\cdot\zeta(s-a)\cdot\zeta(s-b)\zeta(s-a-b)}{\zeta(2s-a-b)}=\sum_{n=1}^{\infty}\frac{\sigma_a(n)\sigma_b(n)}{n^s}\tag{1.10}$$

$$\sigma>\max\{1,\operatorname{Re}a+1,\operatorname{Re}b+1,\operatorname{Re}(a+b)+1\}$$

If $a=b=0$, then

$$\frac{\{\zeta(s)\}^4}{\zeta(2s)}=\sum_{n=1}^{\infty}\frac{\{d(n)^2\}}{n^s},\quad \sigma>1.\tag{1.11}$$

If α is real, and $\alpha\neq 0$, we write αi for a and $-\alpha i$ for b in (1.10), and get

$$\frac{\zeta^2(s)\zeta(s-\alpha i)\zeta(s+\alpha i)}{\zeta(2s)}=\sum_{n=1}^{\infty}\frac{|\sigma_{\alpha i}(n)|^2}{n^s},\ \sigma>1,\tag{1.12}$$

where $\sigma_{\alpha i}(n)=\sum\limits_{d/n}d^{\alpha i}$.

Lecture 11

The Zeta Function of Riemann (Contd)

3 Analytic continuation of $\zeta(s)$. First method [16, p.18]

We have, for $\sigma > 0$, 92

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx.$$

Writing nx for x, and summing over n, we get

$$\begin{aligned}\sum_{n=1}^\infty \frac{\Gamma(s)}{n^s} &= \sum_{n=1}^\infty \int_{n=0}^\infty x^{s-1} e^{-nx} dx\\ &= \int_0^\infty \left(\sum_{n=1}^\infty e^{-nx}\right) x^{s-1} dx, \text{ if } \sigma > 1.\end{aligned}$$

since

$$\sum_{n=1}^\infty \left| \int_0^\infty x^{s-1} e^{-nx} dx \right| \leq \sum_{n=1}^\infty \int_0^\infty x^{\sigma-1} e^{-nx} dx = \sum_{n=1}^\infty \frac{\Gamma(\sigma)}{n^\sigma} < \infty,$$

if $\sigma > 1$. Hence

$$\sum_{n=1}^{\infty} \frac{\Gamma(s)}{n^s} = \int_0^{\infty} \frac{e^{-x}}{1-e^{-x}} x^{s-1} dx = \int_0^{\infty} \frac{x^{s-1} dx}{e^x - 1}$$

or

$$\boxed{\Gamma(s)\zeta(s) = \int_0^{\infty} \frac{x^{s-1}}{e^x - 1} dx, \quad \sigma > 1}$$

In order to continue $\zeta(s)$ analytically all over the s-plane, consider the complex integral

$$I(s) = \int_C \frac{z^{s-1}}{e-1} dz$$

where C is a contour consisting of the real axis from $+\infty$ to ρ, $0 < \rho <$
93 2π, the circle $|z| = \rho$; and the real axis from ρ to ∞. $I(s)$, if convergent,
is independent of ρ, by *Cauchy's* theorem.

Now, on the circle $|z| = \rho$, we have

$$\begin{aligned} |z^{s-1}| = |e^{(s-1)\log z}| &= |e^{\{(\sigma-1)+it\}\{\log|z|+i\arg z\}}| \\ &= e^{(\sigma-1)\log|z| - t\arg z} \\ &= |z|^{\sigma-1} e^{2\pi|t|}, \end{aligned}$$

while

$$|e^z - 1| > A|z|;$$

Hence, for fixed s,

$$\Big| \int_{|z|=\rho} \Big| \leq \frac{2\pi\rho \cdot \rho^{\sigma-1}}{A\rho} \cdot e^{2\pi|t|} \to 0 \text{ as } \rho \to 0, \text{ if } \sigma > 1.$$

Thus, on letting $\rho \to 0$, we get, if $\sigma > 1$,

$$I(s) = -\int_0^{\infty} \frac{x^{s-1}}{e^x - 1} dx + \int_0^{\infty} \frac{(xe^{2\pi i})^{s-1}}{e^x - 1}\, dx$$

$$= -\Gamma(s)\zeta(s) + e^{2\pi is}\Gamma(s)\zeta(s)$$
$$= \Gamma(s)\zeta(s)(e^{2\pi s} - 1).$$

Using the result
$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin\pi s}$$
we get **94**
$$I(s) = \frac{\zeta(s)}{\Gamma(1-s)} \cdot 2\pi i . \frac{e^{2\pi is} - 1}{e^{\pi is} - e^{-\pi is}}$$
$$= \frac{\zeta(s)}{\Gamma(1-s)} \cdot 2\pi i \cdot e^{\pi is},$$
or
$$\boxed{\zeta(s) = \frac{e^{-i\pi s}\Gamma(1-s)}{2\pi i} \int\limits_C \frac{z^{s-1}dz}{e^z - 1}, \quad \sigma > 1.}$$
The integral on the right is uniformly convergent in any finite region of the s-plane (by obvious majorization of the integrand), and so defines an entire function. Hence the above formula, proved first for $\sigma > 1$, defines $\zeta(s)$, as a meromorphic function, all over the s-plane. This only possible poles are the poles of $\Gamma(1-s)$, namely $s = 1, 2, 3, \ldots$. We know that $\zeta(s)$ is regular for $s = 2, 3, \ldots$ (As a matter of fact, $I(s)$ vanishes at these points). Hence the only possible pole is at $s = 1$.

Hence
$$I(1) = \int\limits_C \frac{dz}{e^z - 1} = 2\pi i,$$
while
$$\Gamma(1-s) = -\frac{1}{s-1} + \cdots$$
Hence *the residue of* $\zeta(s)$ *at* $s = 1$ *is* 1.

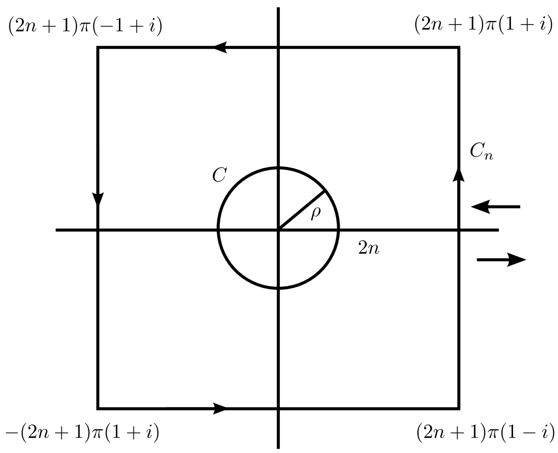
We see in passing, since
$$\frac{1}{e^z - 1} = \frac{1}{z} - \frac{1}{2} + B_1\frac{z}{2!} - B_2\frac{z^3}{4!} + \cdots$$
that **95**
$$\zeta(0) = -\frac{1}{2}, \zeta(-2m) = 0, \zeta(1-2m) = \frac{(-1)^m B_m}{2m}, m = 1, 2, 3, \ldots$$

4 Functional Equation (First method) [16, p.18]

Consider the integral

$$\int \frac{z^{s-1}dz}{e^z - 1}$$

taken along C_n as in the diagram.



Between C and C_n, the integrand has poles at

$$\pm 2i\pi, \ldots, \pm 2ni\pi.$$

The residue at

$$2m\pi i \text{ is } (2m\pi e^{\frac{\pi}{2}i})^{s-1}$$

while the residue at $-2m\pi i$ is $(2m\pi e^{3/2\pi i})^{s-1}$; taken together they amount to

$$\begin{aligned}
&(2m\pi)e^{\pi i(s-1)}\left[e^{\frac{\pi i}{2}(s-1)} + e^{-\frac{\pi i}{2}(s-1)}\right] \\
&= (2m\pi)^{s-1}e^{\pi i(s-1)}2\cos\frac{\pi}{2}(s-1) \\
&= -2(2m\pi)^{s-1}e^{\pi is}\sin\frac{\pi}{2}s
\end{aligned}$$

96 Hence

$$I(s) = \int\limits_{C_n} \frac{z^{s-1}dz}{e^z - 1} + 4\pi i \frac{\sin \pi s}{2} e^{\pi i s} \sum_{m=1}^{n} (2m\pi)^{s-1}$$

by the theorem of residues.

Now let $\sigma < 0$, and $n \to \infty$. Then, on C_n,

$$|z^{s-1}| = O(|z|^{\sigma-1}) = O(n^{\sigma-1}),$$

and

$$\frac{1}{e^z - 1} = O(1),$$

for

$$\begin{aligned} |e^z - 1|^2 = |e^{x+iy} - 1|^2 &= |e^x(\cos y + i \sin y) - 1|^2 \\ &= e^{2x} - 2e^x \cos y + 1, \end{aligned}$$

which, on the vertical lines, is $\geq (e^x - 1)^2$ and, on the horizontal lines, $= (e^x + 1)^2$. (since $\cos y = -1$ there).

Also the length of the square-path is $O(n)$. Hence the integral round the square $\to 0$ as $n \to \infty$.

Hence

$$\begin{aligned} I(s) &= 4\pi i e^{\pi i s} \frac{\sin \pi s}{2} \sum_{m=1}^{\infty} (2m\pi)^{s-1} \\ &= 4\pi i e^{\pi i s} \sin \frac{\pi s}{2} \cdot (2\pi)^{s-1} \zeta(1-s), \text{ if } \sigma < 0. \end{aligned}$$

or 97

$$\begin{aligned} \zeta(s)\Gamma(s)(e^{2\pi i s} - 1) &= (4\pi i)(2\pi)^{s-1} e^{\pi i s} \sin \frac{\pi s}{2} \zeta(1-s) \\ &= 2\pi i e^{\pi i s} \frac{\zeta(s)}{\Gamma(1-s)} \end{aligned}$$

Thus

$$\boxed{\zeta(s) = \Gamma(1-s)\zeta(1-s) 2^s \pi^{s-1} \sin \frac{\pi s}{2},}$$

for $\sigma < 0$, and hence, by analytic continuation, for all values of s (each side is regular except for poles!).

This is the functional equation.

Since
$$\Gamma\left(\frac{x}{2}\right)\Gamma\left(\frac{x+1}{2}\right) = \frac{\sqrt{\pi}}{2^{x-1}}\Gamma(x),$$
we get, on writing $x = 1 - s$,
$$2^{-s}\Gamma\left(\frac{1-s}{2}\right)\Gamma\left(1-\frac{s}{2}\right) = \sqrt{\pi}\Gamma(1-s);$$
also
$$\Gamma\left(\frac{s}{2}\right)\Gamma\left(1-\frac{s}{2}\right) = \frac{\pi}{\sin\frac{\pi s}{2}}$$
Hence
$$\Gamma(1-s) = 2^{-s}\Gamma\left(\frac{1-s}{2}\right)\left\{\Gamma\left(\frac{s}{2}\right)\right\}^{-1}\sqrt{\pi}\{\sin\frac{\pi s}{2}\}^{-1}$$
Thus
$$\boxed{\pi^{-s/2}\Gamma(s/2)\zeta(s) = \pi^{-\frac{(1-s)}{2}}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s)}$$

98 If
$$\begin{aligned}\xi(s) &\equiv \frac{1}{2}s(s-1)\pi^{-s/2}\Gamma(s/2)\zeta(s)\\ &\equiv \frac{1}{2}s(s-1)\eta(s),\end{aligned}$$
then $\eta(s) = \eta(1-s)$ and $\xi(s) = \xi(1-s)$. If $\equiv (z) = \xi(\frac{1}{2}+iz)$, then $\equiv (z) \equiv (-z)$.

5 Functional Equation (Second Method) [16, p.13]

Consider the lemma given in Lecture 9, and write $\lambda_n = n$, $a_n = 1$, $\phi(x) = x^{-s}$ in it. We then get
$$\sum_{n\leq x} n^{-s} = s\int_1^X \frac{[x]}{x^{s+1}}dx + \frac{[X]}{X^s}, \text{ if } X \geq 1.$$

$$= \frac{s}{s-1} - \frac{s}{(s-1)X^{s-1}} - s\int_1^X \frac{x-[x]}{x^{s+1}}dx + \frac{1}{X^{s-1}} - \frac{X-[X]}{X^s}$$

Since

$$\left|\frac{1}{X^{s-1}}\right| = 1/X^{\sigma-1}, \text{ and } \left|\frac{X-[X]}{X^s} \le 1/X^{\sigma},\right|$$

we deduce, on making $X \to \infty$,

$$\boxed{\zeta(s) = \frac{s}{s-1} - s\int_1^\infty \frac{x-[x]}{x^{s+1}}dx, \text{ if } \sigma > 1}$$

or

$$\boxed{\zeta(s) = s\int_1^\infty \frac{[x]-x+\frac{1}{2}}{x^{s+1}}dx + \frac{1}{s-1} + \frac{1}{2}, \text{ if } \sigma > 1}$$

5.1 Since $[x] - x + \dfrac{1}{2}$ is bounded, the integral on the right hand side converges for $\sigma > 0$, and uniformly in any finite region to the right of $\sigma = 0$. Hence it represents an analytic function of *s regular for* $\sigma > 0$, and so provides the continuation of $\zeta(s)$ up to $\sigma = 0$, and $s = 1$ is clearly a simple pole with residue 1. **99**

For $0 < \sigma < 1$, we have, however,

$$\int_0^1 \frac{[x]-x}{x^{s+1}}dx = -\int_0^1 x^{-s}dx = \frac{1}{s-1},$$

and

$$\frac{s}{2} = \int_1^\infty \frac{dx}{x^{s+1}} = \frac{1}{2}$$

Hence

5.2

$$\zeta(s) = s \int_0^\infty \frac{[x]-x}{x^{s+1}} dx, \quad 0 < \sigma < 1$$

We have seen that (5.1) gives the analytic continuation up to $\sigma = 0$. By refining the argument dealing with the integral on the right-hand side of (5.1) we can get the continuation all over the s-plane. For, if

$$f(x) \equiv [x] - x + \frac{1}{2}, \quad f_1(x) \equiv \int_1^x f(y) dy,$$

then $f_1(x)$ is *also* bounded, since $\int\limits_n^{n+1} f(y)dy = 0$ for any integer n.

100 Hence

$$\int_{x_1}^{x_2} \frac{f(x)}{x^{s+1}} dx = \frac{f_1(x)}{x^{s+1}}\Bigg]_{x_1}^{x_2} + (s+1) \int_{x_1}^{x_2} \frac{f_1(x)}{x^{s+2}} dx$$

$$\to 0, \text{ as } x_1 \to \infty, x_2 \to \infty,$$

$$\boxed{\text{if } \sigma > -1.}$$

Hence the integral in (5.1) converges for $\sigma > -1$.
Further

$$s \int_0^1 \frac{[x] - x + \frac{1}{2}}{x^{s+1}} dx = \frac{1}{s-1} + \frac{1}{2}, \text{ for } \sigma < 0;$$

$$\boxed{\zeta(s) = s \int_0^\infty \frac{[x]-x+\frac{1}{2}}{x^{s+1}} dx, \quad -1 < \sigma < 0} \qquad (5.3)$$

Now the function $[x] - x + \frac{1}{2}$ has the Fourier expansion

$$\sum_{n=1}^\infty \frac{\sin 2n\pi x}{n\pi}$$

if x is *not* an integer. The series is boundedly convergent. If we substitute it in (5.3), we get

$$\begin{aligned}
\zeta(s) &= \frac{s}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\int\limits_0^{\infty}\frac{\sin 2n\pi x}{x^{s+1}}dx\\
&= \frac{s}{\pi}\sum_{n=1}^{\infty}\frac{(2n\pi)^s}{n}\int\limits_0^{\infty}\frac{\sin y}{y^{s+1}}dy\\
\zeta(s) &= \frac{s}{\pi}(2\pi)^s\{-\Gamma(1-s)\}\sin\frac{s\pi}{2}\zeta(1-s),
\end{aligned}$$

5.4 If termwise integration is permitted, *for* $-1 < \sigma < 0$. The right hand side is, however, analytic for all values of s such that $\sigma < 0$. Hence (5.4) provides the analytic continuation (not only for $-1 < \sigma < 0$) all over the s-plane. **101**

The term-wise integration preceding (5.4) is certainly justified over any *finite* range since the concerned series is boundedly convergent. We have therefore only to prove that

$$\lim_{X\to\infty}\sum_{n=1}^{\infty}\frac{1}{n}\int\limits_X^{\infty}\frac{\sin 2n\pi x}{x^{s+1}}dx = 0, -1 < \sigma < 0$$

Now

$$\begin{aligned}
\int\limits_X^{\infty}\frac{\sin 2n\pi x}{x^{s+1}}dx &= \left[-\frac{\cos 2n\pi x}{2n\pi x^{s+1}}\right]_X^{\infty} - \frac{s+1}{2n\pi}\int\limits_X^{\infty}\frac{2n\pi x}{x^{s+2}}dx\\
&= O\left(\frac{1}{nX^{\sigma+1}}\right) + O\left(\frac{1}{n}\int\limits_X^{\infty}\frac{dx}{x^{\sigma+2}}\right)\\
&= O\left(\frac{1}{nX^{\sigma+1}}\right)
\end{aligned}$$

and hence

$$\lim_{X\to\infty}\sum_{n=1}^{\infty}\frac{1}{n}\int\limits_{X}^{\infty}\frac{\sin 2n\pi x}{x^{s+1}}dx = 0, \text{ if } -1<\sigma<0$$

This completes the analytic continuation and the proof of the Functional equation by the second method.

As a consequence of (5.1), we get

$$\begin{aligned}\lim_{s\to 1}\left\{\zeta(s)-\frac{1}{s-1}\right\} &= \int\limits_{1}^{\infty}\frac{[x]-x+\frac{1}{2}}{x^2}dx+\frac{1}{2}\\ &= \lim_{n\to\infty}\int\limits_{1}^{n}\frac{[x]-x}{x^2}dx+1\\ &= \lim_{n\to\infty}\left\{\sum_{m=1}^{n-1} m\int\limits_{m}^{m+1}\frac{dx}{x^2}-\log n+1\right\}\\ &= \lim_{n\to\infty}\left\{\sum_{m=1}^{n-1}\frac{1}{m+1}+1-\log n\right\}\\ &= \gamma\end{aligned}$$

102 Hence, near $s=1$, we have

$$\zeta(s)=\frac{1}{s-1}+\gamma+O(|s-1|)$$

2.7 Error Function and Applications

2

THE PROBABILITY INTEGRAL AND RELATED FUNCTIONS

2.1. The Probability Integral and Its Basic Properties

By the *probability integral* is meant the function defined for any complex z by the integral

$$\Phi(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt, \tag{2.1.1}$$

evaluated along an arbitrary path joining the origin to the point $t = z$. The form of this path does not matter, since the integrand is an entire function of the complex variable t, and in fact we can assume that the integration is along the line segment joining the points $t = 0$ and $t = z$. According to a familiar theorem of complex variable theory,[1] $\Phi(z)$ is an entire function and hence can be expanded in a convergent power series for any value of z. To find this expansion, we need only replace e^{-t^2} by its power series in (2.1.1), and then integrate term by term (this is always permissible for power series[2]), obtaining

$$\Phi(z) = \frac{2}{\sqrt{\pi}} \int_0^z \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k}}{k!}\, dt = \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k+1}}{k!(2k+1)}, \qquad |z| < \infty. \tag{2.1.2}$$

---

[1] If $f(t)$ is analytic in a simply connected domain D, then the integral

$$\varphi(z) = \int_a^z f(t)\, dt,$$

evaluated along any rectifiable path contained in D, defines an analytic function in D. See A. I. Markushevich, *op. cit.*, Theorem 13.5, p. 282. The theorem remains true if $f(a) = \infty$ or $a = \infty$, provided that the improper integral exists.

[2] *Ibid.*, Theorems 16.3 and 15.4, pp. 348 and 325.

It follows from (2.1.2) that $\Phi(z)$ is an odd function of z. For real values of its argument, $\Phi(z)$ is a real monotonically increasing function, whose graph is shown in Figure 1. At zero we have $\Phi(0) = 0$, and as z increases, $\Phi(z)$ rapidly approaches the limiting value $\Phi(\infty) = 1$, since

$$\int_0^\infty e^{-t^2}\,dt = \frac{\sqrt{\pi}}{2}. \tag{2.1.3}$$

The difference between $\Phi(z)$ and this limit can be written in the form

$$1 - \Phi(z) = \frac{2}{\sqrt{\pi}}\int_0^\infty e^{-t^2}\,dt - \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt = \frac{2}{\sqrt{\pi}}\int_z^\infty e^{-t^2}\,dt. \tag{2.1.4}$$

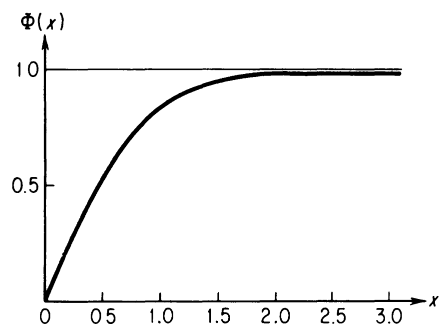


FIGURE 1

The probability integral is encountered in many branches of applied mathematics, e.g., probability theory, the theory of errors, the theory of heat conduction, and various branches of mathematical physics (see Secs. 2.5–2.7). In the literature, one often finds two functions related to the probability integral, i.e., the *error function*

$$\operatorname{Erf} z = \int_0^z e^{-t^2}\,dt = \frac{\sqrt{\pi}}{2}\,\Phi(z), \tag{2.1.5}$$

and its *complement*

$$\operatorname{Erfc} z = \int_z^\infty e^{-t^2}\,dt = \frac{\sqrt{\pi}}{2}\,[1 - \Phi(z)]. \tag{2.1.6}$$

Many more complicated integrals can be expressed in terms of the probability integral. For example, by differentiation of the parameter z it can be shown that

$$\frac{2}{\pi}\int_0^\infty \frac{e^{-zt^2}}{1 + t^2}\,dt = e^z[1 - \Phi(\sqrt{z})]. \tag{2.1.7}$$

2.2. Asymptotic Representation of the Probability Integral for Large $|z|$

To find an asymptotic representation of the function $\Phi(z)$ for large $|z|$, we apply repeated integration by parts to the integral in (2.1.4), obtaining

$$\int_z^\infty e^{-t^2}\,dt = -\frac{1}{2}\int_z^\infty \frac{1}{t}\,d(e^{-t^2}) = \frac{e^{-z^2}}{2z} - \frac{1}{2}\int_z^\infty \frac{e^{-t^2}}{t^2}dt$$

$$= \frac{e^{-z^2}}{2z} - \frac{e^{-z^2}}{2^2z^3} + \frac{1\cdot 3}{z^2}\int_z^\infty \frac{e^{-t^2}}{t^4}\,dt$$

$$= e^{-z^2}\left[\frac{1}{2z} - \frac{1}{2^2z^3} + \frac{1\cdot 3}{2^3z^5} - \frac{1\cdot 3\cdot 5}{2^4z^7} + \cdots + (-1)^n\frac{1\cdot 3\cdots(2n-1)}{2^{n+1}z^{2n+1}}\right]$$

$$+ (-1)^{n+1}\frac{1\cdot 3\cdots(2n+1)}{2^{n+1}}\int_z^\infty \frac{e^{-t^2}}{t^{2n+2}}\,dt.$$

It follows that

$$1 - \Phi(z) = \frac{e^{-z^2}}{\sqrt{\pi}z}\left[1 + \sum_{k=1}^{n}(-1)^k\frac{1\cdot 3\cdots(2k-1)}{(2z^2)^k} + r_n(z)\right], \qquad (2.2.1)$$

where

$$r_n(z) = (-1)^{n+1}\frac{1\cdot 3\cdots(2n+1)}{2^n}\,ze^{z^2}\int_z^\infty \frac{e^{-t^2}}{t^{2n+2}}\,dt. \qquad (2.2.2)$$

Now let

$$|\arg z| \leqslant \frac{\pi}{2} - \delta,$$

where δ is an arbitrarily small positive number, and choose the path of integration in (2.2.2) to be the infinite line segment beginning at the point $t = z$ and parallel to the real axis. If $z = x + iy = re^{i\varphi}$, then this segment has the equation $t = u + iy$ $(x \leqslant u < \infty)$, and on the segment we have

$$|e^{z^2-t^2}| = e^{x^2-u^2}, \qquad |t|^{-(2n+3)} \leqslant |z|^{-(2n+3)}, \qquad |t| \leqslant u\sec\varphi.$$

Therefore

$$|r_n(z)| \leqslant \frac{1\cdot 3\cdots(2n+1)}{2^n|z|^{2n+2}}\sec\varphi\int_x^\infty e^{x^2-u^2}u\,du = \frac{1\cdot 3\cdots(2n+1)}{(2|z|^2)^{n+1}}\sec\varphi,$$

which implies

$$|r_n(z)| \leqslant \frac{1\cdot 3\cdots(2n+1)}{(2|z|^2)^{n+1}}\sec\varphi \leqslant \frac{1\cdot 3\cdots(2n+1)}{(2|z|^2)^{n+1}\sin\delta}. \qquad (2.2.3)$$

It follows from (2.2.3) that as $|z| \to \infty$ the product $z^{2n}r_n(z)$ converges uniformly to zero in the indicated sector, i.e.,

$$1 - \Phi(z) \approx \frac{e^{-z^2}}{\sqrt{\pi}z}\left[1 + \sum_{n=1}^{\infty}(-1)^n \frac{1\cdot 3\cdots(2n-1)}{(2z^2)^n}\right], \qquad |z| \to \infty, \quad |\arg z| \leqslant \frac{\pi}{2} - \delta. \tag{2.2.4}$$

Thus the series on the right is the asymptotic series (see Sec. 1.4) of the function $1 - \Phi(z)$, and a bound on the error committed in approximating $1 - \Phi(z)$ by the sum of a finite number of terms of the series is given by (2.2.3). For positive real z this error does not exceed the first neglected term in absolute value.

An asymptotic representation of the probability integral in the sector

$$\frac{\pi}{2} + \delta \leqslant \arg z \leqslant \frac{3\pi}{2} - \delta$$

can be obtained from (2.2.1) by using the relation $\Phi(z) = -\Phi(-z)$, but the construction of an asymptotic representation in the sector

$$\frac{\pi}{2} - \delta \leqslant \arg z \leqslant \frac{\pi}{2} + \delta$$

requires a separate argument [cf. (2.3.5)].

2.3. The Probability Integral of Imaginary Argument. The Function $F(z)$

In the applications, one often encounters the case where the argument of the probability integral is a complex number. We now examine the particularly simple case where $z = ix$ is a pure imaginary. Choosing a segment of the imaginary axis as the path of integration, and making the substitution $t = iu$, we find from (2.1.1) that

$$\frac{\Phi(ix)}{i} = \frac{2}{\sqrt{\pi}}\int_0^x e^{u^2}\,du. \tag{2.3.1}$$

The integral in the right increases without limit as $x \to \infty$, and therefore it is more convenient to consider the function

$$F(z) = e^{-z^2}\int_0^z e^{u^2}\,du, \tag{2.3.2}$$

which remains bounded for all real z. In the general case of complex z, $F(z)$ is an entire function, and the choice of the path of integration in (2.3.2) is completely arbitrary.

To expand $F(z)$ in power series, we note that $F(z)$ satisfies the linear differential equation

$$F'(z) + 2zF(z) = 1, \qquad (2.3.3)$$

with initial condition $F(0) = 0$. Substituting the series

$$F(z) = \sum_{k=0}^{\infty} a_k z^k$$

into (2.3.3), and comparing coefficients of identical powers of z, we obtain the recurrence relation

$$a_0 = 0, \quad a_1 = 1, \quad (k+1)a_{k+1} + 2a_{k-1} = 0.$$

After some simple calculations, this leads to the expansion

$$F(z) = \sum_{k=0}^{\infty} \frac{(-1)^k 2^k z^{2k+1}}{1\cdot 3\cdots(2k+1)}, \qquad |z| < \infty. \qquad (2.3.4)$$

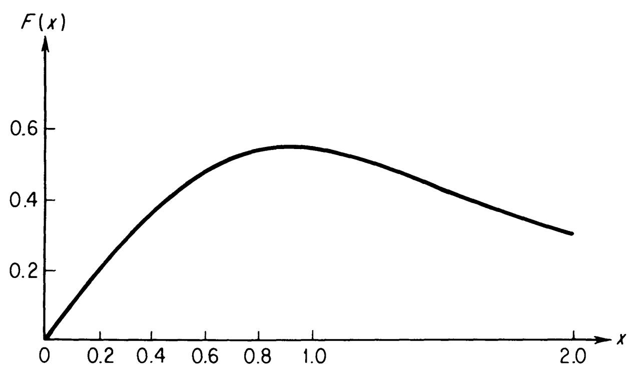


FIGURE 2

To study the behavior of $F(z)$ as $z \to \infty$ for real z, we apply L'Hospital's rule twice to the ratio

$$\frac{2z \displaystyle\int_0^z e^{u^2}\,du}{e^{z^2}},$$

and then use (2.3.2) to deduce that

$$\lim_{z\to\infty} 2zF(z) = 1,$$

i.e.,

$$F(z) \approx \frac{1}{2z}, \qquad z \to \infty. \qquad (2.3.5)$$

In Figure 2 we show the graph of the function $F(z)$ for real $z \geqslant 0$. The maximum of the function occurs at $z = 0.924\ldots$ and equals $F_{\max} = 0.541\ldots$. The function $F(z)$ comes up in the theory of propagation of electromagnetic waves along the earth's surface, and in other problems of mathematical physics.

2.4. The Probability Integral of Argument $\sqrt{i}x$. The Fresnel Integrals

Another interesting case from the standpoint of the applications occurs when the argument of the probability integral is the complex number

$$z = \sqrt{i}x = \frac{x}{\sqrt{2}}(1 + i),$$

where x is real. In this case, we choose the path of integration in (2.1.1) to be a segment of the bisector of the angle between the real and imaginary axes. Then, using the formula $t = \sqrt{i}u$ to introduce the new real variable u, we find from (1.1.1) that

$$\frac{\Phi(\sqrt{i}x)}{\sqrt{i}} = \frac{2}{\sqrt{\pi}}\int_0^x e^{-iu^2}\,du = \frac{2}{\sqrt{\pi}}\int_0^x \cos u^2\,du - i\frac{2}{\sqrt{\pi}}\int_0^x \sin u^2\,du. \tag{2.4.1}$$

The integrals on the right can be expressed in terms of the functions

$$C(z) = \int_0^z \cos\frac{\pi t^2}{2}\,dt, \qquad S(z) = \int_0^z \sin\frac{\pi t^2}{2}\,dt, \tag{2.4.2}$$

where the integration is along any path joining the origin to the point $t = z$. The functions $C(z)$ and $S(z)$ are known as the *Fresnel integrals*. Since the integrands in (2.4.2) are entire functions of the complex variable t, the choice of the path of integration does not matter, and both $C(z)$ and $S(z)$ are entire functions of z.

For real $z = x$, the Fresnel integrals are real, with the graphs shown in Figure 3. Both $C(x)$ and $S(x)$ vanish for $x = 0$, and have an oscillatory character, as follows from the formulas

$$C'(x) = \cos\frac{\pi x^2}{2}, \qquad S'(x) = \sin\frac{\pi x^2}{2},$$

which show that $C(x)$ has extrema at $x = \pm\sqrt{2n+1}$, while $S(x)$ has extrema at $x = \pm\sqrt{2n}$ $(n = 0, 1, 2, \ldots)$. The largest maxima are $C(1) = 0.779893\ldots$ and $S(\sqrt{2}) = 0.713972\ldots$, respectively. As $x \to \infty$, each of the functions approaches the limit

$$C(\infty) = S(\infty) = \tfrac{1}{2},$$

as implied by the familiar formula[3]

$$\int_0^\infty \cos t^2\,dt = \int_0^\infty \sin t^2\,dt = \frac{\sqrt{\pi}}{2\sqrt{2}}. \tag{2.4.3}$$

[3] D. V. Widder, *op. cit.*, p. 382.

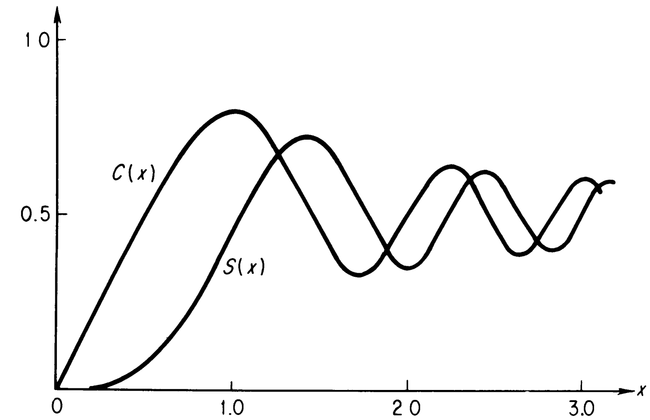


FIGURE 3

Replacing the trigonometric functions in the integrands in (2.4.2) by their power series expansions, and integrating term by term, we obtain the following series expansions for the Fresnel integrals, which converge for arbitrary z:

$$C(z) = \int_0^z \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!}\left(\frac{\pi t^2}{2}\right)^{2k} dt = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!}\left(\frac{\pi}{2}\right)^{2k} \frac{z^{4k+1}}{4k+1},$$

$$S(z) = \int_0^z \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!}\left(\frac{\pi t^2}{2}\right)^{2k+1} dt = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!}\left(\frac{\pi}{2}\right)^{2k+1} \frac{z^{4k+3}}{4k+3}. \tag{2.4.4}$$

The relation between the Fresnel integrals and the probability integral is given by the formula

$$\begin{aligned} C(z) \pm iS(z) &= \int_0^z e^{\pm \pi i t^2/2}\, dt = \sqrt{\frac{2}{\pi}}\, e^{\pm \pi i/4} \int_0^{\sqrt{\pi/2}\, z e^{\mp \pi i/4}} e^{-u^2} du \\ &= \frac{1}{\sqrt{2}}\, e^{\pm \pi i/4} \Phi\left(\sqrt{\frac{\pi}{2}}\, z e^{\mp \pi i/4}\right), \end{aligned} \tag{2.4.5}$$

which implies

$$C(z) = \frac{1}{2\sqrt{2}}\left[e^{\pi i/4}\Phi\left(\sqrt{\frac{\pi}{2}}\, z e^{-\pi i/4}\right) + e^{-\pi i/4}\Phi\left(\sqrt{\frac{\pi}{2}}\, z e^{\pi i/4}\right)\right],$$

$$S(z) = \frac{1}{2i\sqrt{2}}\left[e^{\pi i/4}\Phi\left(\sqrt{\frac{\pi}{2}}\, z e^{-\pi i/4}\right) - e^{-\pi i/4}\Phi\left(\sqrt{\frac{\pi}{2}}\, z e^{\pi i/4}\right)\right]. \tag{2.4.6}$$

Using (2.4.6), we can derive the properties of $C(z)$ and $S(z)$ from the corresponding properties of the probability integral. In particular, the results of

Sec. 2.2 lead to the following asymptotic representations of the Fresnel integrals, valid for large $|z|$ in the sector $|\arg z| \leqslant \frac{1}{4}\pi - \delta$,

$$
\begin{aligned}
C(z) &= \frac{1}{2} - \frac{1}{\pi z}\left[B(z)\cos\frac{\pi z^2}{2} - A(z)\sin\frac{\pi z^2}{2}\right], \\
S(z) &= \frac{1}{2} - \frac{1}{\pi z}\left[A(z)\cos\frac{\pi z^2}{2} + B(z)\sin\frac{\pi z^2}{2}\right],
\end{aligned}
\tag{2.4.7}
$$

where

$$
\begin{aligned}
A(z) &= \sum_{k=0}^{N} \frac{(-1)^k \alpha_{2k}}{(\pi z^2)^{2k}} + O(|z|^{-4N-4}), \\
B(z) &= \sum_{k=0}^{N} \frac{(-1)^k \alpha_{2k+1}}{(\pi z^2)^{2k+1}} + O(|z|^{-4N-6}), \\
\alpha_k &= 1\cdot 3\cdots(2k-1), \quad \alpha_0 = 1.
\end{aligned}
$$

The Fresnel integrals come up in various branches of physics and engineering, e.g., diffraction theory, theory of vibrations (see Sec. 2.7), etc. Many integrals of a more complicated type can be expressed in terms of the functions $C(z)$ and $S(z)$.

2.5. Application to Probability Theory

By a *normal* (or *Gaussian*) random variable with mean m and standard deviation σ is meant a random variable ξ such that the probability of ξ lying in the interval $[x, x + dx]$ is given by the expression[4,5]

$$
\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2}\, dx. \tag{2.5.1}
$$

Then the probability

$$
\mathbf{P}\{a \leqslant \xi - m \leqslant b\} \tag{2.5.2}
$$

that $\xi - m$ lies in the interval $[a, b]$ is just the integral

$$
\begin{aligned}
\frac{1}{\sqrt{2\pi}\sigma}\int_{a+m}^{b+m} e^{-(x-m)^2/2\sigma^2}\, dx &= \frac{1}{\sqrt{\pi}}\int_{a/\sqrt{2}\sigma}^{b/\sqrt{2}\sigma} e^{-t^2}\, dt \\
&= \frac{1}{2}\left[\Phi\left(\frac{b}{\sqrt{2}\sigma}\right) - \Phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right],
\end{aligned}
\tag{2.5.3}
$$

---

[4] As usual, $[a, b]$ denotes the *closed* interval $a \leqslant x \leqslant b$, and (a, b) the *open* interval $a < x < b$.

[5] See W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. 1*, second edition, John Wiley and Sons, Inc., New York (1957). If $x_1, \ldots, x_n$ are the results of measurements of ξ, where n is large, then

$$
m \approx \frac{1}{n}\sum_{k=1}^{n} x_k, \qquad \sigma^2 \approx \frac{1}{n}\sum_{k=1}^{n} (x - m)^2.
$$

where $\Phi(x)$ is the probability integral. As one would expect, (2.5.2) equals 1 if $a = -\infty$, $b = \infty$.

Setting $a = -\delta$, $b = \delta$, we obtain the probability that $|\xi - m|$ does not exceed δ:

$$\mathbf{P}\{|\xi - m| \leqslant \delta\} = \Phi\left(\frac{\delta}{\sqrt{2}\sigma}\right). \tag{2.5.4}$$

Then the probability that $|\xi - m|$ exceeds δ is just

$$\mathbf{P}\{|\xi - m| > \delta\} = 1 - \Phi\left(\frac{\delta}{\sqrt{2}\sigma}\right). \tag{2.5.5}$$

The value $\delta = \delta_p$ for which (2.5.4) and (2.5.5) are equal is called the *probable error*, and clearly satisfies the equation

$$\Phi\left(\frac{\delta_p}{\sqrt{2}\sigma}\right) = \frac{1}{2}.$$

Using a table of the function $\Phi(x)$ to solve this equation,[6] we find that

$$\delta_p = 0.67449\sigma.$$

Example. *With standard deviation* 1 mm, *a machine produces parts of average length* 10 cm. *Find the probability that a part is of length* 10 cm *to within a tolerance of* 1 mm.

The required probability is

$$\mathbf{P}\{|\xi - 10| \leqslant 0.1\} = \Phi\left(\frac{1}{\sqrt{2}}\right) \approx 0.683,$$

i.e., some 68 percent of the parts satisfy the specified tolerance. In this case, the probable error is approximately 0.7 mm.

2.6. Application to the Theory of Heat Conduction. Cooling of the Surface of a Heated Object

Consider the following problem in the theory of heat conduction: An object occupying the half-space $x \geqslant 0$ is initially heated to temperature T_0. It then cools off by radiating heat through its surface $x = 0$ into the surrounding medium which is at zero temperature. We want to find the temperature $T(x, t)$ of the object as a function of position x and time t.

Let the object have thermal conductivity k, heat capacity c, density ρ and

---

[6] See E. Jahnke and F. Emde, *Tables of Higher Functions*, sixth edition, revised by F. Lösch, McGraw-Hill Book Co., New York (1960), p. 31.

emissivity λ, and let $\tau = kt/c\rho$. Then our problem reduces to the solution of the equation of heat conduction

$$\frac{\partial T}{\partial \tau} = \frac{\partial^2 T}{\partial x^2}, \tag{2.6.1}$$

subject to the initial condition

$$T|_{\tau=0} = T_0 \tag{2.6.2}$$

and the boundary conditions[7]

$$\left(\frac{\partial T}{\partial x} - hT\right)\bigg|_{x=0} = 0, \qquad T|_{x\to\infty} = T_0, \tag{2.6.3}$$

where $h = \lambda/k > 0$.

To solve the problem, we introduce the *Laplace transform* $\bar{T} = \bar{T}(x, p)$ of $T = T(x, \tau)$, defined by the formula

$$\bar{T} = \int_0^\infty e^{-p\tau}\, T\, d\tau, \qquad \operatorname{Re} p > 0. \tag{2.6.4}$$

A system of equations determining $\bar{T}$ can be obtained from (2.6.1–3) if we multiply the first and third equations by $e^{-p\tau}$ and integrate from 0 to ∞, taking the second equation into account. The result is

$$\begin{aligned} &\frac{d^2\bar{T}}{dx^2} = p\bar{T} - T_0, \\ &\frac{d\bar{T}}{dx} - h\bar{T}|_{x=0} = 0, \qquad \bar{T}|_{x\to\infty} = \frac{T_0}{p}. \end{aligned} \tag{2.6.5}$$

The system (2.6.5) has the solution

$$\bar{T} = \frac{T_0}{p}\left(1 - \frac{h}{h + \sqrt{p}}\, e^{-\sqrt{p}x}\right), \qquad \operatorname{Re} p > 0, \quad \operatorname{Re}\sqrt{p} > 0. \tag{2.6.6}$$

We can now solve for T by inverting (2.6.4). This can be done either by using a table of Laplace transforms,[8] or by applying the *Fourier-Mellin inversion theorem*,[9] which states that

$$T = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} e^{p\tau}\, \bar{T}\, dp, \tag{2.6.7}$$

where a is a constant greater than the real part of all the singular points of $\bar{T}$.

---

[7] For the derivation of equations (2.6.1, 3), see G. P. Tolstov, *Fourier Series* (translated by R. A. Silverman), Prentice-Hall, Inc., Englewood Cliffs, N.J. (1962), Chap. 9, Secs. 20 and 24.

[8] See A. Erdélyi, W. Magnus, F. Oberhettinger and F. G. Tricomi, *Tables of Integral Transforms, Volume 1* (of two volumes), Chaps. 4–5, McGraw-Hill Book Co., New York (1954). This two-volume set (based, in part, on notes left by Harry Bateman) will henceforth be referred to as the Bateman Manuscript Project, *Tables of Integral Transforms*.

[9] H. S. Carslaw and J. C. Jaeger, *Operational Methods in Applied Mathematics*, second edition, Oxford University Press, London (1953), Chap. 4, Secs. 28–31.

The quantity of greatest interest is the *surface temperature* of the object. Setting $x = 0$ in (2.6.6), we find that

$$\bar{T}|_{x=0} = \frac{T_0}{\sqrt{p}(\sqrt{p}+h)} = T_0\left(\frac{1}{p-h^2} - \frac{h}{p-h^2}\frac{1}{\sqrt{p}}\right). \tag{2.6.8}$$

The simplest way to solve (2.6.8) for the original function $T|_{x=0}$ is to use the *convolution theorem*,[10] which states that if $\bar{f}_1$ and $\bar{f}_2$ are the Laplace transforms of f_1 and f_2, then $\bar{f} = \bar{f}_1\bar{f}_2$ is the Laplace transform of the function

$$f(\tau) = \int_0^\tau f_1(t)f_2(\tau - t)\,dt. \tag{2.6.9}$$

Since it is easily verified that

$$\bar{f}_1 = \frac{h}{\sqrt{p}}, \qquad \bar{f}_2 = \frac{1}{p-h^2}$$

are the Laplace transforms of

$$f_1 = \frac{h}{\sqrt{\pi\tau}}, \qquad f_2 = e^{h^2\tau},$$

(2.6.9) implies

$$T|_{x=0} = T_0\left(e^{h^2\tau} - \frac{h}{\sqrt{\pi}}\int_0^\infty e^{h^2(\tau-t)}\frac{dt}{\sqrt{t}}\right) = T_0 e^{h^2\tau}\left(1 - \frac{2}{\sqrt{\pi}}\int_0^{h\sqrt{\tau}} e^{-s^2}\,ds\right),$$

i.e.,

$$T|_{x=0} = T_0 e^{h^2\tau}[1 - \Phi(h\sqrt{\tau})], \tag{2.6.10}$$

where $\Phi(x)$ is the probability integral. It follows from the asymptotic formula (2.2.1) that for large τ the surface temperature falls off like $1/\sqrt{\tau}$:

$$T|_{x=0} \approx \frac{T_0}{h\sqrt{\pi\tau}}, \qquad \tau \to \infty. \tag{2.6.11}$$

The temperature inside the object ($x \neq 0$) can also be expressed in closed form in terms of the probability integral.

2.7. Application to the Theory of Vibrations. Transverse Vibrations of an Infinite Rod under the Action of a Suddenly Applied Concentrated Force

Consider an infinite rod of linear density ρ and Young's modulus E, lying along the positive x-axis. Let I be the moment of inertia of a cross section of the rod about a horizontal axis through the center of mass of the section, and let $\tau = \sqrt{EI/\rho}\,t$. Suppose the end $x = 0$ satisfies a sliding condition, while

[10] H. S. Carslaw and J. C. Jaeger, *op. cit.*, Chap. 4, Sec. 33.

the end $x = \infty$ is clamped, and suppose a constant force Q is suddenly applied at the end $x = 0$. Then the displacement $u = u(x, t)$ at an arbitrary point $x \geqslant 0$ of the rod is described by the system of equations[11]

$$\frac{\partial^2 u}{\partial \tau^2} + \frac{\partial^4 u}{\partial x^4} = 0,$$

$$u|_{\tau=0} = \left.\frac{\partial u}{\partial \tau}\right|_{\tau=0} = 0, \tag{2.7.1}$$

$$\left.\frac{\partial u}{\partial x}\right|_{x=0} = 0, \qquad \left.\frac{\partial^3 u}{\partial x^3}\right|_{x=0} = \frac{Q}{EI}, \qquad u|_{x\to\infty} = 0, \quad \left.\frac{\partial u}{\partial x}\right|_{x\to\infty} = 0.$$

To solve this system, we use the Laplace transform, as in the preceding section. Writing

$$\bar{u} = \int_0^\infty e^{-p\tau} u \, d\tau, \qquad \operatorname{Re} p > 0, \tag{2.7.2}$$

we obtain the following equations for $\bar{u}$:

$$\frac{d^4 \bar{u}}{dx^4} + p^2 \bar{u} = 0,$$

$$\left.\frac{d\bar{u}}{dx}\right|_{x=0} = 0, \qquad \left.\frac{d^3 \bar{u}}{dx^3}\right|_{x=0} = \frac{Q}{EIp}, \tag{2.7.3}$$

$$\bar{u}|_{x\to\infty} = 0, \qquad \left.\frac{d\bar{u}}{dx}\right|_{x\to\infty} = 0.$$

Simple calculations then show that

$$\bar{u} = \frac{Q}{2EIp^2 i}\left(\frac{e^{-\sqrt{-pi}\,x}}{\sqrt{-pi}} - \frac{e^{-\sqrt{pi}\,x}}{\sqrt{pi}}\right), \qquad \operatorname{Re} p > 0, \quad \operatorname{Re}\sqrt{\pm pi} > 0. \tag{2.7.4}$$

To find u, we again use the convolution theorem. Since[12]

$$\bar{f}_1 = \frac{Q}{EIp^2}, \qquad \bar{f}_2 = \frac{1}{2i}\left(\frac{e^{-\sqrt{-pi}\,x}}{\sqrt{-pi}} - \frac{e^{-\sqrt{pi}\,x}}{\sqrt{pi}}\right)$$

are the Laplace transforms of

$$f_1 = \frac{Q}{EI}\tau, \qquad f_2 = \frac{1}{\sqrt{2\pi\tau}}\left(\sin\frac{x^2}{4\tau} + \cos\frac{x^2}{4\tau}\right),$$

(2.6.9) implies

$$u = \frac{Q}{EI\sqrt{2\pi}}\int_0^\tau \left(\sin\frac{x^2}{4t} + \cos\frac{x^2}{4t}\right)\frac{\tau - t}{\sqrt{t}}\,dt = \frac{Qx\tau}{EI} f\left(\frac{x}{2\sqrt{\tau}}\right), \tag{2.7.5}$$

[11] See R. E. D. Bishop and D. C. Johnson, *The Mechanics of Vibration*, Cambridge University Press, London (1960), p. 285.

[12] Bateman Manuscript Project, *Tables of Integral Transforms, Vol. 1*, formula (27), p. 146 or formula (6), p. 246.

where

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty (\sin y^2 + \cos y^2) \frac{1 - (x^2/y^2)}{y^2} \, dy. \tag{2.7.6}$$

The function $f(x)$ can be expressed in terms of the Fresnel integrals $C(z)$ and $S(z)$, introduced in Sec. 2.4. In fact, integrating (2.7.6) by parts twice, we find that

$$f(x) = \left(1 + \frac{2}{3}x^2\right)\left[\frac{1}{2} - C\left(\sqrt{\frac{2}{\pi}}\,x\right)\right] - \left(1 - \frac{2}{3}x^2\right)\left[\frac{1}{2} - S\left(\sqrt{\frac{2}{\pi}}\,x\right)\right]$$
$$+ \frac{2}{3\sqrt{2\pi}}\left[(1 + x^2)\frac{\sin x^2}{x} + (1 - x^2)\frac{\cos x^2}{x}\right]. \tag{2.7.7}$$

PROBLEMS

1. Show that the functions

$$\varphi(z) = \frac{\sqrt{\pi}}{2} e^{z^2} \Phi(z)$$

satisfies the differential equation $\varphi' - 2z\varphi = 1$, and use this fact to derive the expansion

$$\Phi(z) = \frac{2z}{\sqrt{\pi}} e^{-z^2} \sum_{k=0}^{\infty} \frac{(2z^2)^k}{1\cdot 3\cdots(2k+1)}, \qquad |z| < \infty.$$

2. Using formula (2.4.5) and the result of Problem 1, derive the following expansions of the Fresnel integrals

$$C(x) = x\left[\alpha(x)\cos\frac{\pi x^2}{2} + \beta(x)\sin\frac{\pi x^2}{2}\right],$$
$$S(x) = x\left[\alpha(x)\sin\frac{\pi x^2}{2} - \beta(x)\cos\frac{\pi x^2}{2}\right],$$

where

$$\alpha(x) = \sum_{k=0}^{\infty} \frac{(-1)^k(\pi x^2)^{2k}}{1\cdot 3\cdots(4k+1)}, \qquad \beta(x) = \sum_{k=0}^{\infty} \frac{(-1)^k(\pi x^2)^{2k+1}}{1\cdot 3\cdots(4k+3)}.$$

3. Use integration by parts to show that

$$\int \Phi(x)\, dx = x\Phi(x) + \frac{1}{\sqrt{\pi}} e^{-x^2} + C.$$

4. Let $\overline{\Phi}$ be the Laplace transform of the probability integral, i.e.,

$$\overline{\Phi}(p) = \int_0^\infty e^{-px}\, \Phi(x)\, dx.$$

Prove that

$$\overline{\Phi}(p) = \frac{1}{4} e^{p^2/4}\left[1 - \Phi\left(\frac{p}{2}\right)\right].$$

5. Derive the integral representations

$$F(z) = \int_0^\infty e^{-t^2} \sin 2zt \, dt, \qquad \Phi(z) = \frac{2}{\pi} \int_0^\infty e^{-t^2} \frac{\sin 2zt}{t} \, dt.$$

Hint. Replace $\sin 2zt$ by its power series expansion and integrate term by term.

6. Derive the following integral representations for the square of the probability integral:

$$\Phi^2(z) = 1 - \frac{4}{\pi} \int_0^1 \frac{e^{-z^2(1+t^2)}}{1 + t^2} \, dt,$$

$$[1 - \Phi(z)]^2 = \frac{4}{\pi} \int_1^\infty \frac{e^{-z^2(1+t^2)}}{1 + t^2} \, dt, \qquad |\arg z| \leqslant \frac{\pi}{4}.$$

Hint. Represent $\Phi^2(z)$ as a double integral over the region $0 \leqslant s \leqslant z$, $0 \leqslant t \leqslant z$, and transform to polar coordinates.

7. Derive the formulas

$$1 - \Phi(z) = \frac{2}{\sqrt{\pi}} e^{-z^2} \int_0^\infty e^{-t^2 - 2zt} \, dt,$$

$$[1 - \Phi(z)]^2 = \frac{4}{\sqrt{\pi}} e^{-2z^2} \int_0^\infty e^{-t^2 - 2\sqrt{2}zt} \Phi(t) \, dt.$$

Hint. The second formula is obtained from the first after introducing new variables $\alpha = s + t$, $\beta = st$ in the double integral over the region $0 \leqslant s < \infty$, $0 \leqslant t \leqslant s$.

8. Prove that

$$C^2(z) \pm S^2(z) = \frac{2}{\pi} \int_0^1 \frac{\sin \frac{\pi z^2}{2} (1 \mp t^2)}{1 \mp t^2} \, dt.$$

9. Prove that

$$C(x) = \int_0^{\pi x^2/2} J_{-1/2}(t) \, dt, \qquad S(x) = \int_0^{\pi x^2/2} J_{1/2}(t) \, dt,$$

where $J_\nu(x)$ is the Bessel function of order ν (see Sec. 5.8).

2.8 Cylinder Functions

5

CYLINDER FUNCTIONS: THEORY

5.1. Introductory Remarks

By a *cylinder function* we mean a solution of the second-order linear differential equation

$$u'' + \frac{1}{z}u' + \left(1 - \frac{\nu^2}{z^2}\right)u = 0, \tag{5.1.1}$$

where z is a complex variable and ν is a parameter which can take arbitrary real or complex values. Equation (5.1.1), called *Bessel's equation of order* ν, is encountered in studying the boundary value problems of potential theory for cylindrical domains (see Sec. 6.3), which explains the origin of the term *cylinder function*. Certain special kinds of cylinder functions are known in the literature as *Bessel functions*, and this term is sometimes applied to the whole class of cylinder functions.

The cylinder functions, with their manifold applications, have been studied in great detail, and extensive tables of such functions are available. These functions are among the most important special functions, with very diverse applications to physics, engineering and mathematical analysis itself, ranging from abstract number theory and theoretical astronomy to concrete problems of physics and engineering. Some of these applications, mainly from the field of mathematical physics, will be considered in Chapter 6. The present chapter is devoted to a brief exposition of the elementary theory of cylinder functions. The reader who wishes to go further in his study of these functions should consult the special literature devoted to the subject (see the Bibliography on p. 300), notably the classic treatise by Watson,[1] to which we will make frequent reference.

[1] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, second edition, Cambridge University Press, London (1962).

5.2. Bessel Functions of Nonnegative Integral Order

In many applied problems, one need only consider a special class of cylinder functions, corresponding to the case where the parameter ν in equation (5.1.1) is a nonnegative integer n. This case is much simpler than the case of arbitrary ν, and will serve to introduce the general theory.

We begin by showing that one of the solutions of Bessel's equation

$$u'' + \frac{1}{z} u' + \left(1 - \frac{n^2}{z^2}\right) u = 0, \qquad n = 0, 1, 2, \ldots \tag{5.2.1}$$

is the function $u_1 = J_n(z)$, known as the *Bessel function of the first kind of order n*, and defined for arbitrary z by the series

$$J_n(z) = \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{n+2k}}{k!(n+k)!}, \qquad |z| < \infty. \tag{5.2.2}$$

Using the ratio test, we easily verify that this series converges in the whole complex plane, and hence represents an entire function of z. Suppose we denote the left-hand side of (5.2.1) by $l(u)$, and introduce the abbreviated notation

$$\alpha_k = \frac{(-1)^k}{2^{n+2k} k!(n+k)!}$$

for the coefficients of the series (5.2.2). Then we have

$$\begin{aligned} l(u_1) &= \sum_{k=0}^{\infty} [(n+2k)(n+2k-1) + (n+2k) - n^2] \alpha_k z^{n+2k-2} + \sum_{k=0}^{\infty} \alpha_k z^{n+2k} \\ &= \sum_{k=1}^{\infty} 4\alpha_k k(n+k) z^{n+2k-2} + \sum_{k=0}^{\infty} \alpha_k z^{n+2k} \\ &= \sum_{k=0}^{\infty} [4\alpha_{k+1}(k+1)(n+k+1) + \alpha_k] z^{n+2k}, \end{aligned}$$

and therefore $l(u_1) \equiv 0$, since the expression in brackets vanishes. Thus $J_n(z)$ satisfies Bessel's equation (5.2.1), i.e., $J_n(z)$ is a cylinder function. The simplest functions of this kind are the Bessel functions of orders zero and one:

$$\begin{aligned} J_0(z) &= 1 - \frac{(z/2)^2}{(1!)^2} + \frac{(z/2)^4}{(2!)^2} - \frac{(z/2)^6}{(3!)^2} + \cdots, \\ J_1(z) &= \frac{z}{2}\left[1 - \frac{(z/2)^2}{1!2!} + \frac{(z/2)^4}{2!3!} - \frac{(z/2)^6}{3!4!} + \cdots\right]. \end{aligned} \tag{5.2.3}$$

We now show that the Bessel functions of higher order can be expressed in terms of the two functions $J_0(z)$ and $J_1(z)$. Assuming that n is a positive integer, we multiply the series (5.2.2) by z^n and then differentiate with respect to z. This gives

$$\frac{d}{dz}[z^n J_n(z)] = \sum_{k=0}^{\infty} \frac{(-1)^k(2n+2k)}{2^{n+2k}k!(n+k)!} z^{2n+2k-1}$$

$$= z^n \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(n-1+k)!} \left(\frac{z}{2}\right)^{n-1+2k} = z^n J_{n-1}(z),$$

or

$$\frac{d}{dz}[z^n J_n(z)] = z^n J_{n-1}(z), \qquad n = 1, 2, \ldots \tag{5.2.4}$$

Similarly, multiplying (5.2.2) by z^{-n}, we find that

$$\frac{d}{dz}[z^{-n} J_n(z)] = -z^{-n} J_{n+1}(z), \qquad n = 0, 1, 2, \ldots \tag{5.2.5}$$

Performing the differentiation in (5.2.4–5) and dividing by the factors $z^{\pm n}$, we arrive at the formulas

$$J_n'(z) + \frac{n}{z} J_n(z) = J_{n-1}(z), \qquad J_n'(z) - \frac{n}{z} J_n(z) = -J_{n+1}(z), \tag{5.2.6}$$

which immediately imply the following *recurrence relations* satisfied by the Bessel functions:

$$J_{n-1}(z) + J_{n+1}(z) = \frac{2n}{z} J_n(z), \qquad n = 1, 2, \ldots \tag{5.2.7}$$

$$J_{n-1}(z) - J_{n+1}(z) = 2J_n'(z), \qquad n = 1, 2, \ldots \tag{5.2.8}$$

Repeated application of (5.2.7) allows us to express a Bessel function of arbitrary order $\nu = n$ ($n = 0, 1, 2 \ldots$) in terms of $J_0(z)$ and $J_1(z)$, thereby greatly simplifying the effort needed to calculate tables of Bessel functions. Formula (5.2.8) allows us to express derivatives of Bessel functions in terms of other Bessel functions. For $n = 0$, (5.2.8) should be replaced by

$$J_0'(z) = -J_1(z) \tag{5.2.9}$$

[in keeping with (5.2.5)], which is an immediate consequence of the formulas (5.2.3).

The Bessel functions of the first kind $J_n(z)$ are simply related to the coefficients of the Laurent expansion of the function[2]

$$w(z, t) = e^{\frac{1}{2}z(t-t^{-1})} = \sum_{n=-\infty}^{\infty} c_n(z)t^n, \qquad 0 < |t| < \infty. \tag{5.2.10}$$

[2] Regarded as a function of t, $w(z, t)$ is analytic in the annulus $0 < \delta \leqslant t \leqslant A < \infty$, and therefore this expansion exists.

To calculate the coefficients $c_n(z)$, we multiply the power series

$$e^{zt/2} = 1 + \frac{(z/2)}{1!}t + \frac{(z/2)^2}{2!}t^2 + \cdots,$$

$$e^{-z/2t} = 1 - \frac{(z/2)}{1!}\frac{1}{t} + \frac{(z/2)^2}{2!}\frac{1}{t^2} + \cdots,$$

and then combine terms containing identical powers of t. As a result, we obtain

$$\begin{aligned} c_n(z) &= J_n(z), \qquad n = 0, 1, 2, \ldots, \\ c_n(z) &= (-1)^n J_{-n}(z), \qquad n = -1, -2, \ldots, \end{aligned} \tag{5.2.11}$$

which implies

$$w(z, t) = e^{\frac{1}{2}z(t-t^{-1})} = J_0(z) + \sum_{n=1}^{\infty} J_n(z)[t^n + (-1)^n t^{-n}], \quad 0 < |t| < \infty. \tag{5.2.12}$$

The function $w(z, t)$ is called the *generating function* of the Bessel functions of integral order, and formula (5.2.12) plays an important role in the theory of these functions.

To find a general solution of Bessel's equation (5.2.1), thereby obtaining an arbitrary cylinder function of integral order $\nu = n$ ($n = 0, 1, 2, \ldots$), we must construct a second solution of (5.2.1) which is linearly independent of $J_n(z)$. For such a solution we choose $u_2 = Y_n(z)$, called the *Bessel function of the second kind*, which will be defined in Sec. 5.4. It will be shown in Sec. 5.5 that this definition leads to the series expansion

$$\begin{aligned} Y_n(z) = \frac{2}{\pi}J_n(z)\log\frac{z}{2} &- \frac{1}{\pi}\sum_{k=0}^{n-1}\frac{(n-k-1)!}{k!}\left(\frac{z}{2}\right)^{2k-n} \\ &- \frac{1}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k(z/2)^{n+2k}}{k!(n+k)!}[\psi(k+1) + \psi(k+n+1)], \end{aligned} \tag{5.2.13}$$

where

$$\psi(m+1) = -\gamma + 1 + \frac{1}{2} + \ldots + \frac{1}{m}, \qquad \psi(1) = -\gamma,$$

γ is Euler's constant (see Sec. 1.3), and in the case $n = 0$, the first sum in (5.2.13) should be set equal to zero. The function $Y_n(z)$ is analytic in the complex plane cut along the segment $[-\infty, 0]$, and becomes infinite as $z \to 0$. Thus, the general expression for the cylinder function of order $\nu = n$ is a linear combination of Bessel functions of the first and second kinds, i.e.,

$$u = Z_n(z) = AJ_n(z) + BY_n(z), \qquad n = 0, 1, 2, \ldots, \tag{5.2.14}$$

where A and B are constants.

5.3. Bessel Functions of Arbitrary Order

The Bessel functions considered in the preceding section are a special case of the more general Bessel functions of the first kind of *arbitrary* order ν. To define these functions, consider the series

$$\sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{\nu+2k}}{\Gamma(k+1)\Gamma(k+\nu+1)}, \tag{5.3.1}$$

where z is a complex variable belonging to the plane cut along the segment $[-\infty, 0]$, and ν is a parameter which can take arbitrary real or complex values.[3] It is easily seen that (5.3.1) converges for all z and ν, and that the convergence is uniform in each variable in the region $|z| \leqslant R$, $|\nu| \leqslant N$ (where R and N are arbitrarily large). This follows from the fact that starting from some sufficiently large k, the ratio of the absolute value of the $(k+1)$th term to that of the kth term equals

$$\frac{|z|^2}{4(k+1)|k+1+\nu|} \leqslant \frac{R^2}{4(k+1)(k+1-N)},$$

where the right-hand side is positive, independent of z and ν, and approaches zero as $k \to \infty$.[4] Since the terms of (5.3.1) are analytic functions of z in the plane cut along $[-\infty, 0]$, the sum of the series is an analytic function of z in the same region. We call this function the *Bessel function of the first kind of order* ν, and denote it by $J_\nu(z)$, i.e.,

$$J_\nu(z) = \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{\nu+2k}}{\Gamma(k+1)\Gamma(k+\nu+1)}, \qquad |z| < \infty, \quad |\arg z| < \pi. \tag{5.3.2}$$

To show that the function (5.3.2) satisfies Bessel's equation with parameter ν, we write

$$l(u) \equiv u'' + \frac{1}{z}u' + \left(1 - \frac{\nu^2}{z^2}\right)u = 0, \qquad u_1 = J_\nu(z),$$

and repeat the derivation given in Sec. 5.2,[5] obtaining

$$l(u_1) = \sum_{k=0}^{\infty} [4\alpha_{k+1}(k+1)(k+\nu+1) + \alpha_k] z^{\nu+2k},$$

[3] In general, the condition imposed on z is necessary for the function z^ν to be single-valued, but can be omitted if ν is an integer.

[4] A series of functions

$$\sum_{k=0}^{\infty} u_k(z)$$

converges uniformly in a domain D if

$$\left|\frac{u_{k+1}(z)}{u_k(z)}\right| \leqslant q < 1$$

for all z in D and $k \geqslant M$, where q is independent of z. See E. C. Titchmarsh, *op. cit.*, p. 4.

[5] Recall that a uniformly convergent series of analytic functions can be differentiated term by term.

where

$$\alpha_k = \frac{(-1)^k}{2^{\nu+2k}\Gamma(k+1)\Gamma(k+\nu+1)}.$$

Using (1.2.1), we see at once that $l(u_1) \equiv 0$.

Since for fixed z in the plane cut along the segment $[-\infty, 0]$, the terms of the series (5.3.2) are analytic functions of the variable ν (see Sec. 1.1), the fact that (5.3.2) is uniformly convergent implies that the Bessel function of the first kind is an entire function of *its order* ν. For integral $\nu = n$ ($n = 0, 1, 2, \ldots$), $\Gamma(k + \nu + 1) = (n + k)!$ and (5.3.2) reduces to (5.2.2). Therefore the functions defined in this section are the natural generalizations of those studied in the preceding section. For negative integral $\nu = -n$ ($n = 1, 2, \ldots.$), the first n terms of the series (5.3.2) vanish (see Sec. 1.2), and the series becomes

$$J_{-n}(z) = \sum_{k=n}^{\infty} \frac{(-1)^k (z/2)^{-n+2k}}{k!(k-n)!} = \sum_{s=0}^{\infty} \frac{(-1)^{n+s}(z/2)^{n+2s}}{(n+s)!s!},$$

and hence

$$J_{-n}(z) = (-1)^n J_n(z), \qquad n = 1, 2, \ldots \tag{5.3.3}$$

Thus, the Bessel functions of negative integral order differ only by sign from the corresponding functions of positive integral order. It follows that the expansion (5.2.12) can be written in the form

$$w(z, t) = e^{\frac{1}{2}z(t-t^{-1})} = \sum_{n=-\infty}^{\infty} J_n(z)t^n. \tag{5.3.4}$$

Many of the formulas derived earlier for Bessel functions of nonnegative integral order remain the same for Bessel functions of arbitrary order. For example,

$$\frac{d}{dz}[z^\nu J_\nu(z)] = z^\nu J_{\nu-1}(z), \qquad \frac{d}{dz}[z^{-\nu}J_\nu(z)] = -z^{-\nu}J_{\nu+1}(z), \tag{5.3.5}$$

$$J_{\nu-1}(z) + J_{\nu+1}(z) = \frac{2\nu}{z}J_\nu(z), \qquad J_{\nu-1}(z) - J_{\nu+1}(z) = 2J'_\nu(z), \tag{5.3.6}$$

generalize formulas (5.2.4–5, 7–8), and are proved in exactly the same way. We also have

$$\begin{aligned} \left(\frac{d}{z\,dz}\right)^m [z^\nu J_\nu(z)] &= z^{\nu-m}J_{\nu-m}(z), \\ \left(\frac{d}{z\,dz}\right)^m [z^{-\nu}J_\nu(z)] &= (-1)^m z^{-\nu-m}J_{\nu+m}(z), \end{aligned} \tag{5.3.7}$$

which are proved by repeated application of (5.3.6).

5.4. General Cylinder Functions. Bessel Functions of the Second Kind

By definition, a *cylinder function* is an arbitrary solution of the second-order linear differential equation

$$l(u) = u'' + \frac{1}{z}u' + \left(1 - \frac{\nu^2}{z^2}\right)u = 0, \tag{5.4.1}$$

and hence has the general form

$$u = Z_\nu(z) = C_1 u_1(z) + C_2 u_2(z), \tag{5.4.2}$$

where u_1 and u_2 are arbitrary linearly independent solutions of (5.4.1), and C_1, C_2 are constants which, in general, are arbitrary functions of the parameter ν. It is easy to obtain an expression for the general cylinder function in the case where ν is not an integer. In fact, choosing $u_1 = J_\nu(z)$, where $J_\nu(z)$ is the Bessel function defined in Sec. 5.3, we take the second function to be $u_2 = J_{-\nu}(z)$, which is also a solution of (5.4.1), since (5.4.1) does not change if ν is replaced by $-\nu$. For nonintegral ν, the asymptotic behavior of these solutions as $z \to 0$ is given by

$$u_1 \approx \frac{(z/2)^\nu}{\Gamma(1+\nu)}, \qquad u_2 \approx \frac{(z/2)^{-\nu}}{\Gamma(1-\nu)}, \tag{5.4.3}$$

and therefore these solutions are linearly independent.[6] Thus, the desired expression for the general cylinder function can be written as

$$u = Z_\nu(z) = C_1 J_\nu(z) + C_2 J_{-\nu}(z), \qquad \nu \neq 0, \pm 1, \pm 2, \ldots \tag{5.4.4}$$

If ν is an integer, then, because of (5.3.3), the particular solutions u_1 and u_2 are linearly dependent, and (5.4.4) is no longer a general solution of Bessel's equation (5.4.1). To obtain an expression for the general cylinder function which is suitable for arbitrary ν, we introduce the *Bessel functions of the second kind*, denoted by $Y_\nu(z)$ and defined by the formula

$$Y_\nu(z) = \frac{J_\nu(z)\cos \nu\pi - J_{-\nu}(z)}{\sin \nu\pi} \tag{5.4.5}$$

for arbitrary z belonging to the plane cut along the segment $[-\infty, 0]$.[7] For integral ν, the right-hand side of (5.4.5) becomes indeterminate [cf. (5.3.3)], and in this case we define $Y_n(z)$ as the limit

$$Y_n(z) = \lim_{\nu \to n} Y_\nu(z). \tag{5.4.6}$$

[6] This argument breaks down if ν is an integer (including zero).

[7] The function we denote by $Y_\nu(z)$ is sometimes denoted by $N_\nu(z)$ in the literature on Bessel functions.

Since both the numerator and denominator are entire functions of ν, and since

$$\frac{d}{d\nu}\sin \nu\pi = \pi \cos \nu\pi \neq 0 \qquad \text{if} \quad \nu = n,$$

this limit exists and can be calculated by L'Hospital's rule, application of which gives

$$Y_n(z) = \frac{1}{\pi}\left[\frac{\partial J_\nu(z)}{\partial \nu}\bigg|_{\nu=n} - (-1)^n \frac{\partial J_{-\nu}(z)}{\partial \nu}\bigg|_{\nu=n}\right]. \tag{5.4.7}$$

It follows from its definition that $Y_\nu(z)$ is an analytic function of z in the plane cut along $[-\infty, 0]$, and an entire function of the parameter ν for fixed z.

In view of (5.4.4), the fact that $Y_\nu(z)$ is a cylinder function, i.e., satisfies Bessel's equation (5.4.1), is obvious for nonintegral ν. To show that $Y_\nu(z)$ is a cylinder function for integral ν, we use the principle of analytic continuation, noting that since $l(Y_\nu)$ is an entire function of ν, $l(Y_\nu) \equiv 0$ for $\nu \neq n$ implies $l(Y_\nu)$ for all ν. The fact that the solutions $u_1 = J_\nu(z)$ and $u_2 = Y_\nu(z)$ are linearly independent follows from the linear independence of the solutions $J_\nu(z)$ and $J_{-\nu}(z)$ for nonintegral ν, and from a comparison of the behavior of u_1 and u_2 as $z \to 0$ [cf. (5.4.3) and (5.5.4), proved below] for integral ν. Thus, finally, the expression

$$u = Z_\nu(z) = C_1 J_\nu(z) + C_2 Y_\nu(z) \tag{5.4.8}$$

for the general cylinder function $Z_\nu(z)$ is suitable for arbitrary ν.

The Bessel functions of the second kind satisfy the same recurrence relations as the functions of the first kind, e.g.,

$$\begin{gathered}\frac{d}{dz}[z^\nu Y_\nu(z)] = z^\nu Y_{\nu-1}(z), \qquad \frac{d}{dz}[z^{-\nu} Y_\nu(z)] = -z^{-\nu} Y_{\nu+1}(z), \\ Y_{\nu-1}(z) + Y_{\nu+1}(z) = \frac{2\nu}{z} Y_\nu(z), \qquad Y_{\nu-1}(z) - Y_{\nu+1}(z) = 2Y_\nu'(z).\end{gathered} \tag{5.4.9}$$

For nonintegral ν, the validity of these formulas follows from the definition (5.4.5) and the corresponding formulas for $J_\nu(z)$. To obtain the same formulas for integral ν, we need only pass to the limit $\nu \to n$, observing that all the functions involved are continuous with respect to the index ν. We also note that (5.4.7) implies the relation

$$Y_{-n}(z) = (-1)^n Y_n(z), \qquad n = 0, 1, 2, \ldots, \tag{5.4.10}$$

which allows us to reduce the calculation of functions of negative integral order to that of functions of positive integral order.

By making changes of variables in Bessel's equation (5.4.1), we can easily obtain a number of other differential equations whose general solutions can be expressed in terms of cylinder functions. Of the various equations obtained in this way, those of greatest practical interest are

$$u'' + \frac{1-2\alpha}{z}u' + \left[(\beta\gamma z^{\gamma-1})^2 + \frac{\alpha^2 - \nu^2\gamma^2}{z^2}\right]u = 0,$$
$$u'' + \alpha z^\gamma u = 0, \tag{5.4.11}$$

with solutions

$$u = z^\alpha Z_\nu(\beta z^\gamma), \qquad u = z^{1/2} Z_{1/(\gamma+2)}\left(\frac{2\alpha^{1/2}}{\gamma+2} z^{1+(\gamma/2)}\right), \tag{5.4.12}$$

where $Z_\nu(z)$ denotes an arbitrary cylinder function.

5.5. Series Expansion of the Function $Y_n(z)$

To derive a series expansion of the function $Y_n(z)$, we use the expansion (5.3.2) to calculate the derivatives with respect to the index ν which appear in (5.4.7). Because of (5.4.10), we need only consider the case $\nu = n$ ($n = 0, 1, 2, \ldots$). Since, as already shown, the series (5.3.1) converges uniformly in ν, we can differentiate it term by term, obtaining[8]

$$\left.\frac{\partial J_\nu(z)}{\partial \nu}\right|_{\nu=n} = \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{n+2k}}{k!(n+k)!}\left[\log\frac{z}{2} - \psi(k+n+1)\right],$$

where

$$\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$$

is the logarithmic derivative of the gamma function (see Sec. 1.3). Similarly, we have

$$\frac{\partial J_{-\nu}(z)}{\partial \nu} = \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{-\nu+2k}}{k!\Gamma(k-\nu+1)}\left[-\log\frac{z}{2} + \psi(k-\nu+1)\right].$$

For $k = 0, 1, 2, \ldots, n-1$,

$$\Gamma(k-\nu+1) \to \infty, \qquad \psi(k-\nu+1) \to \infty$$

as $\nu \to n$, so that the first n terms of the last series become indeterminate. However, using familiar formulas from the theory of the gamma function [see (1.2.2, 4) and (1.3.4)], we find that

$$\lim_{\nu\to n} \frac{\psi(k-\nu+1)}{\Gamma(k-\nu+1)} = \lim_{\nu\to n}\left[\Gamma(\nu-k)\sin\pi(\nu-k)\,\frac{\psi(\nu-k) + \pi\cot\pi(\nu-k)}{\pi}\right]$$
$$= (-1)^{n-k}(n-k-1)!, \qquad k = 0, 1, \ldots, n-1,$$

[8] The passage to the limit $\nu \to n$ behind the summation sign is legitimate, since a series obtained by term-by-term differentiation of a uniformly convergent series of analytic functions is itself uniformly convergent.

and therefore

$$\left.\frac{\partial J_{-\nu}(z)}{\partial \nu}\right|_{\nu=n} = (-1)^n \sum_{k=0}^{n-1} \frac{(n-k-1)!}{k!} \left(\frac{z}{2}\right)^{2k-n}$$
$$+ (-1)^n \sum_{p=0}^{\infty} \frac{(-1)^p}{(n+p)!p!} \left[-\log\frac{z}{2} + \psi(p+1)\right] \left(\frac{z}{2}\right)^{2p+n},$$

where we have introduced the new summation index $p = k - n$.

It now follows from (5.4.7) that the desired expansion of the function $Y_n(z)$ is

$$Y_n(z) = -\frac{1}{\pi} \sum_{k=0}^{n-1} \frac{(n-k-1)!}{k!} \left(\frac{z}{2}\right)^{2k-n}$$
$$+ \frac{1}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{n+2k}}{k!(n+k)!} \left[2\log\frac{z}{2} - \psi(k+1) - \psi(k+n+1)\right]$$
$$|\arg z| < \pi, \quad n = 0, 1, 2, \ldots, \qquad (5.5.1)$$

where the first sum should be set equal to zero if $n = 0$ [cf. (5.2.13)]. According to (1.3.6–7), the values of the logarithmic derivative of the gamma function are given by

$$\psi(1) = -\gamma, \qquad \psi(m+1) = -\gamma + 1 + \frac{1}{2} + \cdots + \frac{1}{m}, \qquad m = 1, 2, \ldots, \qquad (5.5.2)$$

where $\gamma = 0.57721566\ldots$ is Euler's constant. Using (5.2.2), we can write the expansion (5.5.1) in a somewhat different form:

$$Y_n(z) = \frac{2}{\pi} J_n(z) \log\frac{z}{2} - \frac{1}{\pi} \sum_{k=0}^{n-1} \frac{(n-k-1)!}{k!} \left(\frac{z}{2}\right)^{2k-n}$$
$$- \frac{1}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{n+2k}}{k!(n+k)!} [\psi(k+1) + \psi(k+n+1)]. \qquad (5.5.3)$$

Finally, we note that (5.5.1) implies the asymptotic representations

$$Y_0(z) \approx \frac{2}{\pi} \log\frac{z}{2}, \qquad z \to 0$$
$$Y_n(z) \approx -\frac{(n-1)!}{\pi} \left(\frac{z}{2}\right)^{-n}, \qquad z \to 0, \quad n = 1, 2, \ldots, \qquad (5.5.4)$$

which show that $Y_n(z)$ becomes infinite as $z \to 0$.

5.6. Bessel Functions of the Third Kind

Next we discuss still another class of cylinder functions, i.e., the *Bessel functions of the third kind* or *Hankel functions*, denoted by $H_\nu^{(1)}(z)$ and $H_\nu^{(2)}(z)$.

These functions are defined in terms of the Bessel functions of the first and second kinds by the formulas

$$H_\nu^{(1)}(z) = J_\nu(z) + iY_\nu(z), \qquad H_\nu^{(2)}(z) = J_\nu(z) - iY_\nu(z), \tag{5.6.1}$$

where ν is arbitrary and z is any point of the plane cut along the segment $[-\infty, 0]$. The motivation for introducing the functions (5.6.1) is that these linear combinations of $J_\nu(z)$ and $Y_\nu(z)$ have very simple asymptotic expressions for large $|z|$ (see Sec. 5.11) and are frequently encountered in the applications.

It follows from (5.6.1) that the Hankel functions are entire functions of ν, and analytic functions of z in the plane cut along $[-\infty, 0]$. Clearly, the functions $H_\nu^{(1)}(z)$ and $H_\nu^{(2)}(z)$ are linearly independent of each other, and each is linearly independent of $J_\nu(z)$. Therefore we can write the general solution of Bessel's equation (5.4.1) in any of the forms

$$\begin{aligned} u = Z_\nu(z) &= A_1 J_\nu(z) + A_2 H_\nu^{(1)}(z) \\ &= B_1 J_\nu(z) + B_2 H_\nu^{(2)}(z) = D_1 H_\nu^{(1)}(z) + D_2 H_\nu^{(2)}(z), \end{aligned} \tag{5.6.2}$$

where $A_1, \ldots, D_2$ are arbitrary constants, as well as in the form (5.4.8).

Since the Hankel functions are linear combinations of the functions $J_\nu(z)$ and $Y_\nu(z)$, they satisfy the same recurrence relations as these functions, e.g.,

$$\begin{gathered} \frac{d}{dz}[z^\nu H_\nu^{(p)}(z)] = z^\nu H_{\nu-1}^{(p)}(z), \qquad \frac{d}{dz}[z^{-\nu} H_\nu^{(p)}(z)] = -z^{-\nu} H_{\nu+1}^{(p)}(z), \\ H_{\nu-1}^{(p)}(z) + H_{\nu+1}^{(p)}(z) = \frac{2\nu}{z} H_\nu^{(p)}(z), \quad H_{\nu-1}^{(p)}(z) - H_{\nu+1}^{(p)}(z) = 2\frac{dH_\nu^{(p)}(z)}{dz}, \end{gathered} \tag{5.6.3}$$

where $p = 1, 2$. Using (5.4.5) to eliminate $Y_\nu(z)$ from (5.6.1), we obtain

$$H_\nu^{(1)}(z) = \frac{J_{-\nu}(z) - e^{-\nu\pi i} J_\nu(z)}{i \sin \nu\pi}, \quad H_\nu^{(2)}(z) = \frac{e^{\nu\pi i} J_\nu(z) - J_{-\nu}(z)}{i \sin \nu\pi}, \tag{5.6.4}$$

which imply the important formulas

$$H_{-\nu}^{(1)}(z) = e^{\nu\pi i} H_\nu^{(1)}(z), \qquad H_{-\nu}^{(2)}(z) = e^{-\nu\pi i} H_\nu^{(2)}(z). \tag{5.6.5}$$

5.7. Bessel Functions of Imaginary Argument

In the applications, one frequently encounters two functions $I_\nu(z)$ and $K_\nu(z)$, which are closely related to the Bessel functions. Let D be the complex plane cut along the negative real axis. Then, for all z in D, $I_\nu(z)$ and $K_\nu(z)$ are defined by the formulas

$$I_\nu(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{\nu+2k}}{\Gamma(k+1)\Gamma(k+\nu+1)}, \qquad |z| < \infty, \quad |\arg z| < \pi, \tag{5.7.1}$$

$$K_\nu(z) = \frac{\pi}{2} \frac{I_{-\nu}(z) - I_\nu(z)}{\sin \nu\pi}, \qquad |\arg z| < \pi, \qquad \nu \neq 0, \pm 1, \pm 2, \ldots \tag{5.7.2}$$

where, for integral $\nu = n$,

$$K_n(z) = \lim_{\nu \to n} K_\nu(z), \qquad n = 0, \pm 1, \pm 2, \dots \tag{5.7.3}$$

Repeating the considerations of Secs. 5.3–4, we find that $I_\nu(z)$ and $K_\nu(z)$ are analytic functions of z for all z in D, and entire functions of ν.

The functions $I_\nu(z)$ and $K_\nu(z)$ are simply related to the Bessel functions of argument $ze^{\pm \pi i/2}$. If

$$-\pi < \arg z < \frac{\pi}{2}, \quad \text{i.e.,} \quad -\frac{\pi}{2} < \arg (ze^{\pi i/2}) < \pi,$$

then (5.3.2) implies

$$J_\nu(ze^{\pi i/2}) = e^{\nu\pi i/2} \sum_{k=0}^{\infty} \frac{(z/2)^{\nu+2k}}{\Gamma(k+1)\Gamma(k+\nu+1)} = e^{\nu\pi i/2} I_\nu(z),$$

so that

$$I_\nu(z) = e^{-\nu\pi i/2} J_\nu(ze^{\pi i/2}), \qquad -\pi < \arg z < \frac{\pi}{2}. \tag{5.7.4}$$

Similarly, according to (5.6.4), for the same values of z we have

$$\begin{aligned} H_\nu^{(1)}(ze^{\pi i/2}) &= \frac{J_{-\nu}(ze^{\pi i/2}) - e^{-\nu\pi i} J_\nu(ze^{\pi i/2})}{i \sin \nu\pi} \\ &= \frac{e^{-\nu\pi i/2} I_{-\nu}(z) - e^{-\nu\pi i/2} I_\nu(z)}{i \sin \nu\pi} = \frac{2}{\pi i} e^{-\nu\pi i/2} K_\nu(z), \end{aligned}$$

and hence

$$K_\nu(z) = \frac{\pi i}{2} e^{\nu\pi i/2} H_\nu^{(1)}(ze^{\pi i/2}), \qquad -\pi < \arg z < \frac{\pi}{2}. \tag{5.7.5}$$

On the other hand, if

$$-\frac{\pi}{2} < \arg z < \pi, \qquad -\pi < \arg (ze^{-\pi i/2}) < \frac{\pi}{2},$$

then it is easily verified that

$$I_\nu(z) = e^{\nu\pi i/2} J_\nu(ze^{-\pi i/2}), \qquad K_\nu(z) = -\frac{\pi i}{2} e^{-\nu\pi i/2} H_\nu^{(2)}(ze^{-\pi i/2}). \tag{5.7.6}$$

Because of (5.7.4–6), $I_\nu(z)$ and $K_\nu(z)$ are often called *Bessel functions of imaginary argument*. However, this term is not too fortunate, and instead we will usually refer to $I_\nu(z)$ as the *modified Bessel function of the first kind* and to $K_\nu(z)$ as *Macdonald's function*.[9]

[9] $K_\nu(z)$ is called the *modified Bessel function of the third kind* in the Bateman Manuscript Project, *Higher Transcendental Functions, Vol. 2*, p. 5.

It is an immediate consequence of the formulas just derived that $I_\nu(z)$ and $K_\nu(z)$ are linearly independent solutions of the differential equation

$$u'' + \frac{1}{z}u' - \left(1 + \frac{\nu^2}{z^2}\right)u = 0, \tag{5.7.7}$$

which differs from Bessel's equation only by the sign of one term, and goes into Bessel's equation if we make the substitution $z = \pm it$. Equation (5.7.7) is often encountered in mathematical physics, and its general solution, for arbitrary ν, can be written in the form

$$u = C_1 I_\nu(z) + C_2 K_\nu(z). \tag{5.7.8}$$

The functions $I_\nu(z)$ and $K_\nu(z)$ satisfy simple recurrence relations, e.g.

$$\begin{aligned}
&\frac{d}{dz}[z^\nu I_\nu(z)] = z^\nu I_{\nu-1}(z), \qquad \frac{d}{dz}[z^{-\nu} I_\nu(z)] = z^{-\nu} I_{\nu+1}(z),\\
&\frac{d}{dz}[z^\nu K_\nu(z)] = -z^\nu K_{\nu-1}(z), \qquad \frac{d}{dz}[z^{-\nu} K_\nu(z)] = -z^{-\nu} K_{\nu+1}(z)\\
&I_{\nu-1}(z) + I_{\nu+1}(z) = 2I'_\nu(z), \qquad I_{\nu-1}(z) - I_{\nu+1}(z) = \frac{2\nu}{z} I_\nu(z),\\
&K_{\nu-1}(z) + K_{\nu+1}(z) = -2K'_\nu(z), \qquad K_{\nu-1}(z) - K_{\nu+1}(z) = -\frac{2\nu}{z} K_\nu(z).
\end{aligned} \tag{5.7.9}$$

The recurrence relations involving $I_\nu(z)$ are proved by substituting from (5.7.1). Then, using these formulas and (5.7.2), we derive the corresponding formulas involving $K_\nu(z)$ for nonintegral ν. Finally, we extend the results to the case of integral ν by using the continuity of $K_\nu(z)$ with respect to the index ν.

Two other useful formulas are

$$\begin{aligned}
I_{-n}(z) &= I_n(z), \qquad n = 0, \pm 1, \pm 2, \ldots,\\
K_{-\nu}(z) &= K_\nu(z),
\end{aligned} \tag{5.7.10}$$

where the first follows from (5.7.1) if we note that the first n terms of the expansion vanish if $\nu = -n$, while the second is an immediate consequence of the definition (5.7.2).

Using (5.7.3) and the method of Sec. 5.5, we can derive a series expansion of the function $K_n(z)$. The result of the calculations is

$$\begin{aligned}
K_n(z) = \frac{1}{2}\sum_{k=0}^{n-1} \frac{(-1)^k(n-k-1)!}{k!}\left(\frac{z}{2}\right)^{2k-n}\\
+ \frac{1}{2}(-1)^{n-1}\sum_{k=0}^{\infty}\frac{(z/2)^{2k+n}}{k!(k+n)!}\left[2\log\frac{z}{2} - \psi(k+1) - \psi(k+n+1)\right],\\
|\arg z| < \pi, \quad n = 0, 1, 2, \ldots,
\end{aligned} \tag{5.7.11}$$

where $\psi(z)$ is the logarithmic derivative of the gamma function [whose values can be found from (5.5.2)], and the first sum should be set equal to zero if $n = 0$. We note that (5.7.11) implies the asymptotic representations

$$\begin{aligned} K_0(z) &\approx \log\frac{2}{z}, \qquad z \to 0, \\ K_n(z) &\approx \frac{1}{2}(n-1)!\left(\frac{z}{2}\right)^{-n}, \qquad z \to 0, \quad n = 1, 2, \ldots, \end{aligned} \tag{5.7.12}$$

which show that $K_n(z)$ becomes infinite as $z \to 0$.

5.8. Cylinder Functions of Half-Integral Order

We now consider the special class of cylinder functions of order $n + \frac{1}{2}$ ($n = 0, \pm 1, \pm 2, \ldots$). In this case, the cylinder functions can be expressed in terms of elementary functions. To see this, we first find the values of the functions $J_{\pm 1/2}(z)$. Setting $\nu = \pm\frac{1}{2}$ in (5.3.1) and using the duplication formula (1.2.3) for the gamma function, we obtain

$$\begin{aligned} J_{1/2}(z) &= \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{2k+\frac{1}{2}}}{\Gamma(k+1)\Gamma(k+\frac{3}{2})} \\ &= \left(\frac{2z}{\pi}\right)^{1/2} \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k}}{\Gamma(2k+2)} = \left(\frac{2}{\pi z}\right)^{1/2} \sin z \end{aligned} \tag{5.8.1}$$

and similarly,

$$J_{-1/2}(z) = \left(\frac{2}{\pi z}\right)^{1/2} \cos z. \tag{5.8.2}$$

The fact that any Bessel function of the first kind of half-integral order can be expressed in terms of elementary functions now follows from the recurrence relation

$$J_{\nu-1}(z) + J_{\nu+1}(z) = \frac{2\nu}{z} J_\nu(z)$$

[see (5.3.6)], repeated application of which gives

$$\begin{aligned} J_{3/2}(z) &= \frac{1}{z} J_{1/2}(z) - J_{-1/2}(z) = \left(\frac{2}{\pi z}\right)^{1/2}\left[\frac{\sin z}{z} - \cos z\right], \\ J_{-3/2}(z) &= -\left(\frac{2}{\pi z}\right)^{1/2}\left[\sin z + \frac{\cos z}{z}\right], \end{aligned}$$

and so on. Using (5.3.7), we can write the general expression for $J_{n+\frac{1}{2}}(z)$ in terms of elementary functions. For example, setting $\nu = \frac{1}{2}$ in the second of the formulas (5.3.7) and taking account of (5.8.1), we find that

$$J_{n+\frac{1}{2}}(z) = (-1)^n \left(\frac{2}{\pi}\right)^{1/2} z^{n+\frac{1}{2}} \left(\frac{d}{z\,dz}\right)^n \frac{\sin z}{z}, \qquad n = 0, 1, 2, \ldots \tag{5.8.3}$$

To derive the corresponding formulas for Bessel functions of the second and third kinds, we start from the expressions (5.4.5) and (5.6.4) of these functions in terms of Bessel functions of the first kind, and use (5.8.1–2). For example,

$$Y_{1/2}(z) = -J_{-1/2}(z) = -\left(\frac{2}{\pi z}\right)^{1/2}\cos z,$$
$$H_{1/2}^{(1)}(z) = -i\left(\frac{2}{\pi z}\right)^{1/2} e^{iz}, \qquad H_{1/2}^{(2)}(z) = i\left(\frac{2}{\pi z}\right)^{1/2} e^{-iz}, \tag{5.8.4}$$

and so on.

Finally, we note that

$$I_{1/2}(z) = \left(\frac{2}{\pi z}\right)^{1/2}\sinh z, \quad I_{-1/2}(z) = \left(\frac{2}{\pi z}\right)^{1/2}\cosh z, \quad K_{1/2}(z) = \left(\frac{\pi}{2z}\right)^{1/2} e^{-z}, \tag{5.8.5}$$

where the formulas for general index $n + \frac{1}{2}$ are obtained from (5.8.5) and the recurrence relations (5.7.9). It has been shown by Liouville that the case of half-integral order is the only case where the cylinder functions reduce to elementary functions.

5.9. Wronskians of Pairs of Solutions of Bessel's Equation

By the *Wronskian* of a pair $u_1(z)$, $u_2(z)$ of solutions of a linear homogeneous second-order differential equation is meant the determinant

$$W\{u_1(z), u_2(z)\} = \begin{vmatrix} u_1(z) & u_2(z) \\ u_1'(z) & u_2'(z) \end{vmatrix},$$

where the prime denotes differentiation with respect to the independent variable z. The solutions u_1 and u_2 are linearly independent if and only if the Wronskian does not vanish identically.[10] We now calculate the Wronskians of various pairs of solutions of Bessel's equation

$$u'' + \frac{1}{z}u' + \left(1 - \frac{\nu^2}{z^2}\right)u = 0,$$

thereby obtaining a number of formulas which are useful in the applications. In particular, these formulas show that the solutions in question are linearly independent, a fact proved earlier by other means.

To calculate the Wronskian, we write the equations for u_1 and u_2 in the form

$$\frac{d}{dz}(zu_1') + \left(z - \frac{\nu^2}{z}\right)u_1 = 0, \qquad \frac{d}{dz}(zu_2') + \left(z - \frac{\nu^2}{z}\right)u_2 = 0,$$

[10] E. A. Coddington, *op. cit.*, Theorem 6, p. 111.

and then subtract the first equation multiplied by u_2 from the second equation multiplied by u_1. The result is

$$\frac{d}{dz}\left[zW\{u_1(z), u_2(z)\}\right] = 0,$$

which implies

$$W\{u_1(z), u_2(z)\} = \frac{C}{z},$$

where C is a constant, independent of z, whose value can be determined, for example, from the relation

$$C = \lim_{z\to 0} zW\{u_1(z), u_2(z)\}.$$

In particular, choosing $u_1 = J_\nu(z)$, $u_2 = J_{-\nu}(z)$, where ν is not an integer, and using the expansion (5.3.2) and formulas (1.2.1–2) from the theory of the gamma function, we find that

$$C = \lim_{z\to 0} \frac{-2\nu}{\Gamma(1+\nu)\Gamma(1-\nu)}\left[1 + O(z)^2\right] = -\frac{2\sin\nu\pi}{\pi},$$

which implies

$$W\{J_\nu(z), J_{-\nu}(z)\} = -\frac{2\sin\nu\pi}{\pi z}. \tag{5.9.1}$$

The validity of (5.9.1) for integral ν follows by continuity, and we have $W \equiv 0$, as must be expected. The Wronskians of other pairs of solutions of Bessel's equation can be found in the same way, or else they can be deduced from (5.9.1) and the relations (5.4.5), (5.6.4). We always begin by considering the case of nonintegral ν, and then use continuity to extend the result to arbitrary values of ν. In this way, we find that

$$W\{J_\nu(z),\ Y_\nu(z)\} = \frac{2}{\pi z}, \tag{5.9.2}$$

$$W\{J_\nu(z), H_\nu^{(2)}(z)\} = -\frac{2i}{\pi z}, \tag{5.9.3}$$

$$W\{H_\nu^{(1)}(z), H_\nu^{(2)}(z)\} = -\frac{4i}{\pi z}, \tag{5.9.4}$$

and so on. For the Bessel functions of imaginary argument we have

$$W\{I_\nu(z), K_\nu(z)\} = -\frac{1}{z}. \tag{5.9.5}$$

5.10. Integral Representations of the Cylinder Functions

The cylinder functions have simple integral representations in terms of definite integrals and contour integrals containing z as a parameter. The representations by contour integrals have greater generality, and are usually valid in larger regions of values of the argument z and parameter ν than the representations by definite integrals, but the latter are more frequently encountered in the applications. Therefore we will be primarily concerned with representations by definite integrals.[11]

One of the simplest integral representations of the Bessel functions is due to Poisson. Consider the identity

$$\frac{1}{\Gamma(k+\nu+1)} = \frac{1}{\Gamma(k+\frac{1}{2})\,\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1} t^{2k}(1-t^2)^{\nu-\frac{1}{2}}\,dt, \qquad \operatorname{Re}\nu > -\tfrac{1}{2}, \tag{5.10.1}$$

implied by (1.5.6). Substituting (5.10.1) into the expansion (5.3.2) and reversing the order of integration and summation,[12] we obtain

$$\begin{aligned} J_\nu(z) &= \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{\nu+2k}}{\Gamma(k+1)} \frac{1}{\Gamma(k+\frac{1}{2})\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1} t^{2k}(1-t^2)^{\nu-\frac{1}{2}}\,dt \\ &= \frac{(z/2)^\nu}{\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1}(1-t^2)^{\nu-\frac{1}{2}}\,dt \sum_{k=0}^{\infty}\frac{(-1)^k(zt)^{2k}}{2^{2k}\Gamma(k+1)\Gamma(k+\frac{1}{2})} \\ &= \frac{(z/2)^\nu}{\Gamma(\frac{1}{2})\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1}(1-t^2)^{\nu-\frac{1}{2}}\cos zt\,dt, \end{aligned} \tag{5.10.2}$$

where we have used the duplication formula (1.2.3) for the gamma function:

$$2^{2k}\Gamma(k+1)\Gamma(k+\tfrac{1}{2}) = \Gamma(\tfrac{1}{2})\Gamma(2k+1) = \Gamma(\tfrac{1}{2})(2k)!.$$

Thus

$$J_\nu(z) = \frac{(z/2)^\nu}{\Gamma(\frac{1}{2})\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1}(1-t^2)^{\nu-\frac{1}{2}}\cos zt\,dt, \qquad \operatorname{Re}\nu > -\tfrac{1}{2}, \quad |\arg z| < \pi, \tag{5.10.3}$$

or equivalently,

$$J_\nu(z) = \frac{(z/2)^\nu}{\Gamma(\frac{1}{2})\Gamma(\nu+\frac{1}{2})}\int_{0}^{\pi}\cos(z\cos\theta)\sin^{2\nu}\theta\,d\theta, \quad \operatorname{Re}\nu > -\tfrac{1}{2}, \quad |\arg z| < \pi, \tag{5.10.4}$$

where we have made the substitution $t = \cos\theta$.

---

[11] The reader with a special interest in integral representations of cylinder functions should consult G. N. Watson, *op. cit.*, Chap. 6.

[12] To justify reversing the order of integration and summation, we note that

$$\sum_{k=0}^{\infty}\frac{|z/2|^{\nu+2k}}{\Gamma(k+1)}\frac{1}{\Gamma(k+\frac{1}{2})\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1}t^{2k}(1-t^2)^{\nu-\frac{1}{2}}\,dt = \sum_{k=0}^{\infty}\frac{|z/2|^{\nu+2k}}{\Gamma(k+1)\Gamma(k+\nu+1)} \equiv I_\nu(|z|) < \infty,$$

if $\operatorname{Re}\nu > -\frac{1}{2}$.

To obtain another important integral representation of $J_\nu(z)$, we start from the formula

$$\frac{1}{\Gamma(k+\nu+1)} = \frac{1}{2\pi i}\int_C e^s s^{-(k+\nu+1)}\,ds, \tag{5.10.5}$$

proved in Problem 9, p. 15, where C is the contour shown in Fig. 13. Substituting (5.10.5) into (5.3.2), we find that

$$\begin{aligned} J_\nu(z) &= \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{\nu+2k}}{\Gamma(k+1)} \frac{1}{2\pi i}\int_C e^s s^{-(k+\nu+1)}\,ds \\ &= \left(\frac{z}{2}\right)^\nu \frac{1}{2\pi i}\int_C e^s s^{-\nu-1}\,ds \sum_{k=0}^{\infty} \frac{(-1)^k (z^2/4s)^k}{\Gamma(k+1)} \\ &= \left(\frac{z}{2}\right)^\nu \frac{1}{2\pi i}\int_C e^{s-(z^2/4s)} s^{-\nu-1}\,ds, \end{aligned} \tag{5.10.6}$$

where reversing the order of integration and summation is again easily justified by an absolute convergence argument. Assuming temporarily that z is a positive real number and setting $s = zt/2$, we can write (5.10.6) in the form

$$J_\nu(z) = \frac{1}{2\pi i}\int_{C'} e^{\frac{1}{2}z(t-t^{-1})} t^{-\nu-1}\,dt, \tag{5.10.7}$$

where C' is a contour resembling C. By the principle of analytic continuation, this result is valid in the whole region $|\arg z| < \pi/2$. Writing $t = \rho e^{i\theta}$ and choosing the radius of the circular part of C' to be 1, we have

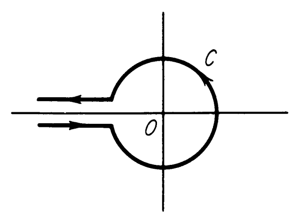


FIGURE 13

$$J_\nu(z) = \frac{1}{\pi}\int_0^\pi \cos(z\sin\theta - \nu\theta)\,d\theta - \frac{\sin\nu\pi}{\pi}\int_1^\infty e^{-\frac{1}{2}z(\rho-\rho^{-1})}\rho^{-\nu-1}\,d\rho,$$

which, after the substitution $\rho = e^\alpha$, becomes

$$J_\nu(z) = \frac{1}{\pi}\int_0^\pi \cos(z\sin\theta - \nu\theta)\,d\theta - \frac{\sin\nu\pi}{\pi}\int_0^\infty e^{-z\sinh\alpha-\nu\alpha}\,d\alpha, \qquad \operatorname{Re} z > 0, \tag{5.10.8}$$

where ν is arbitrary. In the case $\nu = n$ ($n = 0, \pm 1, \pm 2, \dots$), the second term on the right vanishes, and (5.10.8) takes a simpler form.

In many cases, one can derive integral representations of Bessel functions of the second and third kinds from the corresponding integral representations of Bessel functions of the first kind, by using formulas (5.4.5) and (5.6.4).

For example, if $\operatorname{Re} z > 0$ and ν is nonintegral, it follows from (5.4.5) and (5.10.8) that

$$Y_\nu(z) = \frac{\cot \nu\pi}{\pi}\int_0^\pi \cos(z \sin\theta - \nu\theta)\, d\theta - \frac{\cos \nu\pi}{\pi}\int_0^\infty e^{-z\sinh\alpha - \nu\alpha}\, d\alpha$$
$$- \frac{\csc\nu\pi}{\pi}\int_0^\pi \cos(z\sin\theta + \nu\theta)\, d\theta - \frac{1}{\pi}\int_0^\infty e^{-z\sinh\alpha + \nu\alpha}\, d\alpha.$$

Replacing θ by $\pi - \theta$ in the third integral on the right, we find after some simple calculations that

$$Y_\nu(z) = \frac{1}{\pi}\int_0^\pi \sin(z\sin\theta - \nu\theta)\, d\theta - \frac{1}{\pi}\int_0^\infty e^{-z\sinh\alpha}(e^{\nu\alpha} + e^{-\nu\alpha}\cos\nu\pi)\, d\alpha. \tag{5.10.9}$$

In proving (5.10.9), it was assumed that ν is nonintegral, but the formula holds for arbitrary ν by the principle of analytic continuation, since both sides are entire functions of ν.

Integral representations of the Hankel functions can be obtained by using (5.10.8–9) and the definitions (5.6.1). For example, if $\operatorname{Re} z > 0$,

$$\begin{aligned} H_\nu^{(1)}(z) = J_\nu(z) + iY_\nu(z) &= \frac{1}{\pi}\int_0^\pi e^{i(z\sin\theta - \nu\theta)}\, d\theta \\ &\quad + \frac{1}{\pi i}\int_0^\infty e^{-z\sinh\alpha}[e^{\nu\alpha} + e^{-\nu(\alpha+\pi i)}]\, d\alpha \\ &= \frac{1}{\pi i}\int_{-\infty}^0 e^{z\sinh\alpha - \nu\alpha}\, d\alpha + \frac{1}{\pi i}\int_{\theta=0}^\pi e^{z\sinh i\theta - \nu i\theta}\, d(i\theta) \\ &\quad + \frac{1}{\pi i}\int_{\alpha=0}^\infty e^{z\sinh(\alpha+\pi i) - \nu(\alpha+\pi i)}\, d(\alpha + \pi i), \end{aligned}$$

which, after the substitution $t = \alpha + i\theta$, reduces to

$$H_\nu^{(1)}(z) = \frac{1}{\pi i}\int_{C_1} e^{z\sinh t - \nu t}\, dt, \qquad \operatorname{Re} z > 0, \tag{5.10.10}$$

where C_1 is the contour shown in Fig. 14(a). Similarly,

$$H_\nu^{(2)}(z) = -\frac{1}{\pi i}\int_{C_2} e^{z\sinh t - \nu t}\, dt, \qquad \operatorname{Re} z > 0 \tag{5.10.11}$$

where C_2 is the contour shown in Fig. 14(b). Thus (5.10.10) and (5.10.11) are the same, except for the choice of the contour of integration. Substituting $t = u \pm \frac{1}{2}\pi i$ into (5.10.10–11), we find that

$$H_\nu^{(1)}(z) = \frac{e^{-\nu\pi i/2}}{\pi i}\int_{D_1} e^{iz\cosh u - \nu u}\, du, \qquad \operatorname{Re} z > 0, \tag{5.10.12}$$

$$H_\nu^{(2)}(z) = -\frac{e^{\nu\pi i/2}}{\pi i}\int_{D_2} e^{-iz\cosh u - \nu u}\, du, \qquad \operatorname{Re} z > 0, \tag{5.10.13}$$

where the paths of integration D_1 and D_2 are shown in Figure 15.

To further transform these integrals, we assume temporarily that z is a positive real number and that the parameter ν is confined to the strip $-1 < \operatorname{Re} \nu < 1$. Then, according to Cauchy's integral theorem, the integral

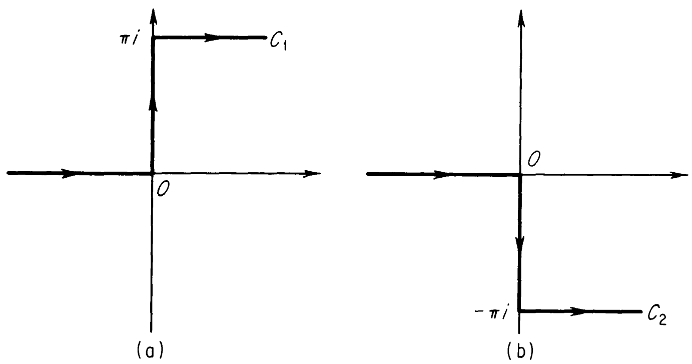


FIGURE 14

along the left-hand part of the broken line D_1 (or D_2), up to the point $u = 0$, can be replaced by an integral along the negative real axis, and the integral

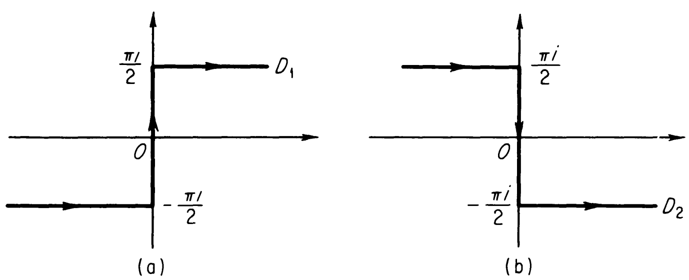


FIGURE 15

along the right-hand part of the broken line can be replaced by an integral along the positive real axis.[13] Thus formulas (5.10.12–13) become

$$H_\nu^{(1)}(z) = \frac{e^{-\nu\pi i/2}}{\pi i} \int_{-\infty}^{\infty} e^{iz \cosh u - \nu u}\, du, \tag{5.10.14}$$

$$H_\nu^{(2)}(z) = -\frac{e^{\nu\pi i/2}}{\pi i} \int_{-\infty}^{\infty} e^{-iz \cosh u - \nu u}\, du, \tag{5.10.15}$$

[13] It is easily verified that the integral along the vertical segment needed to complete each contour to which we apply Cauchy's integral theorem approaches zero as the segment is moved indefinitely far to the left (or to the right) of the imaginary axis. To show that the condition $-1 < \operatorname{Re} \nu < 1$ guarantees the convergence of (5.10, 14–15), consider the substitution $y = e^u$.

where $z > 0$, $-1 < \operatorname{Re} z < 1$. Using the principle of analytic continuation, we easily see that (5.10.14) remains valid for $0 \leqslant \arg z < \pi$, while (5.10.15) remains valid for $-\pi < \arg z \leqslant 0$, since in each case both sides of (5.10.14–15) are analytic functions of z in the indicated region. Moreover, the condition $-1 < \operatorname{Re} \nu < 1$ can be dropped if $\operatorname{Im} z > 0$ in (5.10.14), or if $\operatorname{Im} z < 0$ in (5.10.15). Finally, therefore, we have the integral representations

$$H_\nu^{(1)}(z) = \frac{e^{-\nu\pi i/2}}{\pi i} \int_{-\infty}^{\infty} e^{iz\cosh u - \nu u}\, du, \qquad \operatorname{Im} z > 0, \quad (5.10.16)$$

$$H_\nu^{(2)}(z) = -\frac{e^{\nu\pi i/2}}{\pi i} \int_{-\infty}^{\infty} e^{-iz\cosh u - \nu u}\, du, \qquad \operatorname{Im} z < 0, \quad (5.10.17)$$

where ν is arbitrary.

Formulas (5.10.16–17) are the basic integral representations of the Hankel functions. Other integral representations of the Hankel functions, useful in the applications, can be derived by making suitable transformations of the integrals in (5.10.16–17). For example, consider formula (5.10.16), let $\operatorname{Re} \nu > -\frac{1}{2}$, and for the time being assume that $\arg z = \pi/2$, so that $-iz$ is positive. According to (1.5.1),

$$y^{-\nu-\frac{1}{2}} = \frac{1}{\Gamma(\nu+\frac{1}{2})} \int_0^\infty e^{-xy} x^{\nu-\frac{1}{2}}\, dx, \qquad \operatorname{Re} \nu > -\tfrac{1}{2}, \quad (5.10.18)$$

and hence, setting $y = e^u$ in (5.10.16), we have

$$\begin{aligned} H_\nu^{(1)}(z) &= \frac{e^{-\nu\pi i/2}}{\pi i} \int_0^\infty e^{\frac{1}{2}iz(y+y^{-1})} y^{-\nu-1} dy \\ &= \frac{e^{-\nu\pi i/2}}{\pi i \Gamma(\nu+\frac{1}{2})} \int_0^\infty e^{\frac{1}{2}iz(y+y^{-1})} y^{-1/2}\, dy \int_0^\infty e^{-xy}\, x^{\nu-\frac{1}{2}}\, dx \\ &= \frac{e^{-\nu\pi i/2}}{\pi i \Gamma(\nu+\frac{1}{2})} \int_0^\infty x^{\nu-\frac{1}{2}}\, dx \int_0^\infty \exp\left[-y\left(x - \frac{iz}{2}\right) + \frac{iz}{2y}\right] y^{-1/2}\, dy, \end{aligned}$$

where the reversal of the order of integration is easily justified by proving the absolute convergence of the double integral. To calculate the inner integral, we use the formula[14]

$$\int_0^\infty e^{-av^2-(b/v^2)} dv = \frac{\sqrt{\pi}}{2\sqrt{a}}\, e^{-2\sqrt{ab}}, \qquad a > 0, \quad b > 0. \qquad (5.10.19)$$

This gives

$$H_\nu^{(1)}(z) = \frac{e^{-\nu\pi i/2}}{i\sqrt{\pi}\,\Gamma(\nu+\frac{1}{2})} \int_0^\infty \frac{e^{-2\sqrt{-iz/2}\sqrt{x-(iz/2)}}}{\sqrt{x-(iz/2)}}\, x^{\nu-\frac{1}{2}}\, dx,$$

[14] After making the transformation $\tau = v^2$, the integral (5.10.19) becomes the Laplace transform of the function $\frac{1}{2}\tau^{-1/2}\, e^{-b/\tau}$, evaluated at $p = a$.

or

$$H_\nu^{(1)}(z) = \frac{2e^{-\nu\pi i}}{i\sqrt{\pi}\,\Gamma(\nu+\frac{1}{2})}\left(\frac{z}{2}\right)^\nu \int_1^\infty e^{izt}(t^2-1)^{\nu-\frac{1}{2}}\,dt, \qquad \operatorname{Re}\nu > -\tfrac{1}{2}, \tag{5.10.20}$$

where we introduce the new variable of integration

$$t = \frac{\sqrt{x-(iz/2)}}{\sqrt{-iz/2}}.$$

By the principle of analytic continuation, this formula, proved under the assumption that $-iz > 0$, remains valid for arbitrary complex z belonging to the sector $0 < \arg z < \pi$. In just the same way, we have the formula

$$H_\nu^{(2)}(z) = \frac{-2e^{\nu\pi i}}{i\sqrt{\pi}\,\Gamma(\nu+\frac{1}{2})}\left(\frac{z}{2}\right)^\nu \int_1^\infty e^{-izt}(t^2-1)^{\nu-\frac{1}{2}}\,dt, \tag{5.10.21}$$
$$\operatorname{Re}\nu > -\tfrac{1}{2}, \quad -\pi < \arg z < 0$$

for the second Hankel function. The integral representations (5.10.20–21) play an important role in the derivation of asymptotic representations of the cylinder functions as $|z| \to \infty$.

Integral representations for the Bessel functions of imaginary argument can either be obtained directly by a slight modification of the considerations of this section, or else deduced from (5.7.4–6) and the corresponding integral representations of the Bessel functions and Hankel functions. Thus, it follows from (5.10.3) that

$$I_\nu(z) = \frac{(z/2)^\nu}{\sqrt{\pi}\,\Gamma(\nu+\frac{1}{2})}\int_{-1}^1 (1-t^2)^{\nu-\frac{1}{2}}\cosh zt\,dt, \tag{5.10.22}$$
$$|\arg z| < \pi, \quad \operatorname{Re}\nu > -\tfrac{1}{2},$$

and from (5.10.16, 20) that

$$K_\nu(z) = \frac{1}{2}\int_{-\infty}^\infty e^{-z\cosh u - \nu u}\,du = \int_0^\infty e^{-z\cosh u}\cosh \nu u\,du, \tag{5.10.23}$$
$$\operatorname{Re} z > 0, \quad \nu \text{ arbitrary},$$

$$K_\nu(z) = \frac{\sqrt{\pi}}{\Gamma(\nu+\frac{1}{2})}\left(\frac{z}{2}\right)^\nu \int_1^\infty e^{-zt}(t^2-1)^{\nu-\frac{1}{2}}\,dt, \tag{5.10.24}$$
$$\operatorname{Re} z > 0, \quad \operatorname{Re}\nu > -\tfrac{1}{2}.$$

We also call attention to another integral representation

$$K_\nu(z) = \frac{1}{2}\left(\frac{z}{2}\right)^\nu \int_0^\infty e^{-t-(z^2/4t)}t^{-\nu-1}\,dt, \qquad |\arg z| < \frac{\pi}{4}, \tag{5.10.25}$$

which is useful in the applications, and is obtained from (5.10.23) by changing the variable of integration.

Some other useful integral representations of the cylinder functions and their products are given in Problems 1–9, p. 139.

5.11. Asymptotic Representations of the Cylinder Functions for Large $|z|$

There are simple asymptotic formulas which allow us to approximate the cylinder functions for large $|z|$ and fixed ν. The leading terms of these asymptotic expansions can be derived starting from the differential equations satisfied by the cylinder functions, but to obtain more exact expressions, it is preferable to use the integral representations found in the preceding section.

Asymptotic representations of the cylinder functions for large $|\nu|$ and fixed z can be obtained rather simply from formulas (5.3.2), (5.4.5), (5.6.4) and (5.7.1.–2) by using Stirling's formula (1.4.22). The problem of approximating the cylinder functions when both $|z|$ and $|\nu|$ are large is one of the most difficult problems of the theory. Some basic results along these lines can be found in Chapter 8 of Watson's treatise, and new formulas of this type have been obtained in recent years by Langer[15] and Cherry.[16]

Of all the cylinder functions, the Hankel functions have the simplest asymptotic representations. We now derive an asymptotic representation of the function $H_\nu^{(1)}(z)$, starting from formula (5.10.20). Making the substitution $t = 1 + 2s$, we find that

$$H_\nu^{(1)}(z) = \frac{2^{\nu+1}e^{i(z-\nu\pi)}z^\nu}{i\sqrt{\pi}\,\Gamma(\nu+\frac{1}{2})}\int_0^\infty e^{2zis}s^{\nu-\frac{1}{2}}(1+s)^{\nu-\frac{1}{2}}\,ds, \tag{5.11.1}$$
$$\operatorname{Re}\nu > -\tfrac{1}{2}, \quad 0 > \arg z < \pi.$$

Replacing $(1+s)^{\nu-\frac{1}{2}}$ by its binomial expansion

$$(1+s)^{\nu-\frac{1}{2}} = \sum_{k=0}^{n}\frac{(-1)^k(\frac{1}{2}-\nu)_k}{k!}s^k + \frac{(-1)^{n+1}(\frac{1}{2}-\nu)_{n+1}}{n!}s^{n+1}\int_0^1(1-t)^n(1+st)^{\nu-n-\frac{3}{2}}\,dt \tag{5.11.2}$$

---

[15] R. E. Langer, *On the asymptotic solutions of ordinary differential equations, with an application to the Bessel functions of large order*, Trans. Amer. Math. Soc., **33**, 23 (1931); *On the asymptotic solutions of differential equations, with an application to the Bessel functions of large complex order*, ibid., **34**, 447 (1942).

[16] T. M. Cherry, *Uniform asymptotic expansions*, J. Lond. Math. Soc., **24**, 121 (1949). *On expansion in eigenfunctions, particularly in Bessel functions*, Proc. Lond. Math. Soc., **51**, 14 (1949); *Uniform asymptotic formulae for functions with transition points*, Trans. Amer. Math. Soc., **68**, 224 (1950).

with remainder,[17] and integrating term by term, we obtain

$$H_\nu^{(1)}(z) = \left(\frac{2}{\pi z}\right)^{1/2} e^{i(z-\frac{1}{2}\nu\pi-\frac{1}{4}\pi)}\left[\sum_{k=0}^{n} \frac{(\frac{1}{2}-\nu)_k(\frac{1}{2}+\nu)_k}{k!}(2zi)^{-k} + r_n(z)\right].$$

Here

$$r_n(z) = \frac{(-1)^{n+1}(\frac{1}{2}-\nu)_{n+1}(-2zi)^{\nu+\frac{1}{2}}}{n!\Gamma(\nu+\frac{1}{2})} \times \int_0^\infty e^{2zis}s^{\nu+n+\frac{1}{2}}\,ds \int_0^1 (1-t)^n(1+st)^{\nu-n-\frac{3}{2}}\,dt,$$

and we have used the formula

$$\int_0^\infty e^{2zis}s^{k+\nu-\frac{1}{2}}\,ds = \Gamma(\nu+\tfrac{1}{2})(\nu+\tfrac{1}{2})_k(-2zi)^{-(k+\nu+\frac{1}{2})},$$
$$\operatorname{Re}\nu > -\tfrac{1}{2}, \quad 0 < \arg z < \pi, \quad k = 0, 1, 2, \ldots,$$

implied by (1.5.1).

Now suppose that $\delta \leqslant \arg z \leqslant \pi - \delta$, where δ is an arbitrarily small positive number, and for the time being, assume that $\operatorname{Re}\nu - n - \frac{3}{2} \leqslant 0$. Then, estimating $|r_n(z)|$, we find that[18]

$$|r_n(z)| \leqslant \frac{|(\frac{1}{2}-\nu)_{n+1}|(2|z|)^{\operatorname{Re}\nu+\frac{1}{2}}e^{\pi|\operatorname{Im}\nu|}}{n!|\Gamma(\nu+\frac{1}{2})|} \times \int_0^\infty e^{-2|z|s\sin\delta}s^{\operatorname{Re}\nu+n+\frac{1}{2}}\,ds\int_0^1(1-t)^n\,dt$$
$$= \frac{|(\frac{1}{2}-\nu)_{n+1}|(2|z|)^{\operatorname{Re}\nu+\frac{1}{2}}e^{\pi|\operatorname{Im}\nu|}\Gamma(\operatorname{Re}\nu+n+\frac{3}{2})}{(n+1)!|\Gamma(\nu+\frac{1}{2})|(2|z|\sin\delta)^{\operatorname{Re}\nu+n+\frac{3}{2}}} = O(|z|^{-n-1})$$

for fixed ν. Therefore

$$H_\nu^{(1)}(z) = \left(\frac{2}{\pi z}\right)^{1/2} e^{i(z-\frac{1}{2}\nu\pi-\frac{1}{4}\pi)}\left[\sum_{k=0}^{n} \frac{(\frac{1}{2}-\nu)_k(\frac{1}{2}+\nu)_k}{k!}(2zi)^{-k} + O(|z|^{-n-1})\right],$$
$$\operatorname{Re}\nu > -\tfrac{1}{2}, \quad \delta \leqslant \arg z \leqslant \pi - \delta, \quad n \geqslant \operatorname{Re}\nu - \tfrac{3}{2} \tag{5.11.3}$$

for large $|z|$. Actually, the condition imposed on n can be dropped, since if

$$\operatorname{Re}\nu - n - \tfrac{3}{2} > 0$$

---

[17] Note that

$$(1+\zeta)^\mu = \sum_{k=0}^{n}(-1)^k\frac{(-\mu)_k}{k!}\zeta^k + (-1)^{n+1}\frac{(-\mu)_{n+1}}{n!}\zeta^{n+1}\int_0^1(1-t)^n(1+\zeta t)^{\mu-n-1}\,dt,$$

where

$$|\arg(1+\zeta)| < \pi, \qquad (\lambda)_0 = 1, \qquad (\lambda)_k = \frac{\Gamma(\lambda+k)}{\Gamma(\lambda)} = \lambda(\lambda+1)\cdots(\lambda+k-1).$$

[18] For complex a and b we have

$$|a^b| = |a|^{\operatorname{Re}b}e^{-\operatorname{Im}b\cdot\arg a}.$$

we can always find an integer $m > n$ such that

$$\operatorname{Re} \nu - m - \tfrac{3}{2} \leqslant 0.$$

Then, representing $H_\nu^{(1)}(z)$ by (5.11.3) with n replaced by m, and noting that

$$\sum_{k=0}^{m} \cdots + O(|z|^{-m-1}) = \sum_{k=0}^{n} \cdots + \sum_{k=n+1}^{m} \cdots + O(|z|^{-m-1})$$
$$= \sum_{k=0}^{n} \cdots + O(|z|^{-n-1}),$$

we again arrive at (5.11.3). Moreover, the relation

$$H_\nu^{(1)}(z) = e^{-\nu\pi i} H_{-\nu}^{(1)}(z)$$

[cf. (5.6.5)] allows us to eliminate the condition imposed on the parameter ν, and in fact, by using an integral representation of a somewhat more general type than (5.10.20), it can be shown that the asymptotic formula (5.11.3) remains valid in the larger sector $|\arg z| \leqslant \pi - \delta$.[19] Finally, therefore, we have

$$H_\nu^{(1)}(z) = \left(\frac{2}{\pi z}\right)^{1/2} e^{i(z - \frac{1}{2}\nu\pi - \frac{1}{4}\pi)} \left[\sum_{k=0}^{n} (-1)^k (\nu, k)(2iz)^{-k} + O(|z|^{-n-1})\right],$$
$$|\arg z| \leqslant \pi - \delta \quad (5.11.4)$$

for large $|z|$, where we introduce the notation

$$(\nu, k) = \frac{(-1)^k}{k!} (\tfrac{1}{2} - \nu)_k (\tfrac{1}{2} + \nu)_k = \frac{(4\nu^2 - 1)(4\nu^2 - 3^2) \cdots (4\nu^2 - (2k - 1)^2)}{2^{2k} k!},$$
$$(\nu, 0) = 1.$$

An asymptotic representation of the function $H_\nu^{(2)}(z)$ can be obtained in the same way, starting from formula (5.10.21). The result is

$$H_\nu^{(2)}(z) = \left(\frac{2}{\pi z}\right)^{1/2} e^{-i(z - \frac{1}{2}\nu\pi - \frac{1}{4}\pi)} \left[\sum_{k=0}^{n} (\nu, k)(2iz)^{-k} + O(|z|^{-n-1})\right],$$
$$|\arg z| \leqslant \pi - \delta, \quad (5.11.5)$$

which differs from (5.11.4) only by the sign of i.

Asymptotic representations for the Bessel functions of the first and second kinds can be deduced from formulas (5.11.4–5) and the relations (5.6.1). Thus we find that[20]

$$J_\nu(z) = \left(\frac{2}{\pi z}\right)^{1/2} \cos(z - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi) \left[\sum_{k=0}^{n} (-1)^k (\nu, 2k)(2z)^{-2k} + O(|z|^{-2n-2})\right]$$
$$- \left(\frac{2}{\pi z}\right)^{1/2} \sin(z - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi)$$
$$\times \left[\sum_{k=0}^{n} (-1)^k (\nu, 2k + 1)(2z)^{-2k-1} + O(|z|^{-2n-3})\right],$$
$$|\arg z| \leqslant \pi - \delta, \quad (5.11.6)$$

[19] G. N. Watson, *op. cit.*, p. 196.

[20] In (5.11.6–8) the integer n need not be the same in both sums.

and

$$Y_\nu(z) = \left(\frac{2}{\pi z}\right)^{1/2} \cos\left(z - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi\right) \times \left[\sum_{k=0}^{n} (-1)^k(\nu, 2k+1)(2z)^{-2k-1} + O(|z|^{-2n-3})\right]$$
$$+ \left(\frac{2}{\pi z}\right)^{1/2} \sin\left(z - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi\right)\left[\sum_{k=0}^{n} (-1)^k(\nu, 2k)(2z)^{-2k} + O(|z|^{-2n-2})\right],$$
$$|\arg z| \leqslant \pi - \delta. \quad (5.11.7)$$

Similarly, asymptotic formulas for the Bessel functions of imaginary argument can be derived from the integral representations (5.10.22, 24), or else by using the relations given in Sec. 5.7, in conjunction with formulas (5.11.4–5). In this way, we find that

$$I_\nu(z) = e^z(2\pi z)^{-1/2}\left[\sum_{k=0}^{n} (-1)^k(\nu, k)(2z)^{-k} + O(|z|^{-n-1})\right]$$
$$+ e^{-z \pm \pi i(\nu+\frac{1}{2})}(2\pi z)^{-1/2}\left[\sum_{k=0}^{n} (\nu, k)(2z)^{-k} + O(|z|^{-n-1})\right],$$
$$|\arg z| \leqslant \pi - \delta, \quad (5.11.8)$$

and

$$K_\nu(z) = \left(\frac{\pi}{2z}\right)^{1/2} e^{-z}\left[\sum_{k=0}^{n} (\nu, k)(2z)^{-k} + O(|z|)^{-n-1}\right], \qquad |\arg z| \leqslant \pi - \delta, \quad (5.11.9)$$

where in (5.11.8) we choose the plus sign if $\operatorname{Im} z > 0$ and the minus sign if $\operatorname{Im} z < 0$. The second term in (5.11.8) will be small if $|\arg z| \leqslant \frac{1}{2}\pi - \delta$, and then

$$I_\nu(z) = e^z(2\pi z)^{-1/2}\left[\sum_{k=0}^{n} (-1)^k(\nu, k)(2z)^{-k} + O(|z|^{-n-1})\right], \quad |\arg z| \leqslant \frac{\pi}{2} - \delta. \quad (5.11.10)$$

The divergent series obtained by formally setting $n = \infty$ in each of the formulas (5.11.4–10) is the asymptotic series (see Sec. 1.4) of the function appearing in the left-hand side.

The method used here to derive asymptotic expansions gives only the order of magnitude of the remainder term $r_n(z)$, and does not furnish more exact information about the size of $|r_n(z)|$. With suitable assumptions concerning z and ν, the considerations given above can be modified to yield much more exact results. For example, it can be shown[21] that if z and ν are

[21] G. N. Watson, *op. cit.*, p. 206.

positive real numbers, and if n is so large that $2n \geqslant \nu - \frac{1}{2}$, then the remainder in the asymptotic expansion of $J_\nu(z)$ or $Y_\nu(z)$ is smaller in absolute value than the first neglected term, while the same is true of the asymptotic expansion of $K_\nu(z)$ if $n \geqslant \nu - \frac{3}{2}$.

5.12. Addition Theorems for the Cylinder Functions

Given an arbitrary triangle with sides r_1, r_2 and R, let θ and ψ be the angles opposite the sides R and r_1, respectively (see Figure 16), so that

$$R = \sqrt{r_1^2 + r_2^2 - 2r_1r_2 \cos \theta}, \qquad \sin \psi = \frac{r_1}{R} \sin \theta.$$

By an *addition theorem* for cylinder functions we mean an identity of the form

$$Z_\nu(\lambda R) = f_\nu(r_1, r_2, \theta) \sum_{(m)} \Phi_\nu^{(m)}(\lambda r_1)\Psi_\nu^{(m)}(\lambda r_2)\Theta_\nu^{(m)}(\theta), \tag{5.12.1}$$

where λ is an arbitrary complex number with $|\arg \lambda| < \pi$ (for integral ν, this condition can be dropped), and m ranges over some set of indices. Formula (5.12.1) is an expansion of the general cylinder function $Z_\nu(\lambda R)$ in a series whose terms are obtained by multiplying some function $f_\nu(r_1, r_2, \theta)$, which is independent of the summation index m, by three factors, each of which depends on only one of the variables r_1, r_2, θ.

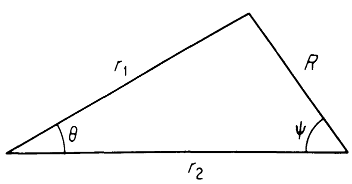


FIGURE 16

Formulas of this kind play an important role in the applications, especially in mathematical physics. The simplest such formula is the following addition theorem for the Bessel function of the first kind of order zero:

$$\begin{aligned} J_0(\lambda R) &= \sum_{m=-\infty}^{\infty} J_m(\lambda r_1)J_m(\lambda r_2)e^{im\theta} \\ &= J_0(\lambda r_1)J_0(\lambda r_2) + 2\sum_{m=1}^{\infty} J_m(\lambda r_1)J_m(\lambda r_2)\cos m\theta. \end{aligned} \tag{5.12.2}$$

To prove (5.12.2), we first note that

$$J_n(z) = \frac{1}{2\pi i}\int_C e^{\frac{1}{2}z(t-t^{-1})}t^{-n-1}\,dt, \qquad n = 0, \pm 1, \pm 2, \ldots \tag{5.12.3}$$

where C is an arbitrary closed contour surrounding the point $t = 0$.[22] Introducing a new variable of integration u by writing

$$t = \frac{r_1e^{i\theta} - r_2}{R}u,$$

[22] Formula (5.12.3) is a special case of (5.10.7) and can be proved immediately by using residues, after recalling the expansion (5.3.4).

and using the fact that

$$R^2 = (r_1 e^{i\theta} - r_2)(r_1 e^{-i\theta} - r_2),$$

we have

$$J_0(\lambda R) = \frac{1}{2\pi i}\int_{C'} \exp\left[\frac{\lambda r_1}{2}\left(ue^{i\theta} - \frac{1}{ue^{i\theta}}\right) - \frac{\lambda r_2}{2}\left(u - \frac{1}{u}\right)\right]\frac{du}{u},$$

where the integration is along a contour C' resembling C. Moreover, according to (5.3.4),

$$\exp\left[\frac{\lambda r_1}{2}\left(ue^{i\theta} - \frac{1}{ue^{i\theta}}\right)\right] = \sum_{m=-\infty}^{\infty} J_m(\lambda r_1)e^{im\theta}u^m, \tag{5.12.4}$$

where the convergence is uniform in u on the contour C'. Therefore, substituting (5.12.4) into (5.12.3) and integrating term by term, we find that

$$\begin{aligned} J_0(\lambda R) &= \sum_{m=-\infty}^{\infty} J_m(\lambda r_1)e^{im\theta}\frac{1}{2\pi i}\int_{C'} \exp\left[-\frac{\lambda r_2}{2}\left(u - \frac{1}{u}\right)\right]u^{m-1}\,du \\ &= \sum_{m=-\infty}^{\infty} J_m(\lambda r_1)J_{-m}(-\lambda r_2)e^{im\theta} = \sum_{m=-\infty}^{\infty} J_m(\lambda r_1)J_m(\lambda r_2)e^{im\theta}, \end{aligned}$$

which proves (5.12.2).

We now give two generalizations of formula (5.12.2) to the case of Bessel functions of arbitrary order ν, referring the reader elsewhere for proofs.[23] The first generalization is of the form[24]

$$J_\nu(\lambda R)\,\begin{matrix}\cos\nu\psi\\ \sin\nu\psi\end{matrix} = \sum_{m=-\infty}^{\infty} J_{\nu+m}(\lambda r_2)J_m(\lambda r_1)\,\begin{matrix}\cos m\theta\\ \sin m\theta\end{matrix}, \tag{5.12.5}$$

where ψ is shown in Figure 16, and $r_2 > r_1$ if ν is nonintegral (for integral ν, this restriction can be dropped, i.e., r_1 and r_2 can be interchanged). The second generalization of (5.12.2) is given by the formula

$$\frac{J_\nu(\lambda R)}{(\lambda R)^\nu} = 2^\nu\Gamma(\nu)\sum_{m=0}^{\infty}(\nu + m)\frac{J_{\nu+m}(\lambda r_1)J_{\nu+m}(\lambda r_2)}{(\lambda r_1)^\nu(\lambda r_2)^\nu}\,C_m^\nu(\cos\theta),$$
$$\nu \neq 0, -1, -2, \ldots, \tag{5.12.6}$$

where r_1 and r_2 are arbitrary. Here the functions $C_m^\nu(x)$, $m = 0, 1, 2, \ldots$, known as the *Gegenbauer polynomials*, are defined as the coefficients in the expansion

$$(1 - 2tx + t^2)^{-\nu} = \sum_{m=0}^{\infty} C_m^\nu(x)t^m, \tag{5.12.7}$$

[so that the function on the left is the generating function of the polynomials $C_m^\nu(x)$], and have the following explicit expressions:

$$C_m^\nu(x) = \sum_{k=0}^{[m/2]}(-1)^k 2^{m-2k}\frac{\Gamma(\nu + m - k)}{\Gamma(\nu)k!(m - 2k)!}x^{m-2k} \tag{5.12.8}$$

---

[23] G. N. Watson, *op. cit.*, Chap. 11.

[24] Formula (5.12.5) is an abbreviated way of writing two formulas, one involving cosines in both sides, the other sines.

$[C_0^\nu(x) = 1]$. For $\nu = \frac{1}{2}$ the expansion (5.12.7) reduces to formula (4.2.3), and then the Gegenbauer polynomials coincide with the Legendre polynomials:

$$C_m^{1/2}(x) = P_m(x). \tag{5.12.9}$$

For $\nu = 0$ we have

$$C_m^0(x) \equiv 0, \qquad m = 1, 2, \ldots,$$

but the product $\Gamma(\nu)C_m^\nu(x)$ approaches a finite limit as $\nu \to 0$:

$$\lim_{\nu \to 0} \Gamma(\nu)(\nu + m)C_m^\nu(x) = 2 \cos (m \operatorname{arc} \cos x), \qquad m = 1, 2, \ldots \tag{5.12.10}$$

Therefore both formulas (5.12.5–6) reduce to (5.12.2) in the limit $\nu \to 0$.

For cylinder functions of other kinds, we have similar addition theorems, among which we cite the following:

$$Z_\nu(\lambda R) \begin{matrix} \cos \nu\psi \\ \sin \nu\psi \end{matrix} = \sum_{m=-\infty}^{\infty} Z_{\nu+m}(\lambda r_2) J_m(\lambda r_1) \begin{matrix} \cos m\theta \\ \sin m\theta \end{matrix}, \tag{5.12.11}$$

$$\frac{Z_\nu(\lambda R)}{(\lambda R)^\nu} = 2^\nu \Gamma(\nu) \sum_{m=0}^{\infty} (\nu + m) \frac{Z_{\nu+m}(\lambda r_2) J_{\nu+m}(\lambda r_1)}{(\lambda r_2)^\nu (\lambda r_1)^\nu} C_m^\nu (\cos \theta), \tag{5.12.12}$$

$$I_\nu(\lambda R) \begin{matrix} \cos \nu\psi \\ \sin \nu\psi \end{matrix} = \sum_{m=-\infty}^{\infty} (-1)^m I_{\nu+m}(\lambda r_2) I_m(\lambda r_1) \begin{matrix} \cos m\theta \\ \sin m\theta \end{matrix}, \tag{5.12.13}$$

$$\frac{I_\nu(\lambda R)}{(\lambda R)^\nu} = 2^\nu \Gamma(\nu) \sum_{m=0}^{\infty} (-1)^m (\nu + m) \frac{I_{\nu+m}(\lambda r_2) I_{\nu+m}(\lambda r_1)}{(\lambda r_2)^\nu (\lambda r_1)^\nu} C_m^\nu (\cos \theta), \tag{5.12.14}$$

$$K_\nu(\lambda R) \begin{matrix} \cos \nu\psi \\ \sin \nu\psi \end{matrix} = \sum_{m=-\infty}^{\infty} K_{\nu+m}(\lambda r_2) I_m(\lambda r_1) \begin{matrix} \cos m\theta \\ \sin m\theta \end{matrix}, \tag{5.12.15}$$

$$\frac{K_\nu(\lambda R)}{(\lambda R)^\nu} = 2^\nu \Gamma(\nu) \sum_{m=0}^{\infty} (\nu + m) \frac{K_{\nu+m}(\lambda r_2) I_{\nu+m}(\lambda r_1)}{(\lambda r_2)^\nu (\lambda r_1)^\nu} C_m^\nu (\cos \theta). \tag{5.12.16}$$

In formulas (5.12.11–13, 15–16), it is assumed that $r_2 > r_1$ unless ν is an integer or $Z_{\nu+m} = J_{\nu+m}$ in (5.12.12).

An important special case of these addition theorems, encountered in mathematical physics, occurs when $\nu = \frac{1}{2}$. The formulas corresponding to this case are easily obtained by using (5.12.9), together with the results of Sec. 5.8.[25]

5.13. Zeros of the Cylinder Functions

In solving many applied problems, one needs information about the location of the zeros of cylinder functions in the complex plane, and in particular,

[25] G. N. Watson, *op. cit.*, p. 368.

one must be able to make approximate calculations of the values of these zeros. Here we cite without proof some important results along these lines.[26] We begin by considering the distribution of zeros of the Bessel functions of the first kind, i.e., roots of the equation

$$J_\nu(z) = 0. \tag{5.13.1}$$

Theorem 1 deals with the case of nonnegative integral ν, and Theorem 2 with the case of arbitrary real ν:

THEOREM 1. *The function $J_n(z)$, $n = 0, 1, 2, \ldots$ has no complex zeros, and has an infinite number of real zeros symmetrically located with respect to the point $z = 0$, which is itself a zero if $n > 0$. All the zeros of $J_n(z)$ are simple, except the point $z = 0$, which is a zero of order n if $n > 0$.*

THEOREM 2. *Let ν be an arbitrary real number, and suppose that $|\arg z| < \pi$. Then the function $J_\nu(z)$ has an infinite number of positive real zeros, and a finite number $2N(\nu)$ of conjugate complex zeros, where*

1. *$N(\nu) = 0$ if $\nu > -1$ or $\nu = -1, -2, \ldots$;*
2. *$N(\nu) = m$ if $-(m + 1) < \nu < -m, \quad m = 1, 2, \ldots$*

(In the second case, if $[-\nu]$ is odd, there is a pair of purely imaginary zeros among the conjugate complex zeros.) Moreover, all the zeros are simple, except possibly the zero at the point $z = 0$.

The following generalization of equation (5.13.1) is often encountered in mathematical physics (A and B are real):

$$AJ_\nu(z) + BzJ'_\nu(z) = 0, \qquad \nu > -1, \qquad |\arg z| < \pi. \tag{5.13.2}$$

It can be shown that this equation has infinitely many positive real roots and no complex roots, unless

$$\frac{A}{B} + \nu < 0,$$

in which case (5.13.2) also has two purely imaginary roots.[27]

The distribution of zeros of the function $I_\nu(z)$ can be deduced from Theorem 2 and the relations of Sec. 5.7. In particular, it should be noted that all the zeros of $I_\nu(z)$ are purely imaginary if $\nu > -1$. If ν is real, Macdonald's function $K_\nu(z)$ has no zeros in the region $|\arg z| \leqslant \pi/2$. In the rest of the z-plane cut along the segment $[-\infty, 0]$, $K_\nu(z)$ has a finite number of zeros.[28]

---

[26] The problem of the distribution of the zeros of cylinder functions is also of considerable theoretical interest, but lies outside the scope of this book. We again refer the reader interested in details to the specialized literature, e.g., Chap. 15 of Watson's treatise. It should be noted that some of the results on zeros of cylinder functions can be derived by arguments of a completely elementary character.

[27] G. N. Watson, *op. cit.*, p. 482.

[28] *Ibid.*, p. 511.

To make approximate calculations of the roots of equations involving cylinder functions, one can use the method of successive approximations, where in many cases a good first approximation is given by the roots of the equations obtained when the cylinder functions are replaced by their asymptotic representations.

5.14. Expansions in Series and Integrals Involving Cylinder Functions

In mathematical physics, it is often necessary to expand a given function in terms of cylinder functions, where the form of the expansion depends on the specific nature of the problem (see Secs. 6.3–6.7). We now consider the most important of these expansions, whose role in various problems involving cylinder functions resembles that of Fourier series and Fourier integrals in problems involving trigonometric functions. Foremost among such expansions are series of the form

$$f(r) = \sum_{m=1}^{\infty} c_m J_\nu\left(x_{\nu m}\frac{r}{a}\right), \qquad 0 < r < a, \quad \nu \geqslant -\tfrac{1}{2}, \tag{5.14.1}$$

where $f(r)$ is a given real function defined in the interval $(0, a)$, $J_\nu(x)$ is a Bessel function of the first kind of real order $\nu \geqslant -\frac{1}{2}$, and

$$0 < x_{\nu 1} < \cdots < x_{\nu m} < \cdots$$

are the positive roots of the equation $J_\nu(x) = 0$. The expansion coefficients c_m can be determined by using an orthogonality property of the system of functions

$$J_\nu\left(x_{\nu m}\frac{r}{a}\right), \qquad m = 1, 2, \ldots, \tag{5.14.2}$$

which is proved as follows: Let α and β be distinct nonzero real numbers, and let

$$u_\alpha'' + \frac{1}{r}u_\alpha' + \left(\alpha^2 - \frac{\nu^2}{r^2}\right)u_\alpha = 0, \qquad u_\beta'' + \frac{1}{r}u_\beta' + \left(\beta^2 - \frac{\nu^2}{r^2}\right)u_\beta = 0$$

be the equations satisfied by the functions $u_\alpha = J_\nu(\alpha r)$ and $u_\beta = J_\nu(\beta r)$. Subtracting the second equation multiplied by ru_α from the first equation multiplied by ru_β, and integrating the result from 0 to a, we find that

$$(\alpha^2 - \beta^2)\int_0^a ru_\alpha u_\beta\, dr = r(u_\alpha u_\beta' - u_\beta u_\alpha')\Big|_0^a,$$

which implies

$$\int_0^a rJ_\nu(\alpha r)J_\nu(\beta r)\, dr = \frac{\alpha\beta J_\nu(\alpha a)J_\nu'(\beta a) - a\alpha J_\nu(\beta a)J_\nu'(\alpha a)}{\alpha^2 - \beta^2} \tag{5.14.3}$$

if $\nu > -1$. Setting $\alpha = x_{\nu m}/a$, $\beta = x_{\nu n}/a$ in (5.14.3), we obtain the formula

$$\int_0^a rJ_\nu\left(x_{\nu m}\frac{r}{a}\right)J_\nu\left(x_{\nu n}\frac{r}{a}\right) dr = 0 \quad \text{if} \quad m \neq n, \tag{5.14.4}$$

which shows that the system (5.14.2) is orthogonal with weight r on the interval $[0, a]$ (see Sec. 4.1).

Taking the limit of (5.14.3) as $\beta \to \alpha$, with the aid of L'Hospital's rule, and using Bessel's equation to eliminate J_ν'', we find that[29]

$$\int_0^a rJ_\nu^2(\alpha r)\, dr = \frac{a^2}{2}\left[J_\nu'^2(\alpha a) + \left(1 - \frac{\nu^2}{\alpha^2 a^2}\right)J_\nu^2(\alpha a)\right], \tag{5.14.5}$$

or, using the relations (5.3.5),

$$\int_0^a rJ_\nu^2\left(x_{\nu n}\frac{r}{a}\right) dr = \frac{a^2}{2}J_\nu'^2(x_{\nu n}) = \frac{a^2}{2}J_{\nu+1}^2(x_{\nu n}). \tag{5.14.6}$$

Then, assuming that an expansion of the form (5.14.1) is possible, multiplying by $rJ_\nu(x_{\nu n}\, r/a)$ and integrating term by term from 0 to a, we obtain the following formal values of the coefficients c_m:

$$c_m = \frac{2}{a^2 J_{\nu+1}^2(x_{\nu m})}\int_0^a rf(r)J_\nu\left(x_{\nu m}\frac{r}{a}\right) dr, \qquad m = 1, 2, \ldots \tag{5.14.7}$$

The series (5.14.1), with coefficients calculated from (5.14.7), is called the *Fourier-Bessel series* of the function $f(r)$.

We now cite a theorem which gives conditions under which the Fourier-Bessel series of the function $f(r)$ actually converges and has the sum $f(r)$:

THEOREM 3.[30] *Suppose the real function $f(r)$ is piecewise continuous in $(0, a)$ and of bounded variation in every subinterval $[r_1, r_2]$,*[31] *where $0 < r_1 < r_2 < a$. Then, if the integral*

$$\int_0^a \sqrt{r}\,|f(r)|\, dr$$

is finite, the Fourier-Bessel series (5.14.1) *converges to $f(r)$ at every continuity point of $f(r)$, and to*

$$\tfrac{1}{2}[f(r + 0) + f(r - 0)]$$

at every discontinuity point of $f(r)$.

Next, we consider an important generalization of the concept of a Fourier-Bessel series. Suppose the function $f(r)$ is expanded in a series of the form (5.14.1), where this time the numbers

$$0 < x_{\nu 1} < \cdots < x_{\nu m} < \cdots$$

---

[29] The details are given in G. P. Tolstov, *op. cit.*, p. 218.

[30] For the proof, see G. N. Watson, *op. cit.*, p. 591.

[31] Concerning *functions of bounded variation*, see E. C. Titchmarsh, *op. cit.*, p. 355.

are the roots of the equation

$$AJ_\nu(x) + BxJ'_\nu(x) = 0, \tag{5.14.8}$$

instead of the equation $J_\nu(x) = 0$. Then it is an immediate consequence of formulas (5.14.3, 5, 8) that

$$\int_0^a rJ_\nu\left(x_{\nu m}\frac{r}{a}\right)J_\nu\left(x_{\nu n}\frac{r}{a}\right)dr = \begin{cases} 0 & \text{if} \quad m \neq n, \\ \dfrac{a^2}{2}\left[J'^2_\nu(x_{\nu n}) + \left(1 - \dfrac{\nu^2}{x^2_{\nu n}}\right)J^2_\nu(x_{\nu n})\right] & \text{if} \quad m = n, \end{cases} \tag{5.14.9}$$

and therefore the coefficients c_m are now given by

$$c_m = \frac{2}{a^2\{J'^2_\nu(x_{\nu m}) + [1 - (\nu^2/x^2_{\nu m})]J^2_\nu(x_{\nu m})\}}\int_0^a rf(r)J_\nu\left(x_{\nu m}\frac{r}{a}\right)dr. \tag{5.14.10}$$

The series (5.14.1), with coefficients calculated from (5.14.10), is called the *Dini series*[32] of the function $f(r)$. If $f(r)$ satisfies the conditions of Theorem 3, and if $AB^{-1} + \nu > 0$, then the Dini series of $f(r)$ actually converges to $f(r)$ at every continuity point.[33] Both Fourier-Bessel series and Dini series play an important role in problems of mathematical physics, and examples of such expansions will be given in Secs. 6.3 and 6.7.

We now turn to expansions of a function $f(r)$ defined in the infinite interval $(0, \infty)$, in terms of *integrals* involving Bessel functions. Among such expansions, the one of greatest practical importance is the *Fourier-Bessel integral*, defined by

$$f(r) = \int_0^\infty \lambda J_\nu(\lambda r)\,d\lambda \int_0^\infty \rho J_\nu(\lambda\rho)f(\rho)\,d\rho, \qquad 0 < r < \infty, \quad \nu > -\tfrac{1}{2}. \tag{5.14.11}$$

Formula (5.14.11) is sometimes called *Hankel's integral theorem*, and is valid at every continuity point of $f(r)$ provided that

1. The function $f(r)$, defined in the infinite interval $(0, \infty)$, is piecewise continuous and of bounded variation in every finite subinterval $[r_1, r_2]$, where $0 < r_1 < r_2 < \infty$;
2. The integral

$$\int_0^\infty \sqrt{r}\,|f(r)|\,dr$$

is finite.[34]

---

[32] Called a *Fourier-Bessel series of the second type* in G. P. Tolstov, *op. cit.*, p. 237.

[33] For the proof, see G. N. Watson, *op. cit.*, p. 596 ff., where one will also find the modifications that must be made in the Dini series if $AB^{-1} + \nu \leqslant 0$.

[34] For the proof, see G. N. Watson, *op. cit.*, p. 456 ff. At discontinuity points, the integral in the right-hand side of (5.4.11) equals

$$\tfrac{1}{2}[f(r + 0) + f(r - 0)].$$

As examples of Fourier-Bessel integrals, consider the expansions

$$\frac{1}{\sqrt{z^2 + r^2}} = \int_0^\infty e^{-\lambda|z|} J_0(\lambda r)\, d\lambda, \tag{5.14.12}$$

$$\frac{e^{-k\sqrt{z^2+r^2}}}{\sqrt{z^2 + r^2}} = \int_0^\infty e^{-|z|\sqrt{\lambda^2+k^2}} \frac{\lambda J_0(\lambda r)}{\sqrt{\lambda^2 + k^2}}\, d\lambda \tag{5.14.13}$$

(with real z and r), implied by formulas (5.15.1, 7) below.

The author has studied another integral expansion of a completely different type, involving integration with respect to the *order* of the cylinder function.[35] This expansion, which turns out to be very useful in solving certain problems of mathematical physics (see Secs. 6.5–6) is of the form

$$f(x) = \frac{2}{\pi^2} \int_0^\infty \tau \sinh \pi\tau \frac{K_{i\tau}(x)}{\sqrt{x}}\, d\tau \int_0^\infty f(\xi) \frac{K_{i\tau}(\xi)}{\sqrt{\xi}}\, d\xi, \qquad x > 0, \tag{5.14.14}$$

where $K_\nu(x)$ is Macdonald's function of imaginary order $\nu = i\tau$. Formula (5.14.14) is valid at every continuity point of $f(x)$ provided that

1. The function $f(x)$, defined in the infinite interval $(0, \infty)$, is piecewise continuous and of bounded variation in every finite subinterval $[x_1,x_2]$, where $0 < x_1 < x_2 < \infty$;
2. The integrals

$$\int_0^{1/2} |f(x)| x^{-1/2} \log \frac{1}{x}\, dx, \qquad \int_{1/2}^\infty |f(x)| dx \tag{5.14.15}$$

are finite.

Example. An expansion of this type is[36]

$$f(x) = \sqrt{x} e^{-x \cos \alpha} = \frac{2}{\pi} \int_0^\infty \frac{\tau \sinh \alpha\tau}{\sin \alpha} \frac{K_{i\tau}(x)}{\sqrt{x}}\, d\tau. \tag{5.14.16}$$

5.15. Definite Integrals Involving Cylinder Functions

In the applications, it is often necessary to evaluate integrals involving cylinder functions in combination with various elementary functions or special

---

[35] N. N. Lebedev, *Sur une formule d'inversion*, Dokl. Akad. Nauk SSSR, **52**, 655 (1946); *Expansion of an arbitrary function in an integral with respect to cylinder functions of imaginary order and argument* (in Russian), Prikl. Mat. Mekh., **13**, 465 (1949); *Some Integral Transformations of Mathematical Physics* (in Russian), Dissertation, Izd. Leningrad. Gos. Univ. (1951). At discontinuity points, the integral in the right-hand side of (5.14.14) equals

$$\tfrac{1}{2}[f(x + 0) + f(x - 0)].$$

[36] To derive (5.14.16), use (5.14.14) and the Bateman Manuscript Project, *Tables of Integral Transforms, Vol. 1*, formula (24), p. 197.

functions of other kinds. Such integrals are usually evaluated by replacing the cylinder function by a series or by a suitable integral representation, and then reversing the order in which the operations are carried out. Since an extremely detailed treatment of this whole topic is available in the literature,[37] we confine ourselves here to a few examples which illustrate the method and lead to some results needed later in the book.

Example 1. *Evaluate the integral*

$$\int_0^\infty e^{-ax} J_0(bx)\, dx, \qquad a > 0, \quad b > 0.$$

Replacing $J_0(bx)$ by its integral representation (5.10.8), we find that

$$\begin{aligned}\int_0^\infty e^{-ax} J_0(bx)\, dx &= \int_0^\infty e^{-ax}\, dx \frac{2}{\pi}\int_0^{\pi/2} \cos(bx \sin\varphi)\, d\varphi \\ &= \frac{2}{\pi}\int_0^{\pi/2} d\varphi \int_0^\infty e^{-ax}\cos(bx\sin\varphi)\, dx \\ &= \frac{2}{\pi}\int_0^{\pi/2} \frac{a\, d\varphi}{a^2 + b^2\sin^2\varphi},\end{aligned}$$

where the absolute convergence of the double integral justifies reversing the order of integration. Evaluating the last integral, we have

$$\int_0^\infty e^{-ax} J_0(bx)\, dx = \frac{1}{\sqrt{a^2+b^2}}, \qquad a > 0, \quad b > 0. \tag{5.15.1}$$

Example 2. *Evaluate Weber's integral*

$$\int_0^\infty e^{-a^2x^2} J_\nu(bx) x^{\nu+1}\, dx, \qquad a > 0, \quad b > 0, \quad \operatorname{Re}\nu > -1.$$

Replacing $J_\nu(bx)$ by its series expansion (5.3.2) and integrating term by term, we find that

$$\begin{aligned}\int_0^\infty e^{-a^2x^2} J_\nu(bx) x^{\nu+1}\, dx &= \int_0^\infty e^{-a^2x^2} x^{\nu+1}\, dx \sum_{k=0}^\infty \frac{(-1)^k (bx/2)^{\nu+2k}}{k!\,\Gamma(k+\nu+1)} \\ &= \sum_{k=0}^\infty \frac{(-1)^k}{k!\,\Gamma(k+\nu+1)} \left(\frac{b}{2}\right)^{\nu+2k} \int_0^\infty e^{-a^2x^2} x^{2\nu+2k+1}\, dx \\ &= \sum_{k=0}^\infty \frac{(-1)^k}{k!\,\Gamma(k+\nu+1)} \left(\frac{b}{2}\right)^{\nu+2k} \frac{1}{2a^{2\nu+2k+2}} \int_0^\infty e^{-t} t^{\nu+k}\, dt \\ &= \frac{b^\nu}{(2a^2)^{\nu+1}} \sum_{k=0}^\infty \frac{(-b^2/4a^2)^k}{k!},\end{aligned}$$

[37] G. N. Watson, *op. cit.*, Chaps. 12–13, the Bateman Manuscript Project, *Higher Transcendental Functions, Vol. 2*, Chap. 7, and ibid., *Tables of Integral Transforms, Vols. 1, 2.* See also F. Oberhettinger, *Tabellen zur Fourier Transformation*, Springer-Verlag, Berlin (1957).

where reversing the order of integration and summation is again justified by an absolute convergence argument. Summing the last series, we have

$$\int_0^\infty e^{-a^2x^2} J_\nu(bx)x^{\nu+1}\,dx = \frac{b^\nu}{(2a^2)^{\nu+1}} e^{-b^2/4a^2}, \qquad (5.15.2)$$
$$a > 0, b > 0, \quad \operatorname{Re}\nu > -1.$$

Example 3. *Evaluate the integral*

$$\int_0^\infty \frac{x^{\nu+1}J_\nu(bx)}{(x^2+a^2)^{\mu+1}}\,dx, \qquad a > 0, \quad b > 0, \quad -1 < \operatorname{Re}\nu < 2\operatorname{Re}\mu + \tfrac{3}{2},$$

often encountered in the applications. First we replace the function $(x^2 + a^2)^{-\mu-1}$ by an integral of the type (1.5.1), i.e.,

$$\frac{1}{(x^2+a^2)^{\mu+1}} = \frac{1}{\Gamma(\mu+1)}\int_0^\infty e^{-(x^2+a^2)t}t^\mu\,dt, \ \operatorname{Re}\mu > -1, \qquad (5.15.3)$$

assuming temporarily that $-1 < \operatorname{Re}\nu < 2\operatorname{Re}\mu + \frac{1}{2}$ (this guarantees absolute convergence of the relevant double integral). Then, using (5.15.2) and the integral representation (5.10.25) of Macdonald's function, we find that

$$\begin{aligned}
\int_0^\infty \frac{x^{\nu+1}J_\nu(bx)}{(x^2+a^2)^{\mu+1}}\,dx &= \frac{1}{\Gamma(\mu+1)}\int_0^\infty e^{-a^2t}t^\mu\,dt\int_0^\infty e^{-x^2t}J_\nu(bx)x^{\nu+1}\,dx \\
&= \frac{b^\nu}{2^{\nu+1}\Gamma(\mu+1)}\int_0^\infty e^{-a^2t-(b^2/4t)}\frac{dt}{t^{\nu+1-\mu}} \\
&= \frac{b^\nu a^{2\nu-2\mu}}{2^{\nu+1}\Gamma(\mu+1)}\int_0^\infty e^{-u-[(ab)^2/4u]}\frac{du}{u^{\nu-\mu+1}} \\
&= \frac{a^{\nu-\mu}b^\mu}{2^\mu\Gamma(\mu+1)}K_{\nu-\mu}(ab).
\end{aligned}$$

The extension of this result to values of the parameter μ satisfying the weaker condition $-1 < \operatorname{Re}\nu < 2\operatorname{Re}\mu + \frac{3}{2}$ is accomplished by using the principle of analytic continuation. Thus we have

$$\int_0^\infty \frac{x^{\nu+1}J_\nu(bx)}{(x^2+a^2)^{\mu+1}}\,dx = \frac{a^{\nu-\mu}b^\mu}{2^\mu\Gamma(\mu+1)}K_{\nu-\mu}(ab), \qquad (5.15.4)$$
$$a > 0, b > 0, \quad -1 < \operatorname{Re}\nu < 2\operatorname{Re}\mu + \tfrac{3}{2}.$$

In particular, setting $\mu = -\frac{1}{2}$, $\nu = 0$ and using (5.8.5), we obtain the integral

$$\int_0^\infty \frac{xJ_0(bx)}{\sqrt{x^2+a^2}}\,dx = \frac{e^{-ab}}{b}, \qquad a \geqslant 0, \quad b > 0. \qquad (5.15.5)$$

Example 4. *Evaluate the integral*

$$\int_0^\infty \frac{K_\mu(a\sqrt{x^2+y^2})}{(x^2+y^2)^{\mu/2}}J_\nu(bx)x^{\nu+1}\,dx,$$
$$a > 0, \quad b > 0, \quad y > 0, \quad \operatorname{Re}\nu > -1,$$

which also has numerous applications to mathematical physics. Using the integral representation (5.10.25) and formula (5.15.2), we find that

$$\begin{aligned}
\int_0^\infty & \frac{K_\mu(a\sqrt{x^2+y^2})}{(x^2+y^2)^{\mu/2}} J_\nu(bx)x^{\nu+1}\,dx \\
&= \frac{a^\mu}{2^{\mu+1}} \int_0^\infty J_\nu(bx)x^{\nu+1}\,dx \int_0^\infty e^{-t-[a^2(x^2+y^2)/4t]} \frac{dt}{t^{\mu+1}} \\
&= \frac{a^\mu}{2^{\mu+1}} \int_0^\infty e^{-t-(a^2y^2/4t)} \frac{dt}{t^{\mu+1}} \int_0^\infty e^{-a^2x^2/4t} J_\nu(bx)x^{\nu+1}\,dx \\
&= 2^{\nu-\mu}a^{\mu-2\nu-2}b^\nu \int_0^\infty e^{-t(1+b^2/a^2)-(a^2y^2/4t)} \frac{dt}{t^{\mu-\nu}} \\
&= \frac{2^{\nu-\mu}b^\nu}{a^\mu}(a^2+b^2)^{\mu-\nu-1} \int_0^\infty e^{-u-[y^2(a^2+b^2)/4u]} \frac{du}{u^{\mu-\nu}} \\
&= \frac{b^\nu}{a^\mu}\left(\frac{\sqrt{a^2+b^2}}{y}\right)^{\mu-\nu-1} K_{\mu-\nu-1}(y\sqrt{a^2+b^2}).
\end{aligned}$$

By choosing various values of the parameters in the identity

$$\int_0^\infty \frac{K_\mu(a\sqrt{x^2+y^2})}{(x^2+y^2)^{\mu/2}} J_\nu(bx)x^{\nu+1}\,dx = \frac{b^\nu}{a^\mu}\left(\frac{\sqrt{a^2+b^2}}{y}\right)^{\mu-\nu-1} K_{\mu-\nu-1}(y\sqrt{a^2+b^2}),$$

$$a > 0, \quad b > 0, \quad y > 0, \quad \operatorname{Re}\nu > -1, \qquad (5.15.6)$$

we can derive a number of useful formulas encountered in the applications. For example, setting $\mu = \frac{1}{2}$, $\nu = 0$, we have

$$\int_0^\infty \frac{e^{-a\sqrt{x^2+y^2}}}{\sqrt{x^2+y^2}} J_0(bx)x\,dx = \frac{e^{-y\sqrt{a^2+b^2}}}{\sqrt{a^2+b^2}}. \qquad (5.15.7)$$

5.16. Cylinder Functions of Nonnegative Argument and Order

We now collect some elementary and easily verified results pertaining to the very important case of cylinder functions where both the argument x and the order ν are nonnegative real numbers:

1. *Bessel functions of the first kind.* For $x \geqslant 0$ and $\nu \geqslant 0$, the function $J_\nu(x)$ is real and bounded, and has an oscillatory character. Its behavior for small and large values of x is described by the asymptotic formulas

$$\begin{aligned}
J_\nu(x) &\approx \frac{x^\nu}{2^\nu\Gamma(1+\nu)}, \qquad x \to 0, \\
J_\nu(x) &\approx \sqrt{\frac{2}{\pi x}} \cos(x - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi), \qquad x \to \infty.
\end{aligned} \qquad (5.16.1)$$

$J_\nu(x)$ has infinitely many zeros, including the point $x = 0$ if $\nu > 0$. The graphs of $J_0(x)$ and $J_1(x)$ are shown in Figure 17.

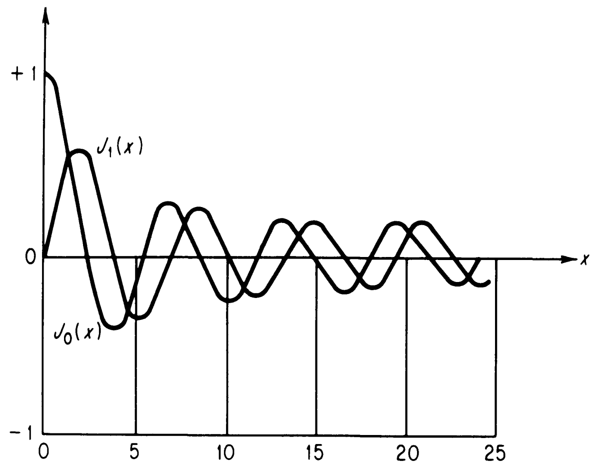


FIGURE 17

2. *Bessel functions of the second kind.* For $x > 0$ and $\nu \geqslant 0$, the function $Y_\nu(x)$ is an oscillatory real function, which is bounded at infinity. Its behavior for small and large values of x is described by the asymptotic formulas

$$
\begin{aligned}
Y_\nu(x) &\approx -\frac{2^\nu\Gamma(\nu)}{\pi x^\nu} \qquad x \to 0, \quad \nu > 0,\\
Y_\nu(x) &\approx \sqrt{\frac{2}{\pi x}}\sin\left(x - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi\right), \qquad x \to \infty, \qquad (5.16.2)\\
Y_0(x) &\approx -\frac{2}{\pi}\log\frac{2}{x}, \qquad x \to 0,
\end{aligned}
$$

which show, in particular, that $Y_\nu(x) \to -\infty$ as $x \to 0$.

3. *Bessel functions of the third kind.* For $x > 0$ and $\nu \geqslant 0$, the Hankei functions $H_\nu^{(1)}(x)$ and $H_\nu^{(2)}(x)$ are conjugate complex functions, which are bounded at infinity. Their behavior for small and large values of x is described by the asymptotic formulas

$$
\begin{aligned}
H_\nu^{(p)}(x) &\approx \mp i\left(\frac{2}{x}\right)^\nu \frac{\Gamma(\nu)}{\pi}, \qquad x \to 0, \quad \nu > 0,\\
H_\nu^{(p)}(x) &\approx \sqrt{\frac{2}{\pi x}}\, e^{\pm i(x - \frac{1}{2}\nu\pi - \frac{1}{4}\pi)}, \qquad x \to \infty, \qquad (5.16.3)\\
H_0^{(p)}(x) &\approx \mp i\frac{2}{\pi}\log\frac{2}{x}, \qquad x \to 0,
\end{aligned}
$$

where the upper sign corresponds to the case $p = 1$, and the lower sign to the case $p = 2$. Obviously, $H_\nu^{(p)}(x) \to \infty$ as $x \to 0$.

4. *Bessel functions of imaginary argument.* For $x > 0$ and $\nu \geqslant 0$, $I_\nu(x)$ is a positive function which increases monotonically as $x \to \infty$, while $K_\nu(x)$ is a positive function which decreases monotonically as $x \to \infty$.[38] For small x we have the asymptotic formulas

$$\begin{aligned} I_\nu(x) &\approx \frac{x^\nu}{2^\nu\Gamma(1+\nu)}, && x \to 0, \\ K_\nu(x) &\approx \frac{2^{\nu-1}\Gamma(\nu)}{x^\nu}, && x \to 0, \\ K_0(x) &\approx \log\frac{2}{x}, && x \to 0, \end{aligned} \tag{5.16.4}$$

and therefore

$$I_\nu(0) = 0 \quad \text{if } \nu > 0, \qquad I_0(0) = 1, \quad K_\nu(0) = \infty.$$

The asymptotic behavior of these functions as $x \to \infty$ is given by

$$\begin{aligned} I_\nu(x) &\approx \frac{e^x}{\sqrt{2\pi x}}, && x \to \infty, \\ K_\nu(x) &\approx \sqrt{\frac{\pi}{2x}}\, e^{-x}, && x \to \infty. \end{aligned} \tag{5.16.5}$$

Clearly, neither function has any zeros for $x > 0$.

5.17. Airy Functions

The solutions of the second-order linear differential equation

$$u'' - zu = 0 \tag{5.17.1}$$

are called *Airy functions.* These functions are closely related to the cylinder functions, and play an important role in the theory of asymptotic representations of various special functions arising as solutions of linear differential equations.[39] In particular, the Airy functions turn out to be useful in deriving asymptotic representations of the cylinder functions for large values of $|z|$ and $|\nu|$, valid in an extended region of values of z and ν. The Airy functions also have a variety of applications to mathematical physics, e.g., the theory of diffraction of radio waves around the earth's surface.[40]

---

[38] This fact about $K_\nu(x)$ follows from the integral representation (5.10.23).

[39] See R. E. Langer, *op. cit.*, T. M. Cherry, *op. cit.*, and V. A. Fock, *Tables of the Airy Fuctions* (in Russian), Izd. Inform. Otdel. Nauchno-Issled. Inst., Moscow (1946).

[40] See V. A. Fock, *Diffraction of Radio Waves Around the Earth's Surface* (in Russian), Izd. Akad. Nauk SSSR, Moscow (1946).

We now present the rudiments of the theory of Airy functions. Choosing $\alpha = -1, \gamma = 1$ in the second of the equations (5.4.11–12), and using the results of Sec. 5.7, we find that the general solution of (5.17.1) can be expressed in terms of Bessel functions of imaginary argument of order $\nu = \pm\frac{1}{3}$. In particular, two linearly independent solutions of (5.17.1) are

$$\begin{aligned} u = u_1 = \mathrm{Ai}(z) &= \frac{z^{1/2}}{3}\left[I_{-1/3}\left(\frac{2z^{3/2}}{3}\right) - I_{1/3}\left(\frac{2z^{3/2}}{3}\right)\right] \\ &\equiv \frac{1}{\pi}\left(\frac{z}{3}\right)^{1/2} K_{1/3}\left(\frac{2z^{3/2}}{3}\right), \qquad |\arg z| < \frac{2\pi}{3}, \\ u = u_2 = \mathrm{Bi}(z) &= \left(\frac{z}{3}\right)^{1/2}\left[I_{-1/3}\left(\frac{2z^{3/2}}{3}\right) + I_{1/3}\left(\frac{2z^{3/2}}{3}\right)\right], \qquad |\arg z| < \frac{2\pi}{3}, \end{aligned} \tag{5.17.2}$$

called the *Airy functions of the first and second kind*, respectively. Replacing $I_{\pm 1/3}$ by the series expansion (5.7.1), we obtain the expansions

$$\begin{aligned} \mathrm{Ai}(z) &= \sum_{k=0}^{\infty} \frac{z^{3k}}{3^{2k+2/3}k!\Gamma(k+\frac{2}{3})} - \sum_{k=0}^{\infty} \frac{z^{3k+1}}{3^{2k+4/3}k!\Gamma(k+\frac{4}{3})}, \qquad |z| < \infty, \\ \mathrm{Bi}(z) &= 3^{1/2}\left[\sum_{k=0}^{\infty} \frac{z^{3k}}{3^{2k+2/3}k!\Gamma(k+\frac{2}{3})} + \sum_{k=0}^{\infty} \frac{z^{3k+1}}{3^{2k+4/3}k!\Gamma(k+\frac{4}{3})}\right], \qquad |z| < \infty, \end{aligned} \tag{5.17.3}$$

which show that the Airy functions are entire functions of z.

We can also write (5.17.3) in another, somewhat more concise form. For example, the first expansion is equivalent to

$$\mathrm{Ai}(z) = \frac{2}{3^{7/6}} \sum_{k=0}^{\infty} \frac{\sin\left[\frac{2\pi}{3}(k+1\right]}{\Gamma\left(\frac{k+2}{3}\right)\Gamma\left(\frac{k+3}{3}\right)} \left(\frac{z}{3^{2/3}}\right)^k, \qquad |z| < \infty. \tag{5.17.4}$$

Using the "triplication formula" for the gamma function [Problem 4, formula (i), p. 14] we can transform (5.17.4) into

$$\mathrm{Ai}(z) = \frac{3^{-2/3}}{\pi} \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{k+1}{3}\right) \sin\frac{2\pi}{3}(k+1)}{k!} (3^{1/3}z)^k, \qquad |z| < \infty. \tag{5.17.5}$$

It follows from (5.17.3) that the Airy functions Ai(z) and Bi(z) can be defined as the solutions of equation (5.17.1) satisfying the initial conditions

$$\begin{aligned} u_1(0) &= \mathrm{Ai}(0) = \frac{3^{-2/3}}{\Gamma(\frac{2}{3})}, \qquad u_1'(0) = \mathrm{Ai}'(0) = -\frac{3^{-4/3}}{\Gamma(\frac{4}{3})}, \\ u_2(0) &= \mathrm{Bi}(0) = \frac{3^{-1/6}}{\Gamma(\frac{2}{3})}, \qquad u_2'(0) = \mathrm{Bi}'(0) = \frac{3^{-5/6}}{\Gamma(\frac{4}{3})}. \end{aligned} \tag{5.17.6}$$

The Wronskian of this pair of solutions is

$$W\{\mathrm{Ai}(z), \mathrm{Bi}(z)\} = W\{\mathrm{Ai}(z), \mathrm{Bi}(z)\}_{z=0} = \frac{1}{\pi}, \tag{5.17.7}$$

where we again use the triplication formula for the gamma function.[41] We can also calculate (5.17.7) directly from (5.17.2) and (5.9.5).

Asymptotic representations of the Airy functions for large $|z|$ can be deduced from the corresponding results of Sec. 5.11. In particular, we have

$$\mathrm{Ai}(z) = \frac{\pi^{-1/2}}{2} z^{-1/4} e^{-\frac{2}{3}z^{3/2}}[1 + O(|z|^{-3/2})], \qquad |\arg z| \leqslant \frac{2\pi}{3} - \delta, \tag{5.17.8}$$

$$\mathrm{Bi}(z) = \pi^{-1/2} z^{-1/4} e^{\frac{2}{3}z^{3/2}}[1 + O(|z|^{-3/2})], \qquad |\arg z| \leqslant \frac{\pi}{3} - \delta. \tag{5.17.9}$$

It follows at once from (5.17.3), (5.7.1) and (5.3.2) that the Airy functions of argument $-z$ can be expressed in terms of Bessel functions of the first kind of order $\nu = \pm\frac{1}{3}$:

$$\begin{aligned} \mathrm{Ai}(-z) &= \frac{z^{1/2}}{3}[J_{-1/3}(\tfrac{2}{3}z^{3/2}) + J_{1/3}(\tfrac{2}{3}z^{3/2})], \qquad |\arg z| < \frac{2\pi}{3}, \\ \mathrm{Bi}(-z) &= \left(\frac{z}{3}\right)^{1/2}[J_{-1/3}(\tfrac{2}{3}z^{3/2}) - J_{1/3}(\tfrac{2}{3}z^{3/2})], \qquad |\arg z| < \frac{2\pi}{3}. \end{aligned} \tag{5.17.10}$$

Then, using the asymptotic representation (5.11.6), we find that

$$\begin{aligned} \mathrm{Ai}(-x) &\approx \pi^{-1/2}x^{-1/4}\cos\left(\frac{2}{3}x^{3/2} - \frac{\pi}{4}\right), \qquad x \to \infty, \\ \mathrm{Bi}(-x) &\approx -\pi^{-1/2}x^{-1/4}\sin\left(\frac{2}{3}x^{3/2} - \frac{\pi}{4}\right), \qquad x \to \infty, \end{aligned} \tag{5.17.11}$$

which shows that the Airy functions have an oscillatory character for large negative values of the argument.

Finally, we note that the definition of $\mathrm{Ai}(x)$ and the integral representation of Macdonald's function given in Problem 6, formula (ii), p. 140, imply

$$\mathrm{Ai}(x) = \frac{2x^{1/2}}{3\pi}\int_0^\infty \cos\left(\frac{2x^{3/2}}{3}\sinh y\right)\cosh\frac{y}{3}\,dy, \qquad x > 0.$$

After making the substitution

$$\sinh\frac{y}{3} = \frac{1}{2}x^{-1/2}t,$$

this gives the following integral representation of $\mathrm{Ai}(x)$:

$$\mathrm{Ai}(x) = \frac{1}{\pi}\int_0^\infty \cos(\tfrac{1}{3}t^3 + xt)\,dt, \qquad x \geqslant 0. \tag{5.17.12}$$

[41] For a proof of the first equality in (5.17.7), cf. E. A. Coddington, *op. cit.*, Theorem 8, p. 113.

A somewhat more complicated argument gives the following integral representation of $\mathrm{Bi}(x)$:[42]

$$\mathrm{Bi}(x) = \frac{1}{\pi}\int_0^\infty [e^{-\frac{1}{3}t^3+xt} + \sin(\tfrac{1}{3}t^3 + xt)]\, dt, \qquad x \geqslant 0.$$

For an integral representation of $[\mathrm{Ai}(x)]^2$, see Problem 22, p. 142.

PROBLEMS

1. Derive the integral representation[43]

$$J_n^2(z) = \frac{2}{\pi}\int_0^{\pi/2} J_{2n}(2z\cos\theta)\, d\theta = (-1)^n \frac{2}{\pi}\int_0^{\pi/2} J_0(2z\cos\theta)\cos 2n\theta\, d\theta,$$
$$n = 0, 1, 2, \ldots$$

2. Derive the following formula involving products of Bessel functions:[44]

$$J_\mu(z)J_\nu(z) = \frac{2}{\pi}\int_0^{\pi/2} J_{\mu+\nu}(2z\cos\theta)\cos(\mu-\nu)\theta\, d\theta, \qquad \mathrm{Re}\,(\mu+\nu) > -1.$$

3. Prove that

$$J_n(z)J_n(z') = \frac{1}{\pi}\int_0^\pi J_0(\sqrt{z^2 + z'^2 - 2zz'\cos\theta})\cos n\theta\, d\theta, \qquad n = 0, 1, 2, \ldots$$

Hint. Use the addition theorem (5.12.2).

4. Derive the integral representations

$$J_\nu(x) = \frac{2}{\pi}\int_0^\infty \sin\left(x\cosh t - \frac{\nu\pi}{2}\right)\cosh \nu t\, dt, \qquad -1 < \mathrm{Re}\,\nu < 1, \quad x > 0,$$

$$Y_\nu(x) = -\frac{2}{\pi}\int_0^\infty \cos\left(x\cosh t - \frac{\nu\pi}{2}\right)\cosh \nu t\, dt, \qquad -1 < \mathrm{Re}\,\nu < 1, \quad x > 0.$$

Hint. Use formulas (5.10.14, 15).

5. Derive the formulas

$$H_\nu^{(1)}(z) = \left(\frac{2}{\pi z}\right)^{1/2} \frac{e^{i(z-\frac{1}{2}\nu\pi-\frac{1}{4}\pi)}}{\Gamma(\nu+\frac{1}{2})}\int_0^\infty e^{-s}s^{\nu-\frac{1}{2}}\left(1 - \frac{s}{2iz}\right)^{\nu-\frac{1}{2}} ds,$$
$$\mathrm{Re}\,\nu > -\tfrac{1}{2}, \quad -\frac{\pi}{2} < \arg z < \pi,$$

$$H_\nu^{(2)}(z) = \left(\frac{2}{\pi z}\right)^{1/2} \frac{e^{-i(z-\frac{1}{2}\nu\pi-\frac{1}{4}\pi)}}{\Gamma(\nu+\frac{1}{2})}\int_0^\infty e^{-s}s^{\nu-\frac{1}{2}}\left(1 + \frac{s}{2iz}\right)^{\nu-\frac{1}{2}} ds,$$
$$\mathrm{Re}\,\nu > -\tfrac{1}{2}, \quad -\pi < \arg z < \frac{\pi}{2}.$$

---

[42] H. Jeffreys and B. S. Jeffreys, *op. cit.*, p. 510.

[43] G. N. Watson, *op. cit.*, p. 32.

[44] *Ibid.*, p. 150.

6. Prove the following integral representations of Macdonald's function:[45]

$$K_\nu(z) = \frac{\sqrt{\pi}\, z^\nu}{2^\nu \Gamma(\nu + \frac{1}{2})} \int_0^\infty e^{-z \cosh t} \sinh^{2\nu} t\, dt, \qquad \operatorname{Re} z > 0, \quad \operatorname{Re} \nu > -\tfrac{1}{2},$$

$$K_\nu(x) = \frac{2^\nu \Gamma(\nu + \frac{1}{2})}{x^\nu \sqrt{\pi}} \int_0^\infty \frac{\cos xt}{(1 + t^2)^{\nu + \frac{1}{2}}}\, dt, \qquad x > 0, \quad \operatorname{Re} \nu > -\tfrac{1}{2}, \tag{i}$$

$$K_\nu(x) = \frac{1}{\cos \frac{\nu\pi}{2}} \int_0^\infty \cos (x \sinh t) \cosh \nu t\, dt, \qquad x > 0, \quad |\operatorname{Re} \nu| < 1, \tag{ii}$$

$$K_\nu(z) = \left(\frac{\pi}{2z}\right)^{1/2} \frac{e^{-z}}{\Gamma(\nu + \frac{1}{2})} \int_0^\infty e^{-s} s^{\nu - \frac{1}{2}} \left(1 + \frac{s}{2z}\right)^{\nu - \frac{1}{2}} ds,$$
$$|\arg z| < \pi, \quad \operatorname{Re} \nu > -\tfrac{1}{2}.$$

7. Prove the following formulas involving products of Macdonald functions:[46]

$$K_\nu(x)K_\nu(y) = \frac{1}{2} \int_0^\infty e^{-\frac{1}{2}[t + (x^2 + y^2)/t]} K_\nu\left(\frac{xy}{t}\right) \frac{dt}{t}$$
$$= \int_0^\infty K_0(\sqrt{x^2 + y^2 + 2xy \cosh t}) \cosh \nu t\, dt, \qquad x > 0, \quad y > 0,$$

$$K_\nu(x)K_\nu(y) = \frac{\pi}{2 \sin \nu\pi} \int_{\log (y/x)}^\infty J_0(\sqrt{2xy \cosh t - x^2 - y^2}) \sinh \nu t\, dt,$$
$$x > 0, \quad y > 0, \quad |\operatorname{Re} \nu| < \tfrac{1}{4}. \tag{iii}$$

8. Derive the integral representation

$$I_\nu(x)K_\nu(y) = \frac{1}{2} \int_{\log (y/x)}^\infty J_0(\sqrt{2xy \cosh t - x^2 - y^2})\, e^{-\nu t}\, dt,$$
$$x > 0, \quad y > 0, \quad \operatorname{Re} \nu > -\tfrac{1}{4}.$$

9. Derive the integral representation

$$K_\mu(x)K_\nu(x) = \int_0^\infty K_{\mu-\nu}\left(2x \cosh \frac{t}{2}\right) \cosh \frac{\mu + \nu}{2} t\, dt, \qquad x > 0, \quad y > 0.$$

10. Derive the following asymptotic representations for large values of the order $|\nu|$:

$$J_\nu(z) \approx \frac{1}{\sqrt{2\pi}} e^{\nu + \nu \log (z/2) - (\nu + \frac{1}{2}) \log \nu}, \qquad |\nu| \to \infty, \quad |\arg \nu| \leqslant \pi - \delta,$$

$$K_{i\tau}(x) \approx \left(\frac{2\pi}{\tau}\right)^{1/2} e^{-\pi\tau/2} \sin \left(\frac{\pi}{4} + \tau \log \tau - \tau - \tau \log \frac{x}{2}\right), \qquad \tau \to \infty.$$

(In the second formula, x is a fixed positive number.)

---

[45] G. N. Watson, *op. cit.*, 172, 183.

[46] Concerning Problems 7–9, see *ibid.*, p. 439. The most detailed investigation of various integral representations of products of cylinder functions is due to A. L. Dixon and W. L. Ferrar, *Integrals for the product of two Bessel functions*, Quart. J. Math. Oxford Ser., **4**, 193 (1933); *Part II*, ibid., **4**, 297 (1933).

11. Prove the formulas

$$\begin{aligned}
J_\nu(-x + i0) - J_\nu(-x - i0) &= 2i \sin \nu\pi \, J_\nu(x),\\
Y_\nu(-x + i0) - Y_\nu(-x - i0) &= 2i[J_\nu(x) \cos \nu\pi + J_{-\nu}(x)],\\
H_\nu^{(1)}(-x + i0) - H_\nu^{(1)}(-x - i0) &= -2[J_{-\nu}(x) + e^{-\nu\pi i} J_\nu(x)],\\
H_\nu^{(2)}(-x + i0) - H_\nu^{(2)}(-x - i0) &= 2[J_\nu(x) + e^{\nu\pi i} J_\nu(x)],
\end{aligned}$$

where $x > 0$, characterizing the behavior of the cylinder functions on the cut $[-\infty, 0]$.

12. Verify that

$$\begin{aligned}
I_\nu(-x + i0) - I_\nu(-x - i0) &= 2i \sin \nu\pi \, I_\nu(x),\\
K_\nu(-x + i0) - K_\nu(-x - i0) &= -\pi i[I_{-\nu}(x) + I_\nu(x)],
\end{aligned}$$

where $x > 0$.

Comment. The formulas given in Problems 11–12 take a particularly simple form if $\nu = n$ ($n = 0, \pm 1, \pm 2, \ldots$).

13. Verify the expansion

$$\int_0^z J_\nu(t)\, dt = 2 \sum_{k=0}^{\infty} J_{\nu+2k+1}(z), \qquad \operatorname{Re} \nu > -1.$$

Hint. Use the recurrence relation (5.3.6) to show that both sides have the same derivative.

14. Derive the recurrence relation

$$\int_0^z t^\mu J_\nu(t)\, dt = z^\mu J_{\nu+1}(z) - (\mu - \nu - 1) \int_0^z t^{\mu-1} J_{\nu+1}(t)\, dt, \quad \operatorname{Re}(\mu + \nu) > -1.$$

Hint. Apply (5.3.5) in the form

$$t^{\nu+1} J_\nu(t) = \frac{d}{dt}[t^{\nu+1} J_{\nu+1}(t)],$$

and then integrate by parts.

15. Using the result of Problem 14, show that the evaluation of integrals of the form

$$\int_0^z t^m J_\nu(t)\, dt, \qquad \operatorname{Re} \nu > -1, \quad m = 0, 1, 2, \ldots$$

reduces to the evaluation of the integral

$$\int_0^z J_{\nu+m}(t)\, dt,$$

whose value was found in Problem 13.

Comment. If $\nu = \pm(m - 1), \ \pm(m - 3), \ \pm(m - 5), \ldots,$ then the coefficient of the last integral vanishes, and the original integral can be expressed in closed form in terms of Bessel functions.

16. Verify the formula[47]

$$\int_0^\infty \frac{J_\nu(x)}{x^\mu}\,dx = \frac{\Gamma\left(\dfrac{\nu+1-\mu}{2}\right)}{2^\mu\Gamma\left(\dfrac{\nu+1+\mu}{2}\right)}, \qquad \operatorname{Re}\mu > \tfrac{1}{2}, \quad \operatorname{Re}(\nu-\mu) > -1.$$

17. Verify the formula

$$\int_0^\infty e^{-ax}J_\nu(bx)\,dx = \frac{[\sqrt{a^2+b^2}-a]^\nu}{b^\nu\sqrt{a^2+b^2}}, \qquad \operatorname{Re}\nu > -1, \quad a > 0, \quad b > 0.$$

18. Show that the Bessel function $J_0(x)$ satisfies the following integral equal tion:

$$J_0(x) = \frac{2}{\pi}\int_0^\infty \frac{\sin(x+y)}{x+y}J_0(y)\,dy, \qquad 0 \leqslant x < \infty.$$

19. The *integral Bessel function of order* ν is defined by the formula

$$\operatorname{Ji}_\nu(z) = \int_\infty^z \frac{J_\nu(t)}{t}\,dt, \qquad |\arg z| < \pi.$$

Show that $\operatorname{Ji}_\nu(z)$ is an entire function of ν and an analytic function of z in the plane cut along the segment $[-\infty, 0]$ (in fact, an entire function of z for $\nu = \pm 1, \pm 2\ldots$). Verify the formulas

$$\nu\operatorname{Ji}_\nu(z) = \nu\int_0^z \frac{J_\nu(t)}{t}\,dt - 1, \tag{iv}$$

$$\nu\operatorname{Ji}_\nu(z) = \int_0^z J_{\nu-1}(t)\,dt - J_\nu(t) - 1, \qquad \operatorname{Re}\nu > 0, \quad |\arg z| < \pi.$$

Hint. Use the results of Problems 14 and 16.

20. Prove the following expansions of the integral Bessel functions:

$$\operatorname{Ji}_0(z) = \log\frac{z}{2} + \gamma + \sum_{k=1}^\infty \frac{(-1)^k(z/2)^{2k}}{(2k)(k!)^2}, \qquad |z| < \infty, \quad |\arg z| < \pi$$

$$\operatorname{Ji}_n(z) = -\frac{1}{n} + \sum_{k=0}^\infty \frac{(-1)^k(z/2)^{2k+n}}{(2k+n)k!(n+k)!}, \qquad |z| < \infty, \quad n = 1, 2, \ldots$$

Hint. Substitute (5.3.2) into Problem 19, formula (iv).

21. Derive the asymptotic formula

$$\operatorname{Ji}_\nu(x) \approx \left(\frac{2}{\pi x}\right)^{1/2}\frac{\sin(x - \frac{1}{2}\nu\pi - \frac{1}{4}\pi)}{x}.$$

22. Prove the integral representation

$$[\operatorname{Ai}(x)]^2 = \frac{1}{4\pi\sqrt{3}}\int_0^\infty J_0\left(\frac{1}{12}t^3 + xt\right)t\,dt, \qquad x \geqslant 0$$

for the square of the Airy function of the first kind.

Hint. Use Problem 7, formula (iii).

[47] G. N. Watson, *op. cit.*, p. 391.

2.9 Cylinder Functions: Aplications

6

CYLINDER FUNCTIONS: APPLICATIONS

6.1. Introductory Remarks

As already noted in Sec. 5.1, the cylinder functions have a very wide range of applications to physics and engineering, which cannot even be touched upon in a book of this size. Instead, we confine ourselves to a discussion of a few selected problems of mathematical physics involving cylinder functions,[1] where the selection has been made with the aim of illustrating the application of the theory of Chapter 6. We are mainly concerned with the solution of boundary value problems for various special domains. In addition to several examples of an elementary character, we include some that are more complicated, e.g., the Dirichlet problem for a wedge (see Sec. 6.5).

6.2. Separation of Variables in Cylindrical Coordinates

Consider the partial differential equation

$$\nabla^2 u = \frac{1}{a^2}\frac{\partial^2 u}{\partial t^2} + b\frac{\partial u}{\partial t} + cu, \tag{6.2.1}$$

where ∇^2 is the Laplacian (operator), t is the time, and a, b, c are given constants. A variety of important differential equations occurring in mathematical physics (e.g., in electrodynamics, the theory of vibrations, the theory of heat conduction) are special cases of (6.2.1). The boundary conditions imposed on the function u often require the use of a system of cylindrical

[1] We assume that the reader has already encountered the simplest problems of this type in a first course on mathematical physics.

coordinates r, φ, z, related to the rectangular coordinates x, y, z by the formulas

$$x = r\cos\varphi, \quad y = r\sin\varphi, \quad z = z,$$

where

$$0 \leqslant r < \infty, \quad -\pi < \varphi \leqslant \pi, \quad -\infty < z < \infty.$$

In cylindrical coordinates, equation (6.2.1) becomes

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u}{\partial\varphi^2} + \frac{\partial^2 u}{\partial z^2} = \frac{1}{a^2}\frac{\partial^2 u}{\partial t^2} + b\frac{\partial u}{\partial t} + cu, \tag{6.2.2}$$

and has infinitely many solutions of the form

$$u = R(r)Z(z)\Phi(\varphi)T(t), \tag{6.2.3}$$

where each of the functions on the right depends on only one variable. Substituting (6.2.3) into (6.2.2) and dividing by $RZ\Phi T$, we obtain

$$\frac{1}{Rr}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \frac{1}{r^2\Phi}\frac{d^2\Phi}{d\varphi^2} + \frac{1}{Z}\frac{d^2 Z}{dz^2} - c = \frac{1}{T}\left(\frac{1}{a^2}\frac{d^2 T}{dt^2} + bT\right)\cdot \tag{6.2.4}$$

Since the variables r, φ, z and t are independent, both sides of (6.2.4) must equal a constant, which we denote by $-\varkappa^2$. This leads to two equations

$$\frac{1}{a^2}\frac{d^2 T}{dt^2} + b\frac{dT}{dt} + \varkappa^2 T = 0 \tag{6.2.5}$$

and

$$\frac{1}{Rr}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \varkappa^2 + \frac{1}{r^2}\frac{d^2\Phi}{d\varphi^2} = c - \frac{1}{Z}\frac{d^2 Z}{dz^2}\cdot$$

The same reasoning shows that both sides of the last equation must equal a constant, which this time we denote by $-\lambda^2$, obtaining the equations

$$\frac{d^2 Z}{dz^2} - (\lambda^2 + c)Z = 0 \tag{6.2.6}$$

and

$$r^2\left[\frac{1}{Rr}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + (\lambda^2 + \varkappa^2)\right] = -\frac{1}{\Phi}\frac{d^2\Phi}{d\varphi^2}\cdot$$

Again, both sides of the last equation must equal a constant, denoted by μ^2, which implies

$$\frac{d^2\Phi}{d\varphi^2} + \mu^2\Phi = 0 \tag{6.2.7}$$

and

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \left(\lambda^2 + \varkappa^2 - \frac{\mu^2}{r^2}\right)R = 0. \tag{6.2.8}$$

The process just described is called *separation of variables*, and leads to infinitely many solutions of the form (6.2.3), depending on the parameters $\varkappa$, λ, μ, which can take real or complex values.[2]

Thus, determining the factors in the product (6.2.3) reduces to the relatively simple problem of solving the ordinary differential equations (6.2.5–8). The first three of these equations can be solved in terms of elementary functions, but if we introduce a new variable proportional to r, the fourth equation becomes Bessel's equation, whose solutions involve cylinder functions. The required solution of the given physical problem is obtained by superposition of the particular solutions (6.2.3), where the specific conditions of the problem dictate the choice of the parameters $\varkappa$, λ, μ and the corresponding solutions of (6.2.5–8).

Finally, we call attention to two important special cases of equation (6.2.1), obtained by making certain choices of the constants a, b and c:

1. *Laplace's equation* $\nabla^2 u = 0$ (corresponding to the choice $a = b = c = 0$). This equation has particular solutions of the form

$$u = R(r)Z(z)\Phi(\varphi), \tag{6.2.9}$$

where

$$\begin{gathered} \frac{1}{r}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \left(\lambda^2 - \frac{\mu^2}{r^2}\right)R = 0, \\ \frac{d^2Z}{dz^2} - \lambda^2 Z = 0, \qquad \frac{d^2\Phi}{d\varphi^2} + \mu^2\Phi = 0. \end{gathered} \tag{6.2.10}$$

In the special case where the conditions of the problem are such that u is independent of the angular coordinate φ, we have

$$u = R(r)Z(z) \tag{6.2.11}$$

where

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \lambda^2 R = 0, \qquad \frac{d^2Z}{dz^2} - \lambda^2 Z = 0. \tag{6.2.12}$$

2. *Helmholtz's equation* $\nabla^2 u + k^2 u = 0$ (corresponding to the choice $a = b = 0$, $c = -k^2$). In this case, application of the method of separation of variables leads to particular solutions of the form

$$u = R(r)Z(z)\Phi(\varphi), \tag{6.2.13}$$

where

$$\begin{gathered} \frac{1}{r}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \left(\lambda^2 - \frac{\mu^2}{r^2}\right)R = 0, \\ \frac{d^2Z}{dz^2} - (\lambda^2 - k^2)Z = 0, \qquad \frac{d^2\Phi}{d\varphi^2} + \mu^2\Phi = 0. \end{gathered} \tag{6.2.14}$$

---

[2] Without loss of generality, we can assume that each of the parameters $\varkappa$, λ, μ belongs to an arbitrarily chosen half-plane, since changing the sign of $\varkappa$, λ, μ does not affect the "separation constants" $-\varkappa^2$, $-\lambda^2$, μ^2.

6.3. The Boundary Value Problems of Potential Theory. The Dirichlet Problem for a Cylinder

A function $u = u(x, y, z)$ is said to be *harmonic* in a domain τ if u and its first and second partial derivatives with respect to x, y and z are continuous and satisfy Laplace's equation $\nabla^2 u = 0$ in τ. Consider the problem of finding a function u which is harmonic in τ and satisfies one of the three boundary conditions

$$u|_\sigma = f, \tag{6.3.1a}$$

$$\left.\frac{\partial u}{\partial n}\right|_\sigma = f, \tag{6.3.1b}$$

$$\left.\left(\frac{\partial u}{\partial n} + hu\right)\right|_\sigma = f, \qquad h > 0, \tag{6.3.1c}$$

where σ is the boundary of τ, f is a given function of a variable point of σ,[3] and $\partial/\partial n$ denotes the derivative with respect to the exterior normal to σ. This problem is called the *first boundary value problem of potential theory* or the *Dirichlet problem* if the boundary condition is of the form (6.3.1a), the *second boundary value problem of potential theory* or the *Neumann problem* if it is of the form (6.3.1b), and the *third* or *mixed boundary value problem of potential theory* if it is of the form (6.3.1c). These problems play a very important role in mathematical physics.[4] We now consider the Dirichlet problem for the case where τ is a cylinder of length l and radius a.

Let r, φ, z be a cylindrical coordinate system, with z-axis along the axis of the cylinder and origin in one face of the cylinder (see Figure 18). To satisfy the boundary condition (6.3.1a), we first solve two simpler problems corresponding to the boundary conditions

$$u|_{r=a} = 0, \qquad u|_{z=0} = f_0, \qquad u|_{z=l} = f_l, \tag{6.3.2a}$$

$$u|_{r=a} = F, \qquad u|_{z=0} = u|_{z=l} = 0. \tag{6.3.2b}$$

(In the first case, f vanishes on the lateral surface of the cylinder, and in the second case, f vanishes on the ends of the cylinder.) Obviously, the sum of the solutions satisfying the boundary conditions (6.3.2a) and (6.3.2b) will then satisfy the more general boundary condition (6.3.1a).[5]

[3] If $f \equiv 0$, the boundary condition is said to be *homogeneous*, and otherwise *inhomogeneous*. Here it is assumed that u is continuous in the closed domain $\tau + \sigma$ (cf. Sec. 8.1).

[4] For a more detailed formulation of boundary value problems, and for conditions guaranteeing the existence and uniqueness of solutions under various assumptions concerning the domain τ and the boundary function f, see the books by Frank and von Mises, Tikhonov and Samarski, Courant and Hilbert, and Smirnov (Vol. IV), cited in the Bibliography on p. 300.

[5] It should be noted that in many problems involving inhomogeneous boundary conditions, repeated use of the superposition method leads to solutions of excessively complicated form. This can often be avoided by using another method, due to G. A. Grinberg, *Selected Topics in the Mathematical Theory of Electric and Magnetic Phenomena* (in Russian), Izd. Akad. Nauk SSSR, Moscow (1948).

For simplicity, we temporarily assume that the boundary conditions are independent of the angular coordinate φ, so that

$$f_0 = f_0(r), \qquad f_l = f_l(r), \qquad F = F(z).$$

Then the solution u will also be independent of φ, and therefore, according to (6.2.11, 12) the particular solutions of Laplace's equation take the form $u = R(r)Z(z)$, where $R(r)$ and $Z(z)$ satisfy the differential equations

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \lambda^2 R = 0, \qquad \frac{d^2Z}{dz^2} - \lambda^2 Z = 0. \tag{6.3.3}$$

Solving these equations, we find that

$$R = AJ_0(\lambda r) + BY_0(\lambda r), \qquad Z = C\cosh\lambda z + D\sinh\lambda z, \tag{6.3.4}$$

where $J_0(x)$ and $Y_0(x)$ are Bessel functions of order zero, of the first and second kinds, respectively.

First we consider the boundary conditions (6.3.2a). Since $J_0(\lambda r) \to 1$, $Y_0(\lambda r) \to \infty$ as $r \to 0$, and since the solution R must satisfy the physical requirement of being bounded on the axis of the cylinder, the constant B must equal zero. Then the homogeneous boundary condition becomes

$$AJ_0(\lambda a) = 0,$$

and hence the admissible values of the parameter λ are $\lambda_n = x_n/a$, where the x_n are the positive zeros of the Bessel function $J_0(x)$ [see Sec. 5.13]. Thus we obtain the following set of particular solutions of Laplace's equation:

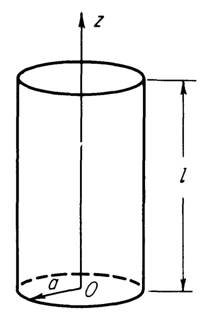


FIGURE 18

$$u = u_n = \left[M_n\cosh\left(x_n\frac{z}{a}\right) + N_n\sinh\left(x_n\frac{z}{a}\right)\right]J_0\left(x_n\frac{r}{a}\right), \qquad n = 1, 2, \ldots \tag{6.3.5}$$

By superposition of these solutions, we can construct a solution of our problem. In fact, suppose each of the functions $f_0(r)$ and $f_l(r)$ can be expanded in a Fourier-Bessel series (see Sec. 5.14), i.e.,

$$f_0(r) = \sum_{n=1}^{\infty} f_{0,n}J_0\left(x_n\frac{r}{a}\right), \qquad f_l(r) = \sum_{n=1}^{\infty} f_{l,n}J_0\left(x_n\frac{r}{a}\right), \tag{6.3.6}$$

where

$$f_{p,n} = \frac{2}{a^2J_1^2(x_n)}\int_0^a rf_p(r)J_0\left(x_n\frac{r}{a}\right)dr, \qquad p = 0, l. \tag{6.3.7}$$

Then the series

$$u = \sum_{n=1}^{\infty}\left[f_{0,n}\frac{\sinh\left(x_n\dfrac{l-z}{a}\right)}{\sinh\left(x_n\dfrac{l}{a}\right)} + f_{l,n}\frac{\sinh\left(x_n\dfrac{z}{a}\right)}{\sinh\left(x_n\dfrac{l}{a}\right)}\right]J_0\left(x_n\frac{r}{a}\right), \tag{6.3.8}$$

whose terms are of the form (6.3.5), clearly satisfies both Laplace's equation and the boundary conditions (6.3.2a).[6]

Next we consider the boundary conditions (6.3.2b). In this case, we must set $C = 0$ and choose

$$\lambda = \frac{n\pi i}{l}, \qquad n = 1, 2, \dots$$

if the homogeneous boundary conditions are to be satisfied. Then the solutions of (6.3.3) take the form

$$\begin{gathered} R = AI_0\left(\frac{n\pi r}{l}\right) + BK_0\left(\frac{n\pi r}{l}\right), \\ Z = D \sin\left(\frac{n\pi z}{l}\right), \end{gathered} \tag{6.3.9}$$

where $I_0(x)$ and $K_0(x)$ are Bessel functions of imaginary argument (see Sec. 5.7). Since $K_0(n\pi r/l) \to \infty$ as $r \to 0$, we must also set $B = 0$. Therefore the particular solutions of Laplace's equation are now

$$u = u_n = M_n I_0\left(\frac{n\pi r}{l}\right) \sin\left(\frac{n\pi z}{l}\right), \qquad n = 1, 2, \dots \tag{6.3.10}$$

Applying the *superposition method* just described,[7] we find that the solution of Laplace's equation satisfying the boundary conditions (6.3.2b) is given by the series

$$u = \sum_{n=1}^{\infty} F_n \frac{I_0\left(\frac{n\pi r}{l}\right)}{I_0\left(\frac{n\pi a}{l}\right)} \sin\frac{n\pi z}{l}, \tag{6.3.11}$$

where the F_n are the Fourier coefficients of $F(z)$ in a series expansion with respect to the functions $\sin(n\pi z/l)$:

$$F_n = \frac{2}{l}\int_0^l F(z) \sin\frac{n\pi z}{l}\, dz. \tag{6.3.12}$$

Remark 1. The solution of the Neumann problem and the mixed problem, involving the boundary conditions (6.3.1b) and (6.3.1c), is obtained in the same way, but now we must use *Dini series* (see Sec. 5.14) instead of Fourier-Bessel series.

Remark 2. To generalize our results to the case of boundary conditions involving the angular coordinate φ, we construct particular solutions of the

[6] Here we have in mind formal solutions, whose validity needs subsequent verification. A somewhat more rigorous point of view is adopted in Chap. 8 (cf. p. 208).

[7] Often called the *Fourier method*, or the *eigenfunction method.*

more general form (6.2.9), satisfying the equations (6.2.10). The values of the parameter μ are now determined by imposing the continuity conditions

$$u|_{\varphi=-\pi} = u|_{\varphi=\pi}, \qquad \left.\frac{\partial u}{\partial \varphi}\right|_{\varphi=-\pi} = \left.\frac{\partial u}{\partial \varphi}\right|_{\varphi=\pi}. \tag{6.3.13}$$

This is equivalent to the physical requirement that the solutions be periodic in φ, and gives $\mu = m$ $(m = 0, 1, 2, \ldots)$. The rest of the analysis differs only slightly from that just given, and leads to the following particular solutions of Laplace's equation

$$u = u_{mn} = \left[M_{mn} \cosh\left(x_{mn}\frac{z}{a}\right) + N_{mn} \sinh\left(x_{mn}\frac{z}{a}\right)\right] J_m\left(x_{mn}\frac{r}{a}\right) \begin{matrix}\cos m\varphi,\\ \sin m\varphi\end{matrix} \tag{6.3.14}$$

$$u = u_{mn} = M_{mn} I_m\left(\frac{n\pi r}{l}\right) \sin\frac{n\pi z}{l} \begin{matrix}\cos m\varphi,\\ \sin m\varphi\end{matrix} \tag{6.3.15}$$

corresponding to (6.3.2a) and (6.3.2b), respectively, where the numbers x_{mn} $(m = 0, 1, 2, \ldots;\ n = 1, 2, \ldots)$ denote the positive zeros of the Bessel function $J_m(x)$. Then the boundary value problems are solved by superpositions of these solutions in the form of double series, with coefficients obtained by expanding the functions

$$f_0 = f_0(r, \varphi), \qquad f_l = f_l(r, \varphi), \qquad F = F(z, \varphi)$$

in appropriate double series.

Example. *Find the stationary distribution of temperature u in a cylinder of length l and radius a, with one end held at temperature u_0, while the rest of the surface is held at temperature zero.*

The desired solution is found at once from (6.3.8) by setting $f_0 = u_0$, $f_l = 0$, and using (5.3.5) to evaluate the integral (6.3.7):

$$u = 2u_0 \sum_{n=1}^{\infty} \frac{\sinh\left(x_n \dfrac{l-z}{a}\right)}{\sinh\left(x_n \dfrac{l}{a}\right)} \frac{J_0\left(x_n \dfrac{r}{a}\right)}{x_n J_1(x_n)}. \tag{6.3.16}$$

6.4 The Dirichlet Problem for a Domain Bounded by Two Parallel Planes

Using the superposition method, we can also solve the boundary value problems of potential theory for the domain consisting of the layer between two parallel planes (see Figure 19). Let the boundary conditions be of the form (6.3.1a), and consider the case of rotational symmetry, where the functions f_0 and f_l appearing in the conditions

$$u|_{z=0} = f_0, \qquad u|_{z=l} = f_l \tag{6.4.1}$$

depend only on the variable r. A function which is harmonic in the domain $0 < z < l$ and satisfies the conditions (6.4.1) can be found by integration with respect to λ of the following particular solutions of Laplace's equation:

$$u = u_\lambda = [M_\lambda \cosh \lambda z + N_\lambda \sinh \lambda z]J_0(\lambda r), \qquad \lambda \geqslant 0. \tag{6.4.2}$$

In fact, assuming that each of the functions f_0 and f_l can be represented as a Fourier-Bessel integral (5.14.11), we find that the formal solution of the problem is given by

$$u = \int_0^\infty \lambda J_0(\lambda r)\left[f_{0,\lambda} \frac{\sinh \lambda(l-z)}{\sinh \lambda l} + f_{l,\lambda} \frac{\sinh \lambda z}{\sinh \lambda l}\right] d\lambda, \tag{6.4.3}$$

where

$$f_{p,\lambda} = \int_0^\infty r f_p(r) J_0(\lambda r)\, dr, \qquad p = 0, l. \tag{6.4.4}$$

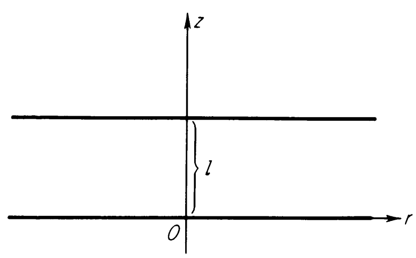


FIGURE 19

The boundary value problem for the half-space $z > 0$ can be solved in the same way. In fact, the solution turns out to be

$$u = \int_0^\infty \lambda J_0(\lambda r) f_\lambda e^{-\lambda z}\, d\lambda,$$

where

$$f_\lambda = \int_0^\infty r f(r) J_0(\lambda r)\, dr,$$

if the boundary condition is of the form

$$u|_{z=0} = f(r).$$

6.5. The Dirichlet Problem for a Wedge

In the case of a wedge-shaped domain, bounded by two intersecting planes (see Figure 20), the boundary value problems of potential theory can also be solved by the superposition method, with the help of cylinder functions. To obtain a suitable set of particular solutions of Laplace's equation $\nabla^2 u = 0$, we introduce a cylindrical coordinate system whose z-axis coincides with the line in which the two planes intersect, and we set

$$\lambda = i\sigma, \qquad 0 \leqslant \sigma < \infty,$$
$$\mu = i\tau, \qquad 0 \leqslant \tau < \infty$$

in the differential equations (6.2.10). Then, according to Sec. 5.7, the solutions of these equations become

$$R = AI_{i\tau}(\sigma r) + BK_{i\tau}(\sigma r),$$
$$\Phi = C \cosh \tau\varphi + D \sinh \tau\varphi,$$
$$Z = E \cos \sigma z + F \sin \sigma z,$$

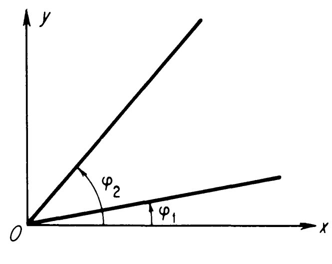


FIGURE 20

where $I_\nu(x)$ and $K_\nu(x)$ are the Bessel functions of imaginary argument, and $A, B, \ldots, F$ are arbitrary constants. Because of the asymptotic behavior of the functions $I_{i\tau}(\sigma r)$ and $K_{i\tau}(\sigma r)$ as $r \to \infty$ (see Sec. 5.11), we must set $A = 0$, which leads to the following set of particular solutions:

$$u = u_{\sigma,\tau} = [M_{\sigma,\tau} \cosh \tau\varphi + N_{\sigma,\tau} \sinh \tau\varphi] K_{i\tau}(\sigma r) \begin{matrix} \cos \sigma z, \\ \sin \sigma z \end{matrix} \qquad (6.5.1)$$
$$0 \leqslant \sigma < \infty, \quad 0 \leqslant \tau < \infty.$$

We now show how to use (6.5.1) to solve the Dirichlet problem for the domain between the two planes $\varphi = \varphi_1$ and $\varphi = \varphi_2$.[8] For simplicity, suppose the functions $f_p = f_p(r, z)$ appearing in the boundary conditions

$$u|_{\varphi=\varphi_p} = f_p, \qquad p = 1, 2 \qquad (6.5.2)$$

are even functions of z, which implies that the same is true of the solution $u = u(r, \varphi, z)$.[9] Assuming that each of the functions f_p can be expanded in a Fourier integral

$$f_p = f_p(r, z) = \int_0^\infty g_p(\sigma, r) \cos \sigma z \, d\sigma, \qquad (6.5.3)$$

where[10]

$$g_p(\sigma, r) = \frac{2}{\pi} \int_0^\infty f_p(r, z) \cos \sigma z \, dz, \qquad (6.5.4)$$

---

[8] It will be assumed that indices are assigned to φ_1, φ_2 in such a way that the domain under consideration corresponds to the interval $\varphi_1 < \varphi < \varphi_2$.

[9] The case where the f_p are odd functions of z is handled in the same way. Then the solution in the general case is represented as the sum of the solutions of the two simpler problems with the following even and odd boundary conditions:

$$u|_{\varphi=\varphi_p} = \tfrac{1}{2}[f_p(r, z) \pm f_p(r, -z)].$$

[10] G. P. Tolstov, *op. cit.*, p. 190.

we try to represent the solution of our problem as a double integral

$$u = \int_0^\infty \cos \sigma z \, d\sigma \int_0^\infty \left[G_1(\sigma, \tau) \frac{\sinh (\varphi_2 - \varphi)\tau}{\sinh (\varphi_2 - \varphi_1)\tau} + G_2(\sigma, \tau) \frac{\sinh (\varphi - \varphi_1)\tau}{\sinh (\varphi_2 - \varphi_1)\tau} \right] K_{i\tau}(\sigma r) \, d\tau, \tag{6.5.5}$$

formed by integrating solutions of the type (6.5.1) with respect to the parameters σ and τ. Clearly, the functions $G_p(\sigma, \tau)$ must satisfy the relation

$$g_p(\sigma, r) = \int_0^\infty G_p(\sigma, \tau) K_{i\tau}(\sigma r) \, d\tau, \qquad 0 < r < \infty, \tag{6.5.6}$$

and hence are the coefficients of the functions $g_p(\sigma, r)$, expanded as integrals with respect to the function $K_{i\tau}(\sigma r)$.

In some cases, we can use formula (5.14.14) to find the functions $G_p(\sigma, \tau)$. In fact, if we write

$$x = \sigma r, \qquad \xi = \sigma \rho, \qquad \sqrt{x}\, f(x) = g(\sigma, r),$$

(5.14.14) becomes

$$g(\sigma, r) = \frac{2}{\pi^2} \int_0^\infty \tau \, K_{i\tau}(\sigma r) \sinh \pi\tau \, d\tau \int_0^\infty g(\sigma, \rho) \frac{K_{i\tau}(\sigma\rho)}{\rho} \, d\rho. \tag{6.5.7}$$

The expansion theorem (6.5.7) is valid if $g(\sigma, r)$, regarded as a function of r, is piecewise continuous and of bounded variation in every finite subinterval $[r_1, r_2]$, where $0 < r_1 < r_2 < \infty$, and if the integrals

$$\int_0^{1/2} |g(\sigma, r)| r^{-1} \log \frac{1}{r} \, dr, \qquad \int_{1/2}^\infty |g(\sigma, r)| r^{-1/2} \, dr \tag{6.5.8}$$

are finite [cf. (5.14.15)]. Provided that the functions $g_p(\sigma, r)$ has these properties, a comparison of (6.5.6) and (6.5.7) shows that

$$G_p(\sigma, \tau) = \frac{2}{\pi^2} \tau \sinh \pi\tau \int_0^\infty g_p(\sigma, r) \frac{K_{i\tau}(\sigma r)}{r} \, dr, \tag{6.5.9}$$

and then (6.5.5) gives a formal solution of the problem. However, it often happens that the first of the integrals (6.5.8) is not finite, since $g_p(\sigma, r)$ generally approaches a nonzero limit $g_p(\sigma, 0)$ as $r \to 0$. To avoid this difficulty, we introduce the modified functions

$$g_p^*(\sigma, r) = g_p(\sigma, r) - g_p(\sigma, 0) e^{-\sigma r}, \qquad p = 1, 2, \tag{6.5.10}$$

and assume, as is usually the case in physical problems, that the conditions for applying formula (6.5.7) are satisfied by $g_p^*(\sigma, r)$. We then have

$$g_p^*(\sigma, r) = \int_0^\infty G_p^*(\sigma, \tau) K_{i\tau}(\sigma r) \, d\tau, \tag{6.5.11}$$

where

$$G_p^*(\sigma, \tau) = \frac{2}{\pi^2}\tau \sinh \pi\tau \int_0^\infty g_p^*(\sigma, r)\frac{K_{i\tau}(\sigma r)}{r}\,dr. \tag{6.5.12}$$

On the other hand, it is easy to prove the formula[11]

$$\frac{2}{\pi}\int_0^\infty K_{i\tau}(x)\,d\tau = e^{-x}, \qquad x > 0, \tag{6.5.13}$$

which implies

$$g_p(\sigma, 0)e^{-\sigma r} = \frac{2}{\pi}g_p(\sigma, 0)\int_0^\infty K_{i\tau}(\sigma r)\,d\tau. \tag{6.5.14}$$

Adding (6.5.11) and (6.5.14), we find the desired representation of $g_p(\sigma, r)$ as an integral with respect to $K_{i\tau}(\sigma r)$. Comparing the result with (6.5.6), we finally obtain

$$G_p(\sigma, \tau) = G_p^*(\sigma, \tau) + \frac{2}{\pi}g_p(\sigma, 0). \tag{6.5.15}$$

and then the solution is given by (6.5.5), as before.

6.6. The Field of a Point Charge near the Edge of a Conducting Sheet

We now illustrate the method developed in the preceding section, by finding the electrostatic field due to a point charge q located near the straight line edge of a thin conducting sheet held at zero potential. To avoid complicating the calculations, we assume that the charge q is at a point A in the same plane as the conducting sheet. Choosing a coordinate system whose z-axis coincides with the edge of the sheet and whose x-axis passes through the point A (see Figure 21), we represent the potential ψ of the electrostatic field as the sum of the potential ψ_0 due to the source and the potential u due to the induced charges:

$$\psi = \psi_0 + u, \qquad \psi_0 = \frac{q}{\sqrt{r^2 + a^2 + 2ar\cos\varphi + z^2}}. \tag{6.6.1}$$

Then the problem reduces to the special case of the general problem of Sec. 6.5 which corresponds to the following choice of angles and boundary conditions:

$$\varphi_1 = 0, \qquad \varphi_2 = 2\pi, \qquad f_1(r, z) = f_2(r, z) = -\frac{q}{\sqrt{(r + a)^2 + z^2}}. \tag{6.6.2}$$

[11] Use (5.10.23) to expand the function $e^{-x\cosh\alpha}$ in a Fourier integral with respect to $\cos\tau\alpha$, obtaining

$$e^{-x\cosh\alpha} = \frac{2}{\pi}\int_0^\infty K_{i\tau}(x)\cos\tau\alpha\,d\tau, \qquad x > 0,$$

and then set $\alpha = 0$.

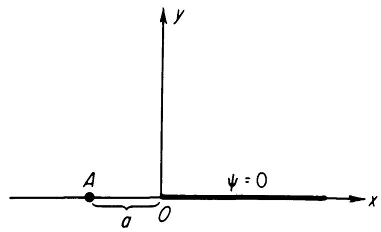


FIGURE 21

Using the integral representation given in Problem 6, formula (i), p. 140, we find that

$$g_p(\sigma, r) = -\frac{2q}{\pi}\int_0^\infty \frac{\cos \sigma z}{\sqrt{(r+a)^2+z^2}}\, dz = -\frac{2q}{\pi} K_0[\sigma(r+a)], \quad (6.6.3)$$

where $K_0(x)$ is Macdonald's function. In the present case,

$$g_p(\sigma, 0) = -\frac{2q}{\pi} K_0(\sigma a),$$

and hence, according to the method of Sec. 6.5, we must first determine the quantity

$$G_p^*(\sigma, \tau) = -\frac{4q}{\pi^3}\tau \sinh \pi\tau \int_0^\infty \frac{K_0[\sigma(r+a)] - K_0(\sigma a)e^{-\sigma r}}{r} K_{i\tau}(\sigma r)\, dr. \quad (6.6.4)$$

Since the evaluation of the integral in (6.6.4) is quite complicated, we omit the details and merely give the final result:

$$G_p^*(\sigma, \tau) = \frac{4q}{\pi^2}[K_0(\sigma a) - K_{i\tau}(\sigma a)]. \quad (6.6.5)$$

Substituting (6.6.5) into (6.5.15), we obtain

$$G_p(\sigma, \tau) = -\frac{4q}{\pi^2} K_{i\tau}(\sigma a), \quad (6.6.6)$$

and then formula (6.5.5) gives

$$u = -\frac{4q}{\pi^2}\int_0^\infty \cos \sigma z\, d\sigma \int_0^\infty \frac{\cosh(\pi - \varphi)\tau}{\cosh \pi\tau} K_{i\tau}(\sigma a) K_{i\tau}(\sigma r)\, d\tau. \quad (6.6.7)$$

The integral in (6.6.7) can be expressed in closed form in terms of elementary functions, and the final result of the calculations turns out to be

$$u = -\frac{q}{\sqrt{r^2 + a^2 + 2ar\cos\varphi + z^2}} \times \left(1 - \frac{2}{\pi}\arctan \frac{2\sqrt{ar}\sin\frac{1}{2}\varphi}{\sqrt{r^2+a^2+2ar\cos\varphi+z^2}}\right) \quad (6.6.8)$$

(we omit the details).[12] It follows from (6.6.8) that

$$\psi = \frac{2q}{\pi\sqrt{r^2 + a^2 + 2ar\cos\varphi + z^2}} \arctan \frac{2\sqrt{ar}\sin\frac{1}{2}\varphi}{\sqrt{r^2 + a^2 + 2ar\cos\varphi + z^2}}. \tag{6.6.9}$$

Finally, we observe that the surface charge density on the sheet is given by the quantity[13]

$$-\frac{1}{4\pi r}\frac{\partial\psi}{\partial\varphi}\bigg|_{\varphi=0} = -\frac{q}{2\pi^2}\sqrt{\frac{a}{r}}\frac{1}{(r+a)^2 + z^2}. \tag{6.6.10}$$

6.7. Cooling of a Heated Cylinder

As an example of the application of cylinder functions to the nonstationary problems of mathematical physics, we now consider the problem of the cooling of an infinitely long cylinder of radius a, heated to the temperature $u_0 = f(r)$ [r is the distance from the axis] and radiating heat into the surrounding medium at zero temperature. From a mathematical point of view, the problem reduces to solving the equation of heat conduction

$$c\rho\frac{\partial u}{\partial t} = k\nabla^2 u, \tag{6.7.1}$$

subject to the boundary condition

$$\left(\frac{\partial u}{\partial r} + hu\right)\bigg|_{r=a} = 0, \tag{6.7.2}$$

and the initial condition

$$u|_{t=0} = u_0 = f(r) \tag{6.7.3}$$

where k, c, ρ, λ and $h = \lambda/k$ have the same meaning as in Sec. 2.6. Separating variables in (6.7.1) by writing $u = R(r)T(t)$, we find the equations

$$b\frac{dT}{dt} + \kappa^2 T = 0, \qquad \frac{1}{r}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \kappa^2 R = 0,$$

where $-\kappa^2$ is the separation constant and $b = c\rho/k$, with solutions

$$R = AJ_0(\kappa r) + BY_0(\kappa r), \qquad T = Ce^{-\kappa^2 t/b}.$$

[12] It should be noted that in the present case, the formula

$$K_0[\sigma(r+a)] = \frac{2}{\pi}\int_0^\infty K_{i\tau}(\sigma a)K_{i\tau}(\sigma r)\,d\tau$$

allows us to derive the solution (6.6.7) without recourse to the general method of expansion as an integral with respect to the functions $K_{i\tau}(\sigma r)$. To obtain this formula, set $\phi = \pi$ in formula (42), p. 55 of the Bateman Manuscript Project, *Higher Transcendental Functions, Vol. 2.*

[13] G. Joos, *op. cit.*, p. 267.

Since $J_0(\varkappa r) \to 1$, $Y_0(\varkappa r) \to \infty$ as $r \to 0$, and since R must satisfy the physical requirement of being bounded on the axis of the cylinder, the constant B must equal zero.

It follows from (6.7.2) that the parameter $\varkappa$ must satisfy the equation

$$hJ_0(\varkappa a) - \varkappa J_1(\varkappa a) = 0. \tag{6.7.4}$$

If we write $x = \varkappa a$, then (6.7.4) becomes

$$haJ_0(x) - xJ_1(x) = 0, \tag{6.7.5}$$

which has only real roots, symmetrically located with respect to the origin (see Sec. 5.13). Let $0 < x_1 < \cdots < x_n < \cdots$ be the positive roots of equation (6.7.5). Then the admissible values of the parameter $\varkappa$ are $\varkappa_n = x_n/a$, and hence the appropriate set of particular solutions of (6.7.1) is

$$u = u_n = M_n J_0\left(x_n \frac{r}{a}\right) e^{-x_n^2 t/a^2 b}, \qquad n = 1, 2, \ldots$$

Superposition of these solutions gives

$$u = \sum_{n=1}^{\infty} M_n J_0\left(x_n \frac{r}{a}\right) e^{-x_n^2 t/a^2 b}, \tag{6.7.6}$$

where, because of the initial condition (6.7.3), the coefficients M_n must be chosen to satisfy the relation

$$f(r) = \sum_{n=1}^{\infty} M_n J_0\left(x_n \frac{r}{a}\right), \qquad 0 \leqslant r < a. \tag{6.7.7}$$

This is just the problem of expanding $f(r)$ in a Dini series, which can be solved by using formulas (5.14.9–10). Thus we have

$$M_n = \frac{2}{a^2[J_0^2(x_n) + J_1^2(x_n)]} \int_0^a rf(r) J_0\left(x_n \frac{r}{a}\right) dr, \tag{6.7.8}$$

and the solution of our heat conduction problem is given by the series (6.7.6), with these values of the coefficients.

6.8 Diffraction by a Cylinder

Finally, we give an example illustrating the application of Bessel functions of the third kind. Consider the diffraction of a plane electromagnetic wave by an infinite conducting cylinder of radius a. Let (r, φ, z) be a system of cylindrical coordinates such that the z-axis coincides with the axis of the cylinder and the angle φ is measured from the direction of propagation of the incident wave. We assume that the time dependence is described by the factor $e^{i\omega t}$, where ω is the angular frequency of the incident radiation, and that the electric vector of the incident wave is parallel to the axis of the cylinder. Then the problem reduces to finding the complex amplitude of the secondary field E satisfying Helmholtz's equation

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial E}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 E}{\partial \varphi^2} + k^2 E = 0, \tag{6.8.1}$$

the boundary condition

$$E|_{r=a} + E_0 e^{-ika\cos\varphi} = 0 \tag{6.8.2}$$

and the *radiation conditions*

$$E = O\left(\frac{1}{\sqrt{r}}\right), \qquad \lim_{r\to\infty} \sqrt{r}\left(\frac{\partial E}{\partial r} + ikE\right) = 0, \tag{6.8.3}$$

where $k = \omega/c$ is the wave number, and E_0 is the amplitude of the incident plane wave.[14]

Applying the method of separation of variables, we find that the particular solutions of (6.8.1), which must also be periodic in φ, are of the form

$$E = E_n = [M_n H_n^{(1)}(kr) + N_n H_n^{(2)}(kr)] \begin{matrix}\cos n\varphi \\ \sin n\varphi\end{matrix}, \qquad n = 0, 1, 2, \ldots, \tag{6.8.4}$$

where $H_n^{(1)}(kr)$, $H_n^{(2)}(kr)$ are the Hankel functions introduced in Sec. 5.6. It follows from the symmetry condition that E is an even function of φ, and hence we need only consider solutions containing $\cos n\varphi$. Moreover, examining the asymptotic behavior of the Hankel functions at infinity, we see that the radiation conditions will be satisfied only if $M_n = 0$ (no incoming waves). Therefore the solution of our problem must have the form

$$E = \sum_{n=0}^{\infty} N_n H_n^{(2)}(kr) \cos n\varphi. \tag{6.8.5}$$

It follows from the boundary condition (6.8.2) that

$$\sum_{n=0}^{\infty} N_n H_n^{(2)}(ka) \cos n\varphi + E_0 e^{-ika\cos\varphi} = 0. \tag{6.8.6}$$

Setting $z = ka$ and $t = -ie^{i\varphi}$ in formula (6.8.4), we obtain

$$e^{-ika\cos\varphi} = J_0(ka) + 2\sum_{n=1}^{\infty} (-i)^n J_n(ka) \cos n\varphi, \tag{6.8.7}$$

which, together with (6.8.5), implies

$$N_0 H_0^{(2)}(ka) = -E_0 J_0(ka), \qquad N_n H_n^{(2)}(ka) = -2E_0(-i)^n J_n(ka).$$

Therefore the required solution is given by

$$E = -E_0\left[\frac{J_0(ka)}{H_0^{(2)}(ka)} H_0^{(2)}(kr) + 2\sum_{n=1}^{\infty} (-i)^n \frac{J_n(ka)}{H_n^{(2)}(ka)} H_n^{(2)}(kr) \cos n\varphi\right]\cdot \tag{6.8.8}$$

---

[14] See A. N. Tikhonov and A. A. Samarski, *Differentialgleichungen der Mathematischen Physik*, VEB Deutscher Verlag der Wissenschaften, Berlin (1959), p. 497.

2.10 Hypergeometric Functions

9

HYPERGEOMETRIC FUNCTIONS

9.1. The Hypergeometric Series and Its Analytic Continuation

By the *hypergeometric series* (already introduced in Sec. 7.2) is meant the power series

$$\sum_{k=0}^{\infty} \frac{(\alpha)_k(\beta)_k}{(\gamma)_k k!} z^k, \tag{9.1.1}$$

where z is a complex variable, α, β and γ are parameters which can take arbitrary real or complex values (provided that $\gamma \neq 0, -1, -2, \ldots$), and the symbol $(\lambda)_k$ denotes the quantity

$$(\lambda)_0 = 1, \quad (\lambda)_k = \frac{\Gamma(\lambda + k)}{\Gamma(\lambda)} = \lambda(\lambda + 1)\cdots(\lambda + k - 1), \quad k = 1, 2, \ldots$$

If either α or β is zero or a negative integer, the series terminates after a finite number of terms, and its sum is then a polynomial in z. Except for this case, the radius of convergence of the hypergeometric series is 1, as is easily seen by using the ratio test.[1]

The sum of the series (9.1.1), i.e., the function

$$F(\alpha, \beta; \gamma; z) = \sum_{k=0}^{\infty} \frac{(\alpha)_k(\beta)_k}{(\gamma)_k k!} z^k, \qquad |z| < 1, \tag{9.1.2}$$

[1] Writing

$$u_k = \frac{(\alpha)_k(\beta)_k}{(\gamma)_k k!} z^k,$$

we have

$$\left|\frac{u_{k+1}}{u_k}\right| = \left|\frac{(\alpha + k)(\beta + k)}{(\gamma + k)(1 + k)} z\right| \to |z|$$

as $k \to \infty$, so that the hypergeometric series converges for $|z| < 1$ and diverges for $|z| > 1$.

is called the *hypergeometric function*, but this definition is only suitable when z lies inside the unit circle. We now show that there exists a complex function which is analytic in the z-plane cut along the segment $[1, \infty]$ and coincides with $F(\alpha, \beta; \gamma; z)$ for $|z| < 1$. This function is the analytic continuation of $F(\alpha, \beta; \gamma; z)$ into the cut plane, and will be denoted by the same symbol. To carry out this analytic continuation, we first assume that $\operatorname{Re}\gamma > \operatorname{Re}\beta > 0$ and use the integral representation

$$\frac{(\beta)_k}{(\gamma)_k} = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)}\int_0^1 t^{\beta-1+k}(1-t)^{\gamma-\beta-1}\,dt, \qquad k = 0, 1, 2, \ldots, \tag{9.1.3}$$

implied by the formulas of Sec. 1.5. Substitution of (9.1.3) into (9.1.2) gives

$$\begin{aligned} F(\alpha, \beta; \gamma; z) &= \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)}\sum_{k=0}^{\infty}\frac{(\alpha)_k}{k!}z^k\int_0^1 t^{\beta-1+k}(1-t)^{\gamma-\beta-1}\,dt \\ &= \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)}\int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}\,dt\sum_{k=0}^{\infty}\frac{(\alpha)_k}{k!}(zt)^k, \end{aligned}$$

where, as usual, reversing the order of summation and integration is justified by an absolute convergence argument.[2] According to the binomial expansion (cf. footnote 17, p. 121),

$$\sum_{k=0}^{\infty}\frac{(\alpha)_k}{k!}(zt)^k = (1-tz)^{-\alpha}, \qquad 0 \leqslant t \leqslant 1, \quad |z| < 1,$$

and hence $F(\alpha, \beta; \gamma; z)$ has the representation

$$F(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)}\int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tz)^{-\alpha}\,dt,$$
$$\operatorname{Re}\gamma > \operatorname{Re}\beta > 0, \quad |z| < 1. \tag{9.1.4}$$

The next step is to show that the integral in (9.1.4) has meaning and represents an analytic function of z in the plane cut along $[1, \infty]$. If z belongs to the closed domain

$$\rho \leqslant |z-1| \leqslant R, \qquad |\arg(1-z)| \leqslant \pi - \delta, \tag{9.1.5}$$

where $R > 0$ is arbitrarily large and $\rho > 0$, $\delta > 0$ are arbitrarily small, and if $0 < t < 1$, then the integrand

$$t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tz)^{-\alpha}$$

is continuous in t for every z and analytic in z for every t, and we need only

[2] In fact, if $\operatorname{Re}\gamma > \operatorname{Re}\beta > 0$ and $|z| < 1$, then

$$\begin{aligned} \sum_{k=0}^{\infty}\frac{|(\alpha)_k|}{k!}|z|^k\int_0^1 |t^{\beta-1+k}(1-t)^{\gamma-\beta-1}|\,dt &\leqslant \sum_{k=0}^{\infty}\frac{(|\alpha|)_k}{k!}|z|^k\int_0^1 t^{\operatorname{Re}\beta-1+k}(1-t)^{\operatorname{Re}\gamma-\operatorname{Re}\beta-1}\,dt \\ &= \frac{\Gamma(\operatorname{Re}\beta)\Gamma(\operatorname{Re}(\gamma-\beta))}{\Gamma(\operatorname{Re}\gamma)}F(|\alpha|, \operatorname{Re}\beta; \operatorname{Re}\gamma; |z|). \end{aligned}$$

show that the integral is uniformly convergent in the indicated region.[3] But this follows at once from the estimate

$$|t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tz)^{-\alpha}| \leqslant Mt^{\mathrm{Re}\,\beta-1}(1-t)^{\mathrm{Re}\,\gamma-\mathrm{Re}\,\beta-1}$$

where M is the maximum value of the continuous function $|(1-tz)|^{-\alpha}$ for t in $[0, 1]$ and z in the domain (9.1.5), and from the fact that the integral

$$M\int_0^1 t^{\mathrm{Re}\,\beta-1}(1-t)^{\mathrm{Re}\,\gamma-\mathrm{Re}\,\beta-1}\,dt$$

converges for $\mathrm{Re}\,\gamma > \mathrm{Re} > \beta > 0$. Therefore the condition $|z| < 1$ can be dropped in (9.1.4), and the desired analytic continuation of the hypergeometric function is given by the formula

$$F(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)}\int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tz)^{-\alpha}\,dt,$$
$$\mathrm{Re}\,\gamma > \mathrm{Re}\,\beta > 0, \quad |\arg(1-z)| < \pi. \quad (9.1.6)$$

In the general case where the parameters have arbitrary values, the analytic continuation of $F(\alpha, \beta; \gamma; z)$ into the plane cut along $[1, \infty]$ can be written as a contour integral obtained by using residue theory to sum the series (9.1.2).[4] A more elementary method of carrying out the analytic continuation, which, however, does not lead to a general analytic expression for the hypergeometric function in explicit form, involves the use of the recurrence relation[5]

$$\begin{aligned}\gamma(\gamma+1)F(\alpha, \beta; \gamma; z) &= \gamma(\gamma-\alpha+1)F(\alpha, \beta+1; \gamma+2; z)\\ &\quad + \alpha[\gamma-(\gamma-\beta)z]F(\alpha+1, \beta+1; \gamma+2; z).\end{aligned} \quad (9.1.7)$$

By repeated application of this identity, we can represent the function $F(\alpha, \beta; \gamma; z)$ with arbitrary parameters ($\gamma \neq 0, -1, -2, \ldots$) as a sum

$$F(\alpha, \beta; \gamma; z) = \sum_{s=0}^{p} a_{sp}(\alpha, \beta; \gamma; z)F(\alpha+s, \beta+p; \gamma+2p; z), \quad (9.1.8)$$

where p is a positive integer and the $a_{sp}(\alpha, \beta; \gamma; z)$ are polynomials in z. If we

---

[3] E. C. Titchmarsh, *op. cit.*, pp. 99–100.

[4] E. T. Whittaker and G. N. Watson, *op. cit.*, p. 288.

[5] To verify (9.1.7), we substitute from (9.1.2), noting that the coefficient of z^k in the right-hand side of (9.1.7) becomes

$$\gamma(\gamma-\alpha+1)\frac{(\alpha)_k(\beta+1)_k}{(\gamma+2)_k k!} + \alpha\gamma\frac{(\alpha+1)_k(\beta+1)_k}{(\gamma+2)_k k!} - \alpha(\gamma-\beta)\frac{(\alpha+1)_{k-1}(\beta+1)_{k-1}}{(\gamma+2)_{k-1}(k-1)!}$$
$$= \frac{(\alpha)_k(\beta)_k}{(\gamma+2)_k k!}\left[\gamma(\gamma-\alpha+1)\frac{\beta+k}{\beta} + \alpha\gamma\frac{\alpha+k}{\alpha}\frac{\beta+k}{\beta} - \alpha(\gamma-\beta)\frac{(\gamma+k+1)k}{\alpha\beta}\right]$$
$$= \frac{(\alpha)_k(\beta)_k}{(\gamma+2)_k k!}(\gamma+k)(\gamma+k+1) \equiv \gamma(\gamma+1)\frac{(\alpha)_k(\beta)_k}{(\gamma)_k k!},$$

choose p so large that $\operatorname{Re}\beta > -p$, $\operatorname{Re}(\gamma - \beta) > -p$, then we can use formula (9.1.6) to make the analytic continuation of each of the functions $F(\alpha + s, \beta + p; \gamma + 2p; z)$ appearing in the right-hand side of (9.1.8). Substituting the corresponding expressions into (9.1.8), we obtain the desired analytic continuation of $F(\alpha, \beta; \gamma; z)$, since the resulting function is analytic in the plane cut along $[1, \infty]$ and coincides with (9.1.2) for $|z| < 1$.

The hypergeometric function $F(\alpha, \beta; \gamma; z)$ plays an important role in mathematical analysis and its applications. Introduction of this function allows us to solve many interesting problems, such as conformal mapping of triangular domains bounded by line segments or circular arcs, various problems of quantum mechanics, etc. Moreover, as will be seen in Sec. 9.8, a number of special functions can be expressed in terms of the hypergeometric function, so that the theory of these functions can be regarded as a special case of the general theory developed in this chapter (cf. footnote 20, p. 176).

9.2. Elementary Properties of the Hypergeometric Function

In this section we consider some properties of the hypergeometric function which are immediate consequences of its definition by the series (9.1.2).[6] First of all, observing that the terms of the series do not change if the parameters α and β are permuted, we obtain the *symmetry property*

$$F(\alpha, \beta; \gamma; z) = F(\beta, \alpha; \gamma; z). \tag{9.2.1}$$

Next, differentiating (9.2.1) with respect to z, we find that

$$\begin{aligned}\frac{d}{dz} F(\alpha, \beta; \gamma; z) &= \sum_{k=1}^{\infty} \frac{(\alpha)_k(\beta)_k}{(\gamma)_k(k-1)!} z^{k-1} = \sum_{k=0}^{\infty} \frac{(\alpha)_{k+1}(\beta)_{k+1}}{(\gamma)_{k+1}k!} z^k \\ &= \frac{\alpha\beta}{\gamma} \sum_{k=0}^{\infty} \frac{(\alpha+1)_k(\beta+1)_k}{(\gamma+1)_k k!} z^k = \frac{\alpha\beta}{\gamma} F(\alpha+1, \beta+1; \gamma+1; z),\end{aligned}$$

and hence[7]

$$\frac{d}{dz} F(\alpha, \beta; \gamma; z) = \frac{\alpha\beta}{\gamma} F(\alpha + 1, \beta + 1; \gamma + 1; z). \tag{9.2.2}$$

Repeated application of (9.2.2) leads to the formula

$$\frac{d^m}{dz^m} F(\alpha, \beta; \gamma; z) = \frac{(\alpha)_m(\beta)_m}{(\gamma)_m} F(\alpha + m, \beta + m; \gamma + m; z), \qquad m = 1, 2, \ldots \tag{9.2.3}$$

[6] It follows from the principle of analytic continuation that all the formulas proved here, under the assumption that $|z| < 1$, remain valid in the whole domain of definition of $F(\alpha, \beta; \gamma; z)$.

[7] Cf. formula (7.12.25), p. 197.

From now on, to simplify the notation, we write

$$F(\alpha, \beta; \gamma; z) \equiv F, \qquad F(\alpha \pm 1, \beta; \gamma; z) \equiv F(\alpha \pm 1),$$
$$F(\alpha, \beta \pm 1; \gamma; z) \equiv F(\beta \pm 1), \qquad F(\alpha, \beta; \gamma \pm 1; z) \equiv F(\gamma \pm 1).$$

Then the functions $F(\alpha \pm 1)$, $F(\beta \pm 1)$ and $F(\gamma \pm 1)$ are said to be *contiguous* to F. The function F and any two functions contiguous to F are connected by recurrence relations whose coefficients are linear functions of the variable z.[8] Among the relations of this type we cite the formulas

$$(\gamma - \alpha - \beta)F + \alpha(1 - z)F(\alpha + 1) - (\gamma - \beta)F(\beta - 1) = 0, \qquad (9.2.4)$$
$$(\gamma - \alpha - 1)F + \alpha F(\alpha + 1) - (\gamma - 1)F(\gamma - 1) = 0, \qquad (9.2.5)$$
$$\gamma(1 - z)F - \gamma F(\alpha - 1) + (\gamma - \beta)zF(\gamma + 1) = 0, \qquad (9.2.6)$$

which can be verified by direct substitution of the series (9.1.2). For example, substituting (9.1.2) into (9.2.4), we obtain

$$\begin{aligned}
&(\gamma - \alpha - \beta)F + \alpha(1 - z)F(\alpha + 1) - (\gamma - \beta)F(\beta - 1)\\
&\quad = \sum_{k=1}^{\infty} \Big[(\gamma - \alpha - \beta)\frac{(\alpha)_k(\beta)_k}{(\gamma)_k k!} + \alpha\frac{(\alpha + 1)_k(\beta)_k}{(\gamma)_k k!}\\
&\qquad\qquad - (\gamma - \beta)\frac{(\alpha)_k(\beta - 1)_k}{(\gamma)_k k!} - \alpha\frac{(\alpha + 1)_{k-1}(\beta)_{k-1}}{(\gamma)_{k-1}(k - 1)!}\Big]z^k\\
&\quad = \sum_{k=1}^{\infty} \frac{(\alpha)_k(\beta)_{k-1}}{(\gamma)_k k!}[(\gamma - \alpha - \beta)(\beta + k - 1) + (\alpha + k)(\beta + k - 1)\\
&\qquad\qquad - (\gamma - \beta)(\beta - 1) - (\gamma + k - 1)k]z^k \equiv 0,
\end{aligned}$$

and similarly for (9.2.5–6). Three other formulas are an immediate consequence of (9.2.4–6) and the symmetry condition (9.2.1):

$$(\gamma - \alpha - \beta)F + \beta(1 - z)F(\beta + 1) - (\gamma - \alpha)F(\alpha - 1) = 0, \qquad (9.2.7)$$
$$(\gamma - \beta - 1)F + \beta F(\beta + 1) - (\gamma - 1)F(\gamma - 1) = 0, \qquad (9.2.8)$$
$$\gamma(1 - z)F - \gamma F(\beta - 1) + (\gamma - \alpha)zF(\gamma + 1) = 0. \qquad (9.2.9)$$

The rest of the recurrence relations can be obtained from (9.2.4–9) by eliminating a common contiguous function from an appropriate pair of formulas. For example, combining (9.2.5) and (9.2.8), or (9.2.6) and (9.2.9), we obtain

$$(\alpha - \beta)F - \alpha F(\alpha + 1) + \beta F(\beta + 1) = 0, \qquad (9.2.10)$$
$$(\alpha - \beta)(1 - z)F + (\gamma - \alpha)F(\alpha - 1) - (\gamma - \beta)F(\beta - 1) = 0, \qquad (9.2.11)$$

and so on.[9]

[8] Obviously, the total number of such relations is

$$\binom{6}{2} = 15.$$

[9] The list of all fifteen recurrence relations involving F and its contiguous functions is given in the Bateman Manuscript Project, *Higher Transcendental Functions, Vol. 1*, p. 103.

Besides the recurrence relations just given, there exist similar relations between the function $F(\alpha, \beta; \gamma; z)$ and any pair of functions of the form $F(\alpha + l, \beta + m; \gamma + n; z)$, where l, m and n are arbitrary integers. Some simple relations of this type are[10]

$$F(\alpha, \beta; \gamma; z) - F(\alpha, \beta; \gamma - 1; z) = -\frac{\alpha\beta z}{\gamma(\gamma - 1)} F(\alpha + 1, \beta + 1; \gamma + 1; z), \quad (9.2.12)$$

$$F(\alpha, \beta + 1; \gamma; z) - F(\alpha, \beta; \gamma; z) = \frac{\alpha z}{\gamma} F(\alpha + 1, \beta + 1; \gamma + 1; z), \quad (9.2.13)$$

$$F(\alpha, \beta + 1; \gamma + 1; z) - F(\alpha, \beta; \gamma; z) = \frac{\alpha(\gamma - \beta)z}{\gamma(\gamma + 1)} F(\alpha + 1, \beta + 1; \gamma + 2; z), \quad (9.2.14)$$

$$F(\alpha - 1, \beta + 1; \gamma; z) - F(\alpha, \beta; \gamma; z) = \frac{(\alpha - \beta - 1)z}{\gamma} F(\alpha, \beta + 1; \gamma + 1; z). \quad (9.2.15)$$

Formulas (9.2.12–15) are proved by direct substitution of (9.1.2), or by repeated use of the relations between $F(\alpha, \beta; \gamma; z)$ and its contiguous functions.

Finally, we recall from Sec. 7.2 that the hypergeometric function $u = F(\alpha, \beta; \gamma; z)$ is a solution of the hypergeometric equation

$$z(1 - z)u'' + [\gamma - (\alpha + \beta + 1)z]u' - \alpha\beta u = 0, \quad (9.2.16)$$

which is analytic in a neighborhood of the point $z = 0$.

9.3. Evaluation of $\lim\limits_{z \to 1-} F(\alpha, \beta; \gamma; z)$ for $\operatorname{Re}(\gamma - \alpha - \beta) > 0$

In developing the theory of the hypergeometric function, it is important to know the limit as $z \to 1-$ of the function (9.1.2), where the parameters satisfy the condition $\operatorname{Re}(\gamma - \alpha - \beta) > 0$.[11] Suppose that besides this condition, $\operatorname{Re} \gamma > \operatorname{Re} \beta > 0$ as well. Then the desired result can be obtained by passing to the limit behind the integral sign in (9.1.6), which gives

$$\lim_{z \to 1-} F(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma - \beta)} \int_0^1 t^{\beta - 1} (1 - t)^{\gamma - \alpha - \beta - 1} \, dt,$$

---

[10] Formula (9.1.7) is also a relation of this type.

[11] It can be shown that if this condition is not satisfied, then, with certain exceptions, the sum of the hypergeometric series becomes infinite as $z \to 1-$.

or, in view of (1.5.2, 6),

$$\lim_{z \to 1-} F(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)}, \tag{9.3.1}$$

where, for the time being, we assume that

$$\operatorname{Re}(\gamma - \alpha - \beta) > 0, \qquad \operatorname{Re}\gamma > \operatorname{Re}\beta > 0. \tag{9.3.2}$$

To justify the passage to the limit, it is sufficient to prove that the conditions (9.3.2) imply that the integral (9.1.6) is uniformly convergent for $0 \leqslant z \leqslant 1$. To this end, we note that

$$1 - t \leqslant |1 - tz| \leqslant 1$$

for $0 \leqslant z \leqslant 1$, $0 \leqslant t \leqslant 1$, and hence

$$|t^{\beta - 1}(1 - t)^{\gamma - \beta - 1}(1 - tz)^{-\alpha}| \leqslant t^{\operatorname{Re}\beta - 1}(1 - t)^{\lambda - 1}, \tag{9.3.3}$$

where

$$\lambda = \begin{cases} \operatorname{Re}(\gamma - \alpha - \beta) & \text{if} \quad \operatorname{Re}\alpha > 0, \\ \operatorname{Re}(\gamma - \beta) & \text{if} \quad \operatorname{Re}\alpha < 0. \end{cases}$$

The estimate (9.3.3) shows that the integral (9.1.6) is uniformly convergent for $0 \leqslant z \leqslant 1$, since the integral

$$\int_0^1 t^{\operatorname{Re}\beta - 1}(1 - t)^{\lambda - 1}\, dt,$$

which majorizes (9.1.6), is convergent if the conditions (9.3.2) hold.

We now show that the second of the conditions (9.3.2) is not essential. Suppose that instead of (9.3.2), the parameters of the hypergeometric functions satisfy the weaker inequalities

$$\operatorname{Re}(\gamma - \alpha - \beta) > 0, \quad \operatorname{Re}(\gamma - \beta) > -1, \quad \operatorname{Re}\beta > -1.$$

Then the restrictions under which we proved (9.3.1) are satisfied by each of the hypergeometric functions in the right-hand side of the recurrence relation (9.1.7). It follows that

$$\begin{aligned} \lim_{z \to 1-} F(\alpha, \beta; \gamma; z) &= \frac{\gamma - \alpha + 1}{\gamma + 1} \frac{\Gamma(\gamma + 2)\Gamma(\gamma - \alpha - \beta + 1)}{\Gamma(\gamma - \alpha + 2)\Gamma(\gamma - \beta + 1)} \\ &\quad + \frac{\alpha\beta}{\gamma(\gamma + 1)} \frac{\Gamma(\gamma + 2)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha + 1)\Gamma(\gamma - \beta + 1)} \\ &\equiv \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)}, \end{aligned}$$

which is just the previous result. Repeating this argument, we can prove by induction that

$$\lim_{z \to 1-} F(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)}, \tag{9.3.4}$$

provided only that $\operatorname{Re}(\gamma - \alpha - \beta) > 0$. Formula (9.3.4) plays an important role in the derivation of various relations satisfied by the hypergeometric function.

9.4 $F(\alpha, \beta; \gamma; z)$ as a Function of its Parameters

In this section we show that the function

$$f(\alpha, \beta; \gamma; z) = \frac{1}{\Gamma(\gamma)} F(\alpha, \beta; \gamma; z) \tag{9.4.1}$$

is an entire function of α, β and γ, for fixed z. If $|z| < 1$, the proof is an immediate consequence of the expansion

$$f(\alpha, \beta; \gamma; z) = \sum_{k=0}^{\infty} \frac{(\alpha)_k(\beta)_k}{\Gamma(\gamma + k)k!} z^k, \qquad |z| < 1, \tag{9.4.2}$$

obtained by substituting (9.1.2) into (9.4.1). In fact, since the terms of the series (9.4.2) are entire functions of α, β, γ, and since the series is uniformly convergent in the region $|\alpha| \leqslant A$, $|\beta| \leqslant B$, $|\gamma| \leqslant C$ (where A, B and C are arbitrarily large),[12] it follows that $f(\alpha, \beta; \gamma; z)$ is an entire function of its parameters.

Now let z be an arbitrary point in the complex plane cut along $[1, \infty]$, and consider the formulas

$$\begin{gathered} f(\alpha, \beta; \gamma; z) = \frac{1}{\Gamma(\beta)\Gamma(\gamma - \beta)} \int_0^1 t^{\beta-1}(1 - t)^{\gamma-\beta-1}(1 - tz)^{-\alpha}\, dt, \\ \operatorname{Re}\gamma > \operatorname{Re}\beta > 0, \quad |\arg(1 - z)| < \pi, \end{gathered} \tag{9.4.3}$$

$$\begin{aligned} f(\alpha, \beta; \gamma; z) = {} & \gamma(\gamma - \alpha + 1)f(\alpha, \beta + 1; \gamma + 2; z) \\ & + \alpha[\gamma - (\gamma - \beta)z]f(\alpha + 1, \beta + 1; \gamma + 2; z), \end{aligned} \tag{9.4.4}$$

which are the analogues of (9.1.6) and (9.1.7). Since the integrand in the right-hand side of (9.4.3) is an entire function of the parameters α, β, γ for any t in $(0, 1)$, and since the integral is uniformly convergent in the region

$$|\alpha| \leqslant A, \quad \delta \leqslant \operatorname{Re}\beta \leqslant B, \quad \delta \leqslant \operatorname{Re}(\gamma - \beta) \leqslant C,$$

---

[12] Use the criterion given in footnote 4, p. 102, noting that if

$$u_k = \frac{(\alpha)_k(\beta)_k}{\Gamma(\gamma + k)k!} z^k,$$

then

$$\left|\frac{u_{k+1}}{u_k}\right| = \left|\frac{(\alpha + k)(\beta + k)}{(\gamma + k)(1 + k)} z\right| \leqslant \frac{(A + k)(B + k)}{(k - C)(1 + k)} |z| \leqslant q < 1$$

for $|z| < 1$ and sufficiently large k.

where $\delta > 0$ is arbitrarily small, it follows that $f(\alpha, \beta; \gamma; z)$ is an analytic function of its parameters in the region

$$|\alpha| < \infty, \quad \operatorname{Re} \beta > 0, \quad \operatorname{Re}(\gamma - \beta) > 0.$$

By repeated application of the recurrence relation (9.4.4), we can represent the function $f(\alpha, \beta; \gamma; z)$ as a sum

$$f(\alpha, \beta; \gamma; z) = \sum_{s=0}^{p} b_{sp}(\alpha, \beta; \gamma; z) f(\alpha + s, \beta + p; \gamma + 2p; z), \quad (9.4.5)$$

where the $b_{sp}(\alpha, \beta; \gamma; z)$ are polynomials in α, β, γ and z, and p is a positive integer. As just shown, each term of this sum is an analytic function in the region $|\alpha| < \infty$, $\operatorname{Re} \beta > -p$, $\operatorname{Re}(\gamma - \beta) > -p$, and hence $f(\alpha, \beta; \gamma; z)$ is an entire function of its parameters. It follows that for fixed z in the plane cut along $[1, \infty]$, the hypergeometric function $F(\alpha, \beta; \gamma; z)$ is an entire function of α and β, and a meromorphic function of γ, with simple poles at the points $\gamma = 0, -1, -2, \ldots$

9.5. Linear Transformations of the Hypergeometric Function

Consider the class of all fractional linear transformations

$$z' = \frac{az + b}{cz + d}$$

carrying the points $z = 0, 1, \infty$ into the points $z' = 0, 1, \infty$ chosen in any order. It is easy to see that besides the identity transformation $z' = z$, this class consists of the following five transformations:

$$z' = \frac{z}{z - 1}, \quad z' = 1 - z, \quad z' = \frac{1}{1 - z}, \quad z' = \frac{1}{z}, \quad z' = \frac{z - 1}{z}.$$

We now derive various linear relations connecting the hypergeometric functions with variables z and z'. Relations of this kind are among the most important in the theory of the hypergeometric function, and are known as *linear transformations* of the hypergeometric function. In particular, these formulas enable us to make the analytic continuation of $F(\alpha, \beta; \gamma; z)$ into any part of the plane cut along $[1, \infty]$.[13]

We begin by deriving a relation which is useful in the case where one requires the analytic continuation of the hypergeometric function into the half-plane $\operatorname{Re} z < \frac{1}{2}$. Suppose z belongs to the plane cut along $[1, \infty]$, and assume for the time being that $\operatorname{Re} \gamma > \operatorname{Re} \beta > 0$. Then, using the integral

[13] The theoretical possibility of such an analytic continuation has already been proved in Sec. 9.1.

representation (9.1.6), and introducing the new variable of integration $s = 1 - t$, we find that

$$F(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma - \beta)} \int_0^1 s^{\gamma-\beta-1}(1 - s)^{\beta-1}(1 - z + sz)^{-\alpha}\, ds$$

$$= (1 - z)^{-\alpha} \frac{\Gamma(\gamma)}{\Gamma(\beta')\Gamma(\gamma - \beta')} \int_0^1 s^{\beta'-1}(1 - s)^{\gamma-\beta'-1}(1 - sz')\, ds,$$

where

$$\beta' = \gamma - \beta, \qquad z' = \frac{z}{z - 1},$$

and our assumptions imply that $\operatorname{Re}\gamma > \operatorname{Re}\beta' > 0$, while z' belongs to the plane cut along $[1, \infty]$.[14] According to (9.1.6), the expression on the right is just

$$(1 - z)^{-\alpha}F(\alpha, \beta'; \gamma; z'),$$

and hence

$$F(\alpha, \beta; \gamma; z) = (1 - z)^{-\alpha}F\left(\alpha, \gamma - \beta; \gamma; \frac{z}{z - 1}\right), \qquad |\arg(1 - z)| < \pi. \tag{9.5.1}$$

Formula (9.5.1) was proved under the temporary assumption that $\operatorname{Re}\gamma > \operatorname{Re}\beta > 0$, but, as we know from Sec. 9.4, after dividing by $\Gamma(\gamma)$, both sides become entire functions of β and γ.[15] Therefore, by the principle of analytic continuation, (9.5.1) remains valid for arbitrary β and γ, with the exception of the values $\gamma = 0, -1, -2, \ldots$ for which $F(\alpha, \beta; \gamma; z)$ is not defined. Moreover, if $\operatorname{Re} z < \frac{1}{2}$, then

$$\left|\frac{z}{z - 1}\right| < 1,$$

and the hypergeometric function in the right-hand side of (9.5.1) can be replaced by the sum of the hypergeometric series, i.e., (9.5.1) gives the analytic continuation of $F(\alpha, \beta; \gamma; z)$ into the half-plane $\operatorname{Re} z < \frac{1}{2}$.

Permuting α and β in (9.5.1), and using the symmetry property (9.2.1), we arrive at the relation

$$F(\alpha, \beta; \gamma; z) = (1 - z)^{-\beta}F\left(\gamma - \alpha, \beta; \gamma; \frac{z}{z - 1}\right), \qquad |\arg(1 - z)| < \pi, \tag{9.5.2}$$

which can also be used to make the analytic continuation of the hypergeometric function into the half-plane $\operatorname{Re} z < \frac{1}{2}$. To obtain another important

[14] Note that under the transformation $z' = z/(1 - z)$, the plane cut along $[1, \infty]$ goes into itself.

[15] The expression $F[f(\alpha, \beta, \gamma, \ldots), g(\alpha, \beta, \gamma, \ldots), \ldots]$ is an entire function of $\alpha, \beta, \gamma, \ldots$ if $F, f, g, \ldots$ are entire functions of their arguments.

result, we perform the transformations (9.5.1) and (9.5.2) consecutively, obtaining

$$F(\alpha, \beta; \gamma; z) = (1 - z)^{-\alpha}\left(1 - \frac{z}{z-1}\right)^{-(\gamma-\beta)}F(\gamma - \alpha, \ \gamma - \beta; \gamma; z), \qquad |\arg(1 - z)| < \pi,$$

or

$$F(\alpha, \beta; \gamma; z) = (1 - z)^{\gamma-\alpha-\beta}F(\gamma - \alpha, \gamma - \beta; \gamma; z), \qquad |\arg(1 - z)| < \pi. \tag{9.5.3}$$

To derive a relation between the hypergeometric function with variable z and the hypergeometric function with variable $z' = 1 - z$, we use a general method from the theory of linear differential equations. First we note that the general solution of the hypergeometric equation

$$z(1 - z)u'' + [\gamma - (\alpha + \beta + 1)z]u' - \alpha\beta u = 0 \tag{9.5.4}$$

can be written in the form[16]

$$u = A_1F(\alpha, \beta; \gamma; z) + A_2z^{1-\gamma}F(1 - \gamma + \alpha, 1 - \gamma + \beta; 2 - \gamma; z),$$
$$|\arg(1 - z)| < \pi, \quad |\arg z| < \pi, \quad \gamma \neq 0, \pm 1, \pm 2, \ldots \tag{9.5.5}$$

Under the transformation $z' = 1 - z$, the domain $|\arg(1 - z)| < \pi$, $|\arg z| < \pi$ goes into the domain $|\arg(1 - z')| < \pi$, $|\arg z'| < \pi$, and equation (9.5.4) goes into the hypergeometric equation with parameters $\alpha' = \alpha$, $\beta' = \beta$, $\gamma' = 1 + \alpha + \beta - \gamma$. Therefore the expression

$$u = B_1F(\alpha, \beta; 1 + \alpha + \beta - \gamma; 1 - z) + B_2(1 - z)^{\gamma-\alpha-\beta} \times F(\gamma - \alpha, \gamma - \beta; 1 - \alpha - \beta + \gamma; 1 - z), \tag{9.5.6}$$
$$|\arg(1 - z)| < \pi, \quad |\arg z| < \pi, \quad \alpha + \beta - \gamma \neq 0, \pm 1, \pm 2, \ldots$$

is also a general solution of equation (9.5.4). In particular, this implies the existence of a linear relation of the form

$$F(\alpha, \beta; \gamma; z) = C_1F(\alpha, \beta; 1 + \alpha + \beta - \gamma; 1 - z) + C_2(1 - z)^{\gamma-\alpha-\beta}F(\gamma - \alpha, \gamma - \beta; 1 - \alpha - \beta + \gamma; 1 - z),$$
$$\alpha + \beta - \gamma \neq 0, \pm 1, \pm 2, \ldots$$

To determine the constants C_1 and C_2, we assume temporarily that $\operatorname{Re}(\alpha + \beta) < \operatorname{Re}\gamma < 1$, and then take the limit of the last equality, first as $z \to 1-$ and then as $z \to 0+$. Using (9.3.4), we obtain

$$C_1 = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)},$$

$$C_1\frac{\Gamma(1 + \alpha + \beta - \gamma)\Gamma(1 - \gamma)}{\Gamma(1 + \alpha - \gamma)\Gamma(1 + \beta - \gamma)} + C_2\frac{\Gamma(1 - \alpha - \beta + \gamma)\Gamma(1 - \gamma)}{\Gamma(1 - \alpha)\Gamma(1 - \beta)} = 1.$$

[16] See Sec. 7.2, noting that by the principle of analytic continuation, formula (7.2.6) remains valid in the whole domain $|\arg(1 - z)| < \pi$, $|\arg z| < \pi$.

It follows that

$$C_2 = \frac{\Gamma(\gamma)\Gamma(\alpha + \beta - \gamma)}{\Gamma(\alpha)\Gamma(\beta)},$$

after some simple calculations involving the identity (1.2.2). Therefore the required formula is

$$\begin{aligned} F(\alpha, \beta; \gamma; z) = {} & \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)} F(\alpha, \beta; 1 + \alpha + \beta - \gamma; 1 - z) \\ & + (1 - z)^{\gamma - \alpha - \beta} \frac{\Gamma(\gamma)\Gamma(\alpha + \beta - \gamma)}{\Gamma(\alpha)\Gamma(\beta)} \\ & \quad \times F(\gamma - \alpha, \gamma - \beta; 1 - \alpha - \beta + \gamma; 1 - z), \\ & |\arg z| < \pi, \quad |\arg (1 - z)| < \pi, \quad \alpha + \beta - \gamma \neq 0, \pm 1, \pm 2, \ldots \end{aligned} \tag{9.5.7}$$

To get rid of the superfluous restrictions imposed on the parameters α, β and γ, we note that after multiplication by $\sin \pi(\gamma - \alpha - \beta)/\Gamma(\gamma)$, both sides of (9.5.7) are entire functions of the parameters.[17] Therefore, according to the principle of analytic continuation, the relation (9.5.7) is valid for all values of the parameters except those for which $\alpha + \beta - \gamma = 0, \pm 1, \pm 2, \ldots$ Formula (9.5.7) gives the analytic continuation of the hypergeometric function into the domain $|z - 1| < 1$, $|\arg (1 - z)| < \pi$.

The remaining relations between the hypergeometric functions with variables z and z' can be obtained by combining the formulas just derived. For example, consecutive application of (9.5.1) and (9.5.7) leads to the relation[18]

$$\begin{aligned} F(\alpha, \beta; \gamma; z) = {} & (1 - z)^{-\alpha} \frac{\Gamma(\gamma)\Gamma(\beta - \alpha)}{\Gamma(\gamma - \alpha)\Gamma(\beta)} F\left(\alpha, \gamma - \beta; 1 + \alpha - \beta; \frac{1}{1 - z}\right) \\ & + (1 - z)^{-\beta} \frac{\Gamma(\gamma)\Gamma(\alpha - \beta)}{\Gamma(\gamma - \beta)\Gamma(\alpha)} F\left(\gamma - \alpha, \beta; 1 - \alpha + \beta; \frac{1}{1 - z}\right) \\ & |\arg (-z)| < \pi, \quad |\arg (1 - z)| < \pi, \quad \alpha - \beta \neq 0, \pm 1, \pm 2, \ldots \end{aligned} \tag{9.5.8}$$

which enables us to make the analytic continuation of $F(\alpha, \beta; \gamma; z)$ into the domain $|z - 1| > 1$, $|\arg (1 - z)| < \pi$. Then, combining (9.5.8) with (9.5.1–2), we obtain

$$\begin{aligned} F(\alpha, \beta; \gamma; z) = {} & (-z)^{-\alpha} \frac{\Gamma(\gamma)\Gamma(\beta - \alpha)}{\Gamma(\gamma - \alpha)\Gamma(\beta)} F\left(\alpha, 1 + \alpha - \gamma; 1 + \alpha - \beta; \frac{1}{z}\right) \\ & + (-z)^{-\beta} \frac{\Gamma(\gamma)\Gamma(\alpha - \beta)}{\Gamma(\gamma - \beta)\Gamma(\alpha)} F\left(\beta, 1 + \beta - \gamma; 1 + \beta - \alpha; \frac{1}{z}\right), \\ & |\arg (-z)| < \pi, \quad |\arg (1 - z)| < \pi, \quad \alpha - \beta \neq 0, \pm 1, \pm 2, \ldots, \end{aligned} \tag{9.5.9}$$

---

[17] Here we again make use of (1.2.2).

[18] Note that under the transformation $z' = z/(z - 1)$, the domain $|\arg (-z)| < \pi$, $|\arg (1 - z)| < \pi$ goes into the domain $|\arg z'| < \pi$, $|\arg (1 - z')| < \pi$, which guarantees that (9.5.1) and (9.5.7) can be applied consecutively.

which gives the analytic continuation of $F(\alpha, \beta; \gamma; z)$ into the domain $|z| > 1$, $|\arg(1 - z)| < \pi$. Finally, consecutive application of (9.5.7) and (9.5.1) gives

$$F(\alpha, \beta; \gamma; z) = z^{-\alpha} \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)} F\left(\alpha, 1 + \alpha - \gamma; 1 + \alpha + \beta - \gamma; \frac{z-1}{z}\right)$$
$$+ z^{\alpha-\gamma}(1 - z)^{\gamma-\alpha-\beta} \frac{\Gamma(\gamma)\Gamma(\alpha + \beta - \gamma)}{\Gamma(\alpha)\Gamma(\beta)}$$
$$\times F\left(\gamma - \alpha, 1 - \alpha; 1 + \gamma - \alpha - \beta; \frac{z-1}{z}\right),$$
$$|\arg z| < \pi, \quad |\arg(1 - z)| < \pi, \quad \alpha + \beta - \gamma \neq 0, \pm 1, \pm 2, \ldots, \tag{9.5.10}$$

which can be used to make the analytic continuation of $F(\alpha, \beta; \gamma; z)$ into the domain $\operatorname{Re} z > \frac{1}{2}$, $|\arg(1 - z)| < \pi$.

The problem of the analytic continuation of the hypergeometric function into the z-plane cut along $[1, \infty]$ is solved by using formulas (9.5.1–3) and (9.5.7–10). Some exceptional cases, where these formulas are not applicable, will be considered in Sec. 9.7.

9.6. Quadratic Transformations of the Hypergeometric Function

The relations between hypergeometric functions derived in the preceding section are valid for arbitrary values of the parameters α, β, γ (apart from certain exceptional values). One can also consider relations where the parameters satisfy certain constraints; although less general, relations of this type are also useful in making various transformations and carrying out analytic continuation. Among such relations, the most interesting involve hypergeometric functions with two arbitrary parameters. As will be seen below, they also contain expressions like

$$\frac{1 + \sqrt{1 - z}}{2}, \quad \frac{1 - \sqrt{1 - z}}{1 + \sqrt{1 - z}}, \quad \frac{-4z}{(1 - z)^2}, \ldots,$$

and hence are called *quadratic transformations of the hypergeometric function*.

As an example of a formula belonging to this class, consider the relation

$$F(\alpha, \beta; \alpha + \beta + \tfrac{1}{2}; z) = F\left(2\alpha, 2\beta; \alpha + \beta + \tfrac{1}{2}; \frac{1 - \sqrt{1 - z}}{2}\right),$$
$$|\arg(1 - z)| < \pi, \quad \alpha + \beta + \tfrac{1}{2} \neq 0, -1, -2, \ldots, \tag{9.6.1}$$

which can be proved as follows: The left-hand side is a solution of the hypergeometric equation (9.5.4) with parameter $\gamma = \alpha + \beta + \frac{1}{2}$, which is analytic in the domain $|\arg(1 - z)| < \pi$. Under the substitution[19]

$$z' = \tfrac{1}{2}(1 - \sqrt{1 - z}),$$

[19] By $\sqrt{1 - z}$ is meant the branch which is positive for real z in the interval $(0, 1)$.

this equation goes into an equation of the same form with parameters

$$\alpha' = 2\alpha, \quad \beta' = 2\beta, \quad \gamma' = \alpha + \beta + \tfrac{1}{2},$$

and the domain $|\arg(1 - z)| < \pi$ goes into the domain $\operatorname{Re} z' < \frac{1}{2}$, which is part of the domain $|\arg(1 - z')| < \pi$. But according to (7.2.6), the hypergeometric equation cannot have two linearly independent solutions which are analytic in a neighborhood of the point $z = 0$, and hence there must exist a relation of the form

$$F(\alpha, \beta; \alpha + \beta + \tfrac{1}{2}; z) = AF\left(2\alpha, 2\beta; \alpha + \beta + \tfrac{1}{2}; \frac{1 - \sqrt{1 - z}}{2}\right),$$

where A is a constant. Setting $z = 0$, we find that $A = 1$, thereby proving (9.6.1).

A large number of other relations of the same type can be deduced by applying the linear transformations of Sec. 9.5 to formula (9.6.1) and changing the independent variable or the parameters. For example, using (9.5.3) and (9.5.1) to transform the right-hand side of (9.6.1), we find that

$$\begin{aligned} &F(\alpha, \beta; \alpha + \beta + \tfrac{1}{2}; z) \\ &= \left(\frac{1 + \sqrt{1 - z}}{2}\right)^{-2\alpha} F\left(2\alpha, \alpha - \beta + \tfrac{1}{2}; \alpha + \beta + \tfrac{1}{2}; \frac{\sqrt{1 - z} - 1}{\sqrt{1 - z} + 1}\right), \\ &\qquad |\arg(1 - z)| < \pi, \quad \alpha + \beta + \tfrac{1}{2} \neq 0, -1, -2, \dots, \end{aligned} \tag{9.6.2}$$

$$\begin{aligned} &F(\alpha, \beta; \alpha + \beta + \tfrac{1}{2}; z) \\ &= \left(\frac{1 + \sqrt{1 - z}}{2}\right)^{1/2 - \alpha - \beta} F\left(\alpha - \beta + \tfrac{1}{2}, \beta - \alpha + \tfrac{1}{2}; \alpha + \beta + \tfrac{1}{2}; \frac{1 - \sqrt{1 - z}}{2}\right), \\ &\qquad |\arg(1 - z)| < \pi, \quad \alpha + \beta + \tfrac{1}{2} \neq 0, -1, -2, \dots \end{aligned} \tag{9.6.3}$$

Using (9.5.1) to transform the left-hand sides of (9.6.1) and (9.6.2), and then making the substitution

$$\frac{z}{z - 1} \to z, \qquad \alpha + \beta + \tfrac{1}{2} \to \gamma,$$

we obtain two other useful relations:

$$\begin{aligned} F(\alpha, \alpha + \tfrac{1}{2}; \gamma; z) &= (1 - z)^{-\alpha} F\left(2\alpha, 2\gamma - 2\alpha - 1; \gamma; \frac{1}{2} - \frac{1}{2\sqrt{1 - z}}\right), \\ &\qquad |\arg(1 - z)| < \pi, \end{aligned} \tag{9.6.4}$$

$$\begin{aligned} F(\alpha, \alpha + \tfrac{1}{2}; \gamma; z) &= \left(\frac{1 + \sqrt{1 - z}}{2}\right)^{-2\alpha} F\left(2\alpha, 2\alpha - \gamma + 1; \gamma; \frac{1 - \sqrt{1 - z}}{1 + \sqrt{1 - z}}\right), \\ &\qquad |\arg(1 - z)| < \pi. \end{aligned} \tag{9.6.5}$$

Finally, using (9.5.3) to transform the left-hand sides of (9.6.1) and (9.6.2), and then making the substitution

$$\alpha \to \alpha - \tfrac{1}{2}, \qquad \beta \to \beta - \tfrac{1}{2},$$

we arrive at the relations

$$F(\alpha, \beta; \alpha + \beta - \tfrac{1}{2}; z) = \frac{1}{\sqrt{1-z}} F\left(2\alpha - 1, 2\beta - 1; \alpha + \beta - \tfrac{1}{2}; \frac{1 - \sqrt{1-z}}{2}\right),$$
$$|\arg(1-z)| < \pi, \quad \alpha + \beta - \tfrac{1}{2} \neq 0, -1, -2, \ldots, \tag{9.6.6}$$

$$F(\alpha, \beta; \alpha + \beta - \tfrac{1}{2}; z) = \frac{1}{\sqrt{1-z}} \left(\frac{1 + \sqrt{1-z}}{2}\right)^{1-2\alpha} \times F\left(2\alpha - 1, \alpha - \beta + \tfrac{1}{2}; \alpha + \beta - \tfrac{1}{2}; \frac{\sqrt{1-z} - 1}{\sqrt{1-z} + 1}\right),$$
$$|\arg(1-z)| < \pi, \quad \alpha + \beta - \tfrac{1}{2} \neq 0, -1, -2, \ldots \tag{9.6.7}$$

It is interesting to note that formulas (9.6.2, 5, 7) continue the corresponding hypergeometric functions into the plane cut along $[1, \infty]$. In fact,

$$\left|\frac{1 - \sqrt{1-z}}{1 + \sqrt{1-z}}\right| < 1$$

if $|\arg(1-z)| < \pi$, and hence the hypergeometric function in the right-hand side of each of these formulas can be replaced by the sum of the corresponding hypergeometric series.

Further results can be obtained by taking inverses of the formulas just derived. For example, inversion of (9.6.1–3) gives[20]

$$F\{\alpha, \beta; \tfrac{1}{2}(\alpha + \beta + 1\}; z) = F\{\tfrac{1}{2}\alpha, \tfrac{1}{2}\beta; \tfrac{1}{2}(\alpha + \beta + 1); 4z(1-z)\},$$
$$\operatorname{Re} z < \tfrac{1}{2}, \quad \tfrac{1}{2}(\alpha + \beta + 1) \neq 0, -1, -2, \ldots, \tag{9.6.8}$$

$$F(\alpha, \beta; \alpha - \beta + 1; z) = (1-z)^{-\alpha} F\left\{\tfrac{1}{2}\alpha, \tfrac{1}{2}(\alpha + 1) - \beta; \alpha - \beta + 1; -\frac{4z}{(1-z)^2}\right\},$$
$$|z| < 1, \quad \alpha - \beta + 1 \neq 0, -1, -2, \ldots, \tag{9.6.9}$$

$$F(\alpha, 1 - \alpha; \gamma; z) = (1-z)^{\gamma - 1} F\{\tfrac{1}{2}(\gamma - \alpha), \tfrac{1}{2}(\gamma + \alpha - 1); \gamma; 4z(1-z)\},$$
$$\operatorname{Re} z < \tfrac{1}{2}. \tag{9.6.10}$$

---

[20] In particular, (9.6.8) is obtained from (9.6.1) by making the substitution

$$2\alpha \to \alpha, \quad 2\beta \to \beta, \quad \frac{1 - \sqrt{1-z}}{2} \to z.$$

Moreover, combining these formulas with the linear transformations given in Sec. 9.5, we can obtain still another group of formulas. For example, applying the transformation (9.5.7) to the right-hand side of (9.6.8) and making the substitution

$$\alpha \to 2\alpha, \quad \beta \to 2\beta, \quad z \to \frac{1-z}{2},$$

we find that[21]

$$\begin{aligned} F\Big(2\alpha, 2\beta; \alpha + \beta + \tfrac{1}{2}; \frac{1-z}{2}\Big) &= \frac{\Gamma(\alpha+\beta+\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(\alpha+\frac{1}{2})\Gamma(\beta+\frac{1}{2})} F(\alpha, \beta; \tfrac{1}{2}; z^2) \\ &\quad + z\frac{\Gamma(\alpha+\beta+\frac{1}{2})\Gamma(-\frac{1}{2})}{\Gamma(\alpha)\Gamma(\beta)} F(\alpha+\tfrac{1}{2}, \beta+\tfrac{1}{2}; \tfrac{3}{2}; z^2), \\ &\quad |\arg(1 \pm z)| < \pi, \quad \alpha+\beta+\tfrac{1}{2} \neq 0, -1, -2, \ldots \end{aligned} \tag{9.6.11}$$

Formula (9.6.11) plays an important role in the theory of spherical harmonics. For example, the relation (7.6.9) is an immediate consequence of (9.6.11).

We conclude this section by deriving a few formulas of a more complicated nature. The first result is

$$\begin{aligned} F(\alpha, \beta; 2\beta; z) &= \left(\frac{1+\sqrt{1-z}}{2}\right)^{-2\alpha} F\left\{\alpha, \alpha-\beta+\tfrac{1}{2}; \beta+\tfrac{1}{2}; \left(\frac{1-\sqrt{1-z}}{1+\sqrt{1-z}}\right)^2\right\}, \\ &\quad |\arg(1-z)| < \pi, \quad 2\beta \neq -1, -3, -5, \ldots, \end{aligned} \tag{9.6.12}$$

which is proved in the same way as (9.6.1), by noting that under the change of variables

$$z' = \left(\frac{1-\sqrt{1-z}}{1+\sqrt{1-z}}\right)^2, \qquad u = \left(\frac{1+\sqrt{1-z}}{2}\right)^{-2\alpha} v,$$

equation (9.5.4) goes into the hypergeometric equation with the new parameters

$$\alpha' = \alpha, \qquad \beta' = \alpha - \beta + \tfrac{1}{2}, \qquad \gamma' = \beta + \tfrac{1}{2}.$$

Since the verification of this fact is quite tedious, we supply some intermediate

---

[21] In the course of the derivation, it is convenient to assume temporarily that $|\arg(1-z)| < \pi$, $\operatorname{Re} z > 0$. The result can then be extended the whole domain $|\arg(1 \pm z)| < \pi$ by using the principle of analytic continuation.

steps which will serve to keep the reader on the right track during the course of the calculation:

$$z = 1 - \left(\frac{\sqrt{z'} - 1}{\sqrt{z'} + 1}\right)^2, \qquad \frac{dz'}{dz} = -\frac{z'(\sqrt{z'} + 1)^3}{2(\sqrt{z'} - 1)},$$

$$u = (\sqrt{z'} + 1)^{2\alpha} v, \tag{9.6.13}$$

$$\frac{du}{dz} = \frac{dz'}{dz}\frac{du}{dz'} = -\frac{(\sqrt{z'} + 1)^{2\alpha+2}}{2(\sqrt{z'} - 1)}\left[\alpha v + \sqrt{z'}(\sqrt{z'} + 1)\frac{dv}{dz'}\right], \tag{9.6.14}$$

$$\begin{aligned} z(1 - z)\frac{d^2u}{dz^2} &= \sqrt{z'}(\sqrt{z'} + 1)^{2\alpha}\left\{\left[\alpha + 1 - \frac{1}{2}\frac{\sqrt{z'} + 1}{\sqrt{z'} - 1}\right]\left[\alpha v + \sqrt{z'}(\sqrt{z'} + 1)\frac{dv}{dz'}\right]\right. \\ &\quad \left. + \sqrt{z'}(\sqrt{z'} + 1)\left[\left(\alpha + 1 + \frac{1}{2\sqrt{z'}}\right)\frac{dv}{dz'} + \sqrt{z'}(\sqrt{z'} + 1)\frac{d^2v}{dz'^2}\right]\right\}. \end{aligned} \tag{9.6.15}$$

After using (9.6.13–15) to write the hypergeometric equation satisfied by u, we multiply the result by

$$\frac{1 - \sqrt{z'}}{\sqrt{z'}(1 + \sqrt{z'})},$$

obtaining

$$\begin{aligned} &\frac{1 - \sqrt{z'}}{1 + \sqrt{z'}}\left[\alpha + 1 - \frac{1}{2}\frac{\sqrt{z'} + 1}{\sqrt{z'} - 1}\right]\left[\alpha v + \sqrt{z'}(\sqrt{z'} + 1)\frac{dv}{dz'}\right] \\ &\quad + \sqrt{z'}(1 - \sqrt{z'})\left[\left(\alpha + 1 + \frac{1}{2\sqrt{z'}}\right)\frac{dv}{dz'} + \sqrt{z'}(\sqrt{z'} + 1)\frac{d^2v}{dz'^2}\right] \\ &\quad + \frac{\sqrt{z'} + 1}{\sqrt{z'}}\left[\beta - (\alpha + \beta + 1)\frac{2\sqrt{z'}}{(\sqrt{z'} + 1)^2}\right]\left[\alpha v + \sqrt{z'}(\sqrt{z'} + 1)\frac{dv}{dz'}\right] \\ &\qquad\qquad - \frac{\alpha\beta(1 - \sqrt{z'})}{\sqrt{z'}(1 + \sqrt{z'})}v = 0, \end{aligned}$$

which can now be reduced quite easily to the hypergeometric equation

$$z'(1 - z')\frac{d^2v}{dz'^2} + [(\beta + \tfrac{1}{2}) - (2\alpha - \beta + \tfrac{3}{2})z']\frac{dv}{dz'} - \alpha(\alpha - \beta + \tfrac{1}{2})v = 0,$$

satisfied by v. Making the substitution

$$\frac{1 - \sqrt{1 - z}}{1 + \sqrt{1 - z}} \to z$$

in (9.6.12), we obtain the formula

$$F\left\{\alpha, \beta; 2\beta; \frac{4z}{(1+z)^2}\right\} = (1+z)^{2\alpha} F(\alpha, \alpha - \beta + \tfrac{1}{2}; \beta + \tfrac{1}{2}; z^2),$$

$$|z| < 1, \quad 2\beta \neq -1, -3, -5, \ldots \quad (9.6.16)$$

Our final result is

$$F(\alpha, \beta; 2\beta; z) = \left(1 - \frac{z}{2}\right)^{-\alpha} F\left\{\tfrac{1}{2}\alpha, \tfrac{1}{2}(\alpha+1); \beta + \tfrac{1}{2}; \left(\frac{z}{2-z}\right)^2\right\},$$

$$|\arg(1-z)| < \pi, \quad 2\beta \neq -1, -3, -5, \ldots, \quad (9.6.17)$$

which can be derived as follows: Applying the transformation (9.5.1) to the right-hand side of (9.6.9) and replacing β by $\alpha - \beta + \frac{1}{2}$, we obtain

$$F(\alpha, \alpha - \beta + \tfrac{1}{2}; \beta + \tfrac{1}{2}; z) = (1+z)^{-\alpha} F\left\{\tfrac{1}{2}\alpha, \tfrac{1}{2}(\alpha+1); \beta + \tfrac{1}{2}; \frac{4z}{(1+z)^2}\right\},$$

$$|z| < 1, \quad 2\beta \neq -1, -3, -5, \ldots \quad (9.6.18)$$

Then, comparing (9.6.13) and (9.6.15), we find that

$$F\left\{\alpha, \beta; 2\beta; \frac{4z}{(1+z)^2}\right\} = \frac{(1+z)^{2\alpha}}{(1+z^2)^{\alpha}} F\left\{\tfrac{1}{2}\alpha, \tfrac{1}{2}(\alpha+1); \beta + \tfrac{1}{2}; \frac{4z^2}{(1+z^2)^2}\right\},$$

and the desired result is obtained by making the substitution

$$\frac{4z}{(1+z)^2} \to z,$$

which implies

$$\frac{1+z^2}{(1+z)^2} \to \frac{2-z}{2}, \qquad \frac{4z^2}{(1+z^2)^2} \to \left(\frac{z}{2-z}\right)^2.$$

The theory of quadratic transformations of the hypergeometric function was developed by Gauss, Kummer and Goursat, and also from a more general point of view in Riemann's investigations of a class of differential equations including the hypergeometric equation as a special case.[22] We refer the reader to these sources for a more detailed treatment of the subject.[23]

---

[22] See E. Goursat, *Sur l'équation différentielle linéaire, qui admet pour intégrale la série hypergéometrique*, Ann. Sci. École Norm. Sup. (2), **10**, 3 (1881). The relevant references by Gauss, Kummer and Riemann are given on p. 296 of the book by Whittaker and Watson (*op. cit.*).

[23] See also the Bateman Manuscript Project, *Higher Transcendental Functions, Vol. 1*, p. 110 ff., for an extensive list of quadratic transformations of the hypergeometric function.

9.7. Formulas for Analytic Continuation of $F(\alpha, \beta; \gamma; z)$ in Exceptional Cases

The formulas derived in Sec. 9.5 allow us to obtain the analytic continuation of the hypergeometric function into any part of the z-plane cut along $[1, \infty]$. However, some of these formulas are no longer meaningful for certain values of the parameters, and must therefore be modified in a way we now indicate. The general approach is to start from the formulas of Sec. 9.5 and then carry out appropriate passages to the limit.

For example, suppose we want to find the analytic continuation of the function $F(\alpha, \beta; \gamma; z)$ into the domain $|z - 1| < 1$, $|\arg(1 - z)| < \pi$. If $\alpha + \beta - \gamma \neq 0, \pm 1, \pm 2, \ldots$, we can use (9.5.7), but this formula is not applicable if $\gamma = \alpha + \beta \pm n$ $(n = 0, 1, 2, \ldots)$. To derive a formula allowing us to carry out the analytic continuation in the latter case, we replace the hypergeometric functions in the right-hand side of (9.5.7) by the corresponding series, and use (1.2.2) to transform the result, obtaining

$$\begin{aligned}
&\frac{1}{\Gamma(\gamma)} F(\alpha, \beta; \gamma; z) \\
&= \frac{\pi}{\sin \pi(\gamma - \alpha - \beta)} \left[\frac{1}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)} \sum_{k=0}^{\infty} \frac{(\alpha)_k(\beta)_k}{\Gamma(1 + \alpha + \beta - \gamma + k)k!} (1 - z)^k \right. \\
&\qquad \left. - \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \sum_{k=0}^{\infty} \frac{(\gamma - \alpha)_k(\gamma - \beta)_k}{\Gamma(1 - \alpha - \beta + \gamma + k)} \frac{(1 - z)^{k + \gamma - \alpha - \beta}}{k!}\right] \\
&= \frac{\pi}{\sin \pi(\gamma - \alpha - \beta)} (g_1 - g_2). \qquad (9.7.1)
\end{aligned}$$

It is easily verified that

$$\lim_{\gamma \to \alpha + \beta + n} g_1 = \lim_{\gamma \to \alpha + \beta + n} g_2 = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \sum_{k=0}^{\infty} \frac{(\alpha + n)_k(\beta + n)_k}{(n + k)!k!} (1 - z)^{k+n},$$

and hence the right-hand side of (9.7.1) becomes indeterminate for $\gamma = \alpha + \beta + n$. Using L'Hospital's rule to eliminate this indeterminacy, we have

$$\frac{1}{\Gamma(\alpha + \beta + n)} F(\alpha, \beta; \alpha + \beta + n; z) = (-1)^n \left[\left.\frac{\partial g_1}{\partial \gamma}\right|_{\gamma = \alpha + \beta + n} - \left.\frac{\partial g_2}{\partial \gamma}\right|_{\gamma = \alpha + \beta + n}\right]. \qquad (9.7.2)$$

After some calculations resembling those made in Sec. 5.5, we find that[24]

$$\begin{aligned}
\left.\frac{\partial g_1}{\partial \gamma}\right|_{\gamma = \alpha + \beta + n} &= \frac{1}{\Gamma(\alpha + n)\Gamma(\beta + n)} \sum_{k=0}^{n-1} \frac{(-1)^{n-k}(n - k - 1)!(\alpha)_k(\beta)_k}{k!} (1 - z)^k \\
&\quad + \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \sum_{k=0}^{\infty} \frac{(\alpha + n)_k(\beta + n)_k}{(n + k)!k!} \\
&\quad \times [\psi(k + 1) - \psi(\alpha + n) - \psi(\beta + n)](1 - z)^{k+n}, \qquad (9.7.3)
\end{aligned}$$

---

[24] In differentiating g_2, we use the formula

$$\frac{d}{d\lambda} (\lambda)_k = (\lambda)_k[\psi(\lambda + k) - \psi(\lambda)].$$

From now on, we assume that $\alpha, \beta \neq 0, -1, -2, \ldots$

$$\frac{\partial g_2}{\partial \gamma}\bigg|_{\gamma=\alpha+\beta+n} = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \sum_{k=0}^{\infty} \frac{(\alpha+n)_k(\beta+n)_k}{(n+k)!k!}$$
$$\times [\psi(\alpha+n+k) - \psi(\alpha+n) + \psi(\beta+n+k)$$
$$- \psi(\beta+n) - \psi(1+n+k) + \log(1-z)](1-z)^{k+n}, \quad (9.7.4)$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the logarithmic derivative of the gamma function. Substituting (9.7.3–4) into (9.7.2), we obtain

$$F(\alpha, \beta; \alpha+\beta+n; z)$$
$$= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+n)\Gamma(\beta+n)} \sum_{k=0}^{n-1} \frac{(-1)^k(n-k-1)!(\alpha)_k(\beta)_k}{k!}(1-z)^k$$
$$+ \frac{(-1)^n\Gamma(\alpha+\beta+n)}{\Gamma(\alpha)\Gamma(\beta)} \sum_{k=0}^{\infty} \frac{(\alpha+n)_k(\beta+n)_k}{(n+k)!k!} [\psi(k+1) + \psi(n+k+1)$$
$$- \psi(\alpha+n+k) - \psi(\beta+n+k) - \log(1-z)](1-z)^{n+k},$$
$$|z-1| < 1, \ |\arg(1-z)| < \pi, \quad n = 0, 1, 2, \dots, \quad \alpha, \beta \neq 0, -1, -2, \dots \quad (9.7.5)$$

As usual, the meaningless sum

$$\sum_{k=0}^{-1} \cdots,$$

which appears when $n = 0$, is set equal to zero.

Formula (9.7.5) is no longer applicable if α or β equals $0, -1, -2, \dots$, but then $F(\alpha, \beta; \alpha+\beta+n; z)$ reduces to a polynomial, and there is no need for analytic continuation. Moreover, the case $\gamma = \alpha + \beta - n$ reduces to that just considered by using the transformation (9.5.3), which becomes

$$F(\alpha, \beta; \alpha+\beta-n; z) = (1-z)^{-n}F(\alpha', \beta'; \alpha'+\beta'+n; z) \quad (9.7.6)$$

if $\alpha' = \alpha - n$, $\beta' = \beta - n$.

Similar considerations apply to the other formulas of Secs. 9.5–6. To give another example, we derive a formula suitable for making the analytic continuation of $F(\alpha, \beta; \gamma; z)$ into the domain $|z| > 1$, $|\arg(-z)| < \pi$ in the case where $\alpha - \beta = 0, \pm 1, \pm 2, \dots$ Here we have to pass to the limit $\beta \to \alpha \pm n$ $(n = 0, 1, 2, \dots)$ in (9.5.9). A calculation like that given above leads to the following formula (for the case $\beta = \alpha + n$):[25]

$$F(\alpha, \alpha+n; \gamma; z)$$
$$= \frac{\Gamma(\gamma)(-z)^{-\alpha}}{\Gamma(\gamma-\alpha)\Gamma(\alpha+n)} \sum_{k=0}^{n-1} \frac{(n-k-1)!(\alpha)_k(1-\gamma+\alpha)_k}{k!}(-z)^{-k}$$
$$+ \frac{\Gamma(\gamma)(-z)^{-\alpha}}{\Gamma(\alpha)\Gamma(\gamma-\alpha-n)} \sum_{k=0}^{\infty} \frac{(\alpha+n)_k(1+\alpha-\gamma+n)_k}{(n+k)!k!}$$
$$\times [\psi(k+1) + \psi(n+k+1) - \psi(\alpha+n+k)$$
$$- \psi(\gamma-\alpha-n-k) + \log(-z)]z^{-n-k},$$
$$|z| > 1, \ |\arg(1-z)| < \pi, \quad n = 0, 1, 2, \dots, \quad \alpha \neq 0, -1, -2, \dots,$$
$$\gamma - \alpha \neq 0, \pm 1, \pm 2, \dots, \quad \gamma \neq 0, -1, -2, \dots \quad (9.7.7)$$

[25] In the last step of the calculation, use formula (1.3.4).

We now examine the cases where formula (9.7.7) is not applicable. If $\alpha = 0, -1, -2, \ldots$, the function $F(\alpha, \alpha + n; \gamma; z)$ reduces to a polynomial, and there is no need for analytic continuation. According to (9.5.3),

$$F(\alpha, \alpha + n; \gamma; z) = (1 - z)^{\gamma - 2\alpha - n} F(\gamma - \alpha, \gamma - \alpha - n; \gamma; z), \quad (9.7.8)$$

and therefore $F(\alpha, \alpha + n; \gamma; z)$ reduces to an algebraic function if $\gamma - \alpha = 0, -1, -2, \ldots$ or $\gamma - \alpha = 1, 2, \ldots, n$, and analytic continuation is again unnecessary. If $\gamma - \alpha = n + 1, n + 2, \ldots$ and $\alpha \neq 0, \pm 1, \pm 2, \ldots$, then the hypergeometric function in the right-hand side of (9.7.8) satisfies the conditions allowing it to be continued by using formula (9.7.7). If $\gamma - \alpha = n + 1, n + 2, \ldots$ and $\alpha = 1, 2, \ldots$, the hypergeometric function can be represented by an integral of the type (9.1.6) with a rational integrand, i.e., $F(\alpha, \alpha + n; \gamma; z)$ can be expressed in finite form in terms of rational functions. Finally, we note that the case $\beta = \alpha - n$ reduces to that just considered if we again use the transformation (9.5.3).

9.8. Representation of Various Functions in Terms of the Hypergeometric Function

As we now show, various familiar functions of mathematical analysis are special cases of the hypergeometric function $F(\alpha, \beta; \gamma; z)$, corresponding to suitable choices of the parameters α, β, γ and the variable z:[26]

1. *Elementary functions.* The hypergeometric function $F(\alpha, \beta; \gamma; z)$ reduces to a polynomial if $\alpha = 0, -1, -2, \ldots$ or $\beta = 0, -1, -2, \ldots$ For example,

$$F(\alpha, 0; \gamma; z) = 1, \qquad F(\alpha, -2; \gamma; z) = 1 - 2\frac{\alpha}{\gamma} z + \frac{\alpha(\alpha + 1)}{\gamma(\gamma + 1)} z^2,$$

and so on. The transformation

$$F(\alpha, \beta; \gamma; z) = (1 - z)^{\gamma - \alpha - \beta} F(\gamma - \alpha, \gamma - \beta; \gamma; z), \quad |\arg(1 - z)| < \pi$$

[cf. (9.5.3)] shows that $F(\alpha, \beta; \gamma; z)$ reduces to an algebraic function if $\gamma - \alpha = 0, -1, -2, \ldots$ or $\gamma - \beta = 0, -1, -2, \ldots$ In particular,

$$F(\alpha, \beta; \beta; z) = (1 - z)^{-\alpha}, \qquad |\arg(1 - z)| < \pi \quad (9.8.1)$$

for any value of β, and

$$\begin{aligned} (1 - z)^{\nu} = F(-\nu, 1; 1; z), \qquad (1 - z)^{-1/2} = F(\tfrac{1}{2}, 1; 1; z), \\ z^n = F(-n, 1; 1; 1 - z), \qquad n = 0, 1, 2, \ldots \end{aligned} \quad (9.8.2)$$

---

[26] Further examples are given in the Bateman Manuscript Project, *Higher Transcendental Functions, Vol. 1*, pp. 89, 101.

Other representations of this type can be derived from the formulas of Sec. 9.6. Thus, setting $\beta = \alpha + \frac{1}{2}$ in (9.6.2) and (9.6.7), we obtain

$$F(\alpha, \alpha + \tfrac{1}{2}; 2\alpha + 1; z) = \left(\frac{1 + \sqrt{1 - z}}{2}\right)^{-2\alpha}, \qquad |\arg(1 - z)| < \pi,$$

$$F(\alpha, \alpha + \tfrac{1}{2}; 2\alpha; z) = \frac{1}{\sqrt{1 - z}}\left(\frac{1 + \sqrt{1 - z}}{2}\right)^{1-2\alpha}, \ |\arg(1 - z)| < \pi. \tag{9.8.3}$$

By starting from the series expansion

$$\log(1 - z) = -\sum_{k=0}^{\infty} \frac{z^{k+1}}{k + 1} = -z \sum_{k=0}^{\infty} \frac{(1)_k(1)_k}{(2)_k k!} z^k, \qquad |z| < 1$$

of the logarithm, we find that

$$\log(1 - z) = -zF(1, 1; 2; z), \qquad |\arg(1 - z)| < \pi. \tag{9.8.4}$$

Similarly, we deduce the following formulas for the inverse trigonometric functions:

$$\begin{aligned} \arctan z &= zF(\tfrac{1}{2}, 1; \tfrac{3}{2}; -z^2), \qquad |\arg(1 \pm zi)| < \pi, \\ \arcsin z &= zF(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{3}{2}; z^2), \qquad |\arg(1 \pm z)| < \pi. \end{aligned} \tag{9.8.5}$$

2. *Elliptic integrals*. The complete elliptic integrals

$$K(z) = \int_0^{\pi/2} (1 - z^2 \sin^2 \varphi)^{-1/2} \, d\varphi, \quad E(z) = \int_0^{\pi/2} (1 - z^2 \sin^2 \varphi)^{1/2} \, d\varphi$$

of the first and second kinds [cf. (7.10.11)], where z is a complex variable belonging to the domain $|\arg(1 \pm z)| < \pi$, can also be represented in terms of the hypergeometric function. Assuming temporarily that $|z| < 1$ and using the binomial expansion, we find that

$$K(z) = \sum_{k=0}^{\infty} \frac{(\frac{1}{2})_k}{k!} z^{2k} \int_0^{\pi/2} \sin^{2k} \varphi \, d\varphi = \frac{\pi}{2} \sum_{k=0}^{\infty} \frac{(\frac{1}{2})_k(\frac{1}{2})_k}{(1)_k k!} z^{2k},$$

which implies

$$K(z) = \frac{\pi}{2} F(\tfrac{1}{2}, \tfrac{1}{2}; 1; z^2), \qquad |\arg(1 \pm z)| < \pi. \tag{9.8.6}$$

Similarly, we have the following representation of the elliptic integral of the second kind:

$$E(z) = \frac{\pi}{2} F(-\tfrac{1}{2}, \tfrac{1}{2}; 1; z^2), \qquad |\arg(1 \pm z)| < \pi. \tag{9.8.7}$$

Starting from these formulas, one can develop the theory of elliptic integrals, regarded as functions of the modulus z.

3. *Spherical harmonics.* One of the most important classes of functions which can be expressed in terms of the hypergeometric function consists of the spherical harmonics studied in Chapter 7. In fact, formulas (7.12.27) and (7.12.29) immediately imply the following representations of the associated Legendre functions:

$$P_\nu^m(z) = \frac{\Gamma(\nu + m + 1)}{\Gamma(\nu - m + 1)} \frac{(z^2 - 1)^{m/2}}{2^m\Gamma(m + 1)} \times F\left(m - \nu, m + \nu + 1; m + 1; \frac{1 - z}{2}\right),$$
$$|\arg (z \pm 1)| < \pi, \qquad m = 0, 1, 2, \ldots, \tag{9.8.8}$$

$$Q_\nu^m(z) = \frac{(-1)^m\sqrt{\pi}\,\Gamma(\nu + m + 1)}{2^{\nu+1}\Gamma(\nu + \frac{3}{2})z^{\nu+m+1}} (z^2 - 1)^{m/2} \times F\left(\frac{m + \nu + 2}{2}, \frac{m + \nu + 1}{2}; \nu + \frac{3}{2}; \frac{1}{z^2}\right),$$
$$|\arg z| < \pi, \qquad |\arg (z \pm 1)| < \pi, \qquad m = 0, 1, 2, \ldots \tag{9.8.9}$$

In particular, the Legendre polynomials (see Sec. 4.2) are given by the formula

$$P_n(z) = F\left(-n, n + 1; 1; \frac{1 - z}{2}\right), \qquad n = 0, 1, 2, \ldots \tag{9.8.10}$$

By regarding (9.8.8–10) as definitions and using the general theory of the hypergeometric function, it is a simple matter to develop the theory of spherical harmonics. This approach is especially convenient for deriving the relations of Sec. 7.6 and their generalizations to the case of arbitrary m.

9.9 The Confluent Hypergeometric Function

Besides the hypergeometric function $F(\alpha, \beta; \gamma; z)$, an important role is played in the theory of special functions by a related function

$$\Phi(\alpha, \gamma; z) = \sum_{k=0}^{\infty} \frac{(\alpha)_k}{(\gamma)_k} \frac{z^k}{k!}, \qquad |z| < \infty, \quad \gamma \neq 0, -1, -2, \ldots, \tag{9.9.1}$$

known as the *confluent hypergeometric function.* Here z is a complex variable, α and γ are parameters which can take arbitrary real or complex values (except that $\gamma \neq 0, -1, -2, \ldots$), and, as always,

$$(\lambda)_0 = 1, \qquad (\lambda)_k = \frac{\Gamma(\lambda + k)}{\Gamma(\lambda)} = \lambda(\lambda + 1)\cdots(\lambda + k - 1), \qquad k = 1, 2, \ldots$$

As indicated, the series (9.9.1) converges for all finite z,[27] and therefore represents an entire function of z.

If we set

$$\varphi(\alpha, \gamma; z) = \frac{1}{\Gamma(\gamma)} \Phi(\alpha, \gamma; z) = \sum_{k=0}^{\infty} \frac{(\alpha)_k}{\Gamma(\gamma + k)} \frac{z^k}{k!}, \tag{9.9.2}$$

then $\varphi(\alpha, \gamma; z)$ is an entire function of α and γ, for fixed z. In fact, the terms of the series (9.9.2) are entire functions of α and γ, and the series is uniformly convergent in the region $|\alpha| \leqslant A$, $|\gamma| \leqslant C$ (where A and C are arbitrarily large).[28] Therefore, for fixed z, $\Phi(\alpha, \gamma; z)$ is an entire function of α and a meromorphic function of γ, with simple poles at the points $\gamma = 0, -1, -2, \ldots$

A comparison of (9.1.2) and (9.1.3) shows at once that

$$\Phi(\alpha, \gamma; z) = \lim_{\beta \to \infty} F\left(\alpha, \beta; \gamma; \frac{z}{\beta}\right). \tag{9.9.3}$$

The function $\Phi(\alpha, \gamma; z)$ is very frequently encountered in analysis, mainly because of the fact that a large number of special functions can be obtained from $\Phi(\alpha, \gamma; z)$ by making suitable choices of the parameters α, γ and the variable z (see Sec. 9.13). This makes it possible to develop the general theory of these functions in a simple and compact form.

The definition of the confluent hypergeometric function immediately implies the identities

$$\frac{d}{dz} \Phi(\alpha, \gamma; z) = \frac{\alpha}{\gamma} \Phi(\alpha + 1, \gamma + 1; z), \tag{9.9.4}$$

$$\frac{d^m}{dz^m} \Phi(\alpha, \gamma; z) = \frac{(\alpha)_m}{(\gamma)_m} \Phi(\alpha + m, \gamma + m; z), \quad m = 1, 2, \ldots, \tag{9.9.5}$$

---

[27] Use the ratio test, noting that if

$$u_k = \frac{(\alpha)_k}{(\gamma)_k} \frac{z^k}{k!},$$

then

$$\left|\frac{u_{k+1}}{u_k}\right| = \left|\frac{\alpha + k}{(\gamma + k)(1 + k)} z\right| \to 0,$$

as $k \to \infty$.

[28] Use the criterion given in footnote 4, p. 102, noting that if

$$v_k = \frac{(\alpha)_k}{\Gamma(\gamma + k)} \frac{z^k}{k!},$$

then

$$\left|\frac{v_{k+1}}{v_k}\right| = \left|\frac{\alpha + k}{(\gamma + k)(1 + k)} z\right| \leqslant \frac{A + k}{(k - C)(1 + k)} |z| \leqslant q < 1,$$

for sufficiently large k.

and the recurrence relations

$$(\gamma - \alpha - 1)\Phi + \alpha\Phi(\alpha + 1) - (\gamma - 1)\Phi(\gamma - 1) = 0, \tag{9.9.6}$$

$$\gamma\Phi - \gamma\Phi(\alpha - 1) - z\Phi(\gamma + 1) = 0, \tag{9.9.7}$$

$$(\alpha - 1 + z)\Phi + (\gamma - \alpha)\Phi(\alpha - 1) - (\gamma - 1)\Phi(\gamma - 1) = 0, \tag{9.9.8}$$

$$\gamma(\alpha + z)\Phi - \alpha\gamma\Phi(\alpha + 1) - (\gamma - \alpha)z\Phi(\gamma + 1) = 0, \tag{9.9.9}$$

$$(\gamma - \alpha)\Phi(\alpha - 1) + (2\alpha - \gamma + z)\Phi - \alpha\Phi(\alpha + 1) = 0, \tag{9.9.10}$$

$$\gamma(\gamma - 1)\Phi(\gamma - 1) - \gamma(\gamma - 1 + z)\Phi + (\gamma - \alpha)z\Phi(\gamma + 1) = 0, \tag{9.9.11}$$

connecting the function $\Phi \equiv \Phi(\alpha, \gamma; z)$ with any two *contiguous* functions $\Phi(\alpha \pm 1) \equiv \Phi(\alpha \pm 1, \gamma; z)$ and $\Phi(\gamma \pm 1) \equiv \Phi(\alpha, \gamma \pm 1; z)$. Formulas (9.9.6–7) can be verified by direct substitution of the series (9.9.1), and then the other recurrence relations can be obtained by simple transformations of (9.9.6–7).

Besides the recurrence relations just given, there exist similar relations between the function $\Phi(\alpha, \gamma; z)$ and any pair of functions of the form $\Phi(\alpha + m, \gamma + n; z)$, where m and n are arbitrary integers. Two simple relations of this kind are[29]

$$\Phi(\alpha, \gamma; z) = \Phi(\alpha + 1, \gamma; z) - \frac{z}{\gamma}\Phi(\alpha + 1, \gamma + 1; z), \tag{9.9.12}$$

$$\Phi(\alpha, \gamma; z) = \frac{\gamma - \alpha}{\gamma}\Phi(\alpha, \gamma + 1; z) + \frac{\alpha}{\gamma}\Phi(\alpha + 1, \gamma + 1; z), \tag{9.9.13}$$

as can be verified by direct substitution of (9.9.1), or by repeated use of the relations between $\Phi(\alpha, \gamma; z)$ and its contiguous functions.

9.10. The Differential Equation for the Confluent Hypergeometric Function and Its Solutions. The Confluent Hypergeometric Function of the Second Kind

It is easy to see that the confluent hypergeometric function is a particular solution of the linear differential equation

$$zu'' + (\gamma - z)u' - \alpha u = 0, \tag{9.10.1}$$

where $\gamma \neq 0, -1, -2, \ldots$ In fact, denoting the left-hand side of this equation by $l(u)$, and setting $u = u_1 = \Phi(\alpha, \gamma; z)$, we have

$$\begin{aligned} l(u_1) &= \sum_{k=2}^{\infty} \frac{k(k-1)(\alpha)_k}{(\gamma)_k k!} z^{k-1} + (\gamma - z)\sum_{k=1}^{\infty} \frac{(\alpha)_k k}{(\gamma)_k k!} z^{k-1} - \alpha \sum_{k=0}^{\infty} \frac{(\alpha)_k}{(\gamma)_k k!} z^k \\ &= \left[\gamma\frac{(\alpha)_1}{(\gamma)_1} - \alpha\right] + \sum_{k=1}^{\infty} \frac{(\alpha)_k z^k}{(\gamma)_k k!}\left[k\frac{\alpha + k}{\gamma + k} + \gamma\frac{\alpha + k}{\gamma + k} - k - \alpha\right] \equiv 0. \end{aligned}$$

---

[29] Note the similarity between formulas (9.9.6–13) and formulas (9.2.4–15).

To obtain a second linearly independent solution of (9.10.1), we assume that $|\arg z| < \pi$ and make the substitution $u = z^{1-\gamma}v$. Then equation (9.10.1) goes into an equation of the same form, i.e.,

$$zv'' + (\gamma' - z)v' - \alpha' v = 0,$$

with new parameters $\alpha' = 1 + \alpha - \gamma$, $\gamma' = 2 - \gamma$. It follows that the function

$$u = u_2 = z^{1-\gamma}\Phi(1 + \alpha - \gamma, 2 - \gamma; z)$$

is also a solution of (9.10.1) if $\gamma \neq 2, 3, \ldots$. Thus, if $\gamma \neq 0, \pm 1, \pm 2, \ldots$, both solutions u_1, u_2 are meaningful and are linearly independent of each other,[30] so that the general solution of (9.10.1) can be written in the form

$$u = A\Phi(\alpha, \gamma; z) + Bz^{1-\gamma}\Phi(1 + \alpha - \gamma; 2 - \gamma; z),$$
$$|\arg z| < \pi, \quad \gamma \neq 0, \pm 1, \pm 2, \ldots \tag{9.10.2}$$

With a view to obtaining an expression for the general solution of (9.10.1) which is suitable for arbitrary $\gamma \neq 0, -1, -2, \ldots$ [see (9.10.11) below], we introduce a new function

$$\Psi(\alpha, \gamma; z) = \frac{\Gamma(1 - \gamma)}{\Gamma(1 + \alpha - \gamma)}\,\Phi(\alpha, \gamma; z) + \frac{\Gamma(\gamma - 1)}{\Gamma(\alpha)}\,z^{1-\gamma}\Phi(1 + \alpha - \gamma, 2 - \gamma; z),$$
$$|\arg z| < \pi, \quad \gamma \neq 0, \pm 1, \pm 2, \ldots, \tag{9.10.3}$$

called the *confluent hypergeometric function of the second kind.* Formula (9.10.3) defines the function $\Psi(\alpha, \gamma; z)$ for arbitrary nonintegral γ, and moreover, as we now show, the right-hand side of (9.10.3) approaches a definite limit as $\gamma \to n + 1$ $(n = 0, 1, 2, \ldots)$. Replacing the Φ functions in (9.10.3) by the appropriate series, and using formula (1.2.2) from the theory of the gamma function, we obtain

$$\begin{aligned}\Psi(\alpha, \gamma; z) = \frac{\pi}{\sin \pi\gamma}\Bigg[&\frac{1}{\Gamma(1 + \alpha - \gamma)}\sum_{k=0}^{\infty}\frac{(\alpha)_k}{\Gamma(\gamma + k)}\frac{z^k}{k!} \\ &- \frac{1}{\Gamma(\alpha)}\sum_{k=0}^{\infty}\frac{(4\alpha - \gamma)}{\Gamma(2 - \gamma + k)}\frac{z^{k+1-\gamma}}{k!}\Bigg] = \frac{\pi}{\sin \pi\gamma}(g_1 - g_2).\end{aligned} \tag{9.10.4}$$

Since

$$\lim_{\gamma\to n+1} g_1 = \frac{1}{\Gamma(\alpha - n)}\sum_{k=0}^{\infty}\frac{(\alpha)_k}{\Gamma(k + n + 1)}\frac{z^k}{k!} = \frac{1}{\Gamma(\alpha - n)}\sum_{k=0}^{\infty}\frac{(\alpha)_k}{(n + k)!}\frac{z^k}{k!},$$

$$\begin{aligned}\lim_{\gamma\to n+1} g_2 &= \frac{1}{\Gamma(\alpha)}\sum_{k=n}^{\infty}\frac{(\alpha - n)_k}{\Gamma(k - n + 1)}\frac{z^{k-n}}{k!} \\ &= \frac{1}{\Gamma(\alpha)}\sum_{k=0}^{\infty}\frac{(\alpha - n)_{k+n}}{\Gamma(k + 1)}\frac{z^k}{(n + k)!} = \frac{1}{\Gamma(\alpha - n)}\sum_{k=0}^{\infty}\frac{(\alpha)_k}{(n + k)!}\frac{z^k}{k!},\end{aligned}$$

---

[30] Note that $u_1 \equiv u_2$ if $\gamma = 1$.

the right-hand side of (9.10.4) becomes indeterminate as $\gamma \to n + 1$, and approaches a limit whose value can be found by using L'Hospital's rule, i.e.,

$$\Psi(\alpha, n+1; z) = \lim_{\gamma \to n+1} \Psi(\alpha, \gamma; z) = (-1)^{n+1}\left[\frac{\partial g_1}{\partial \gamma}\bigg|_{\gamma=n+1} - \frac{\partial g_2}{\partial \gamma}\bigg|_{\gamma=n+1}\right],$$
$$|\arg z| < \pi, \quad n = 0, 1, 2, \ldots \tag{9.10.5}$$

Calculations like those made in Sec. 5.5 show that[31]

$$\frac{\partial g_1}{\partial \gamma}\bigg|_{\gamma=n+1} = \frac{1}{\Gamma(\alpha - n)} \sum_{k=0}^{\infty} \frac{(\alpha)_k z^k}{(n+k)!k!} [\psi(\alpha - n) - \psi(n + k + 1)],$$

$$\frac{\partial g_2}{\partial \gamma}\bigg|_{\gamma=n+1} = \frac{1}{\Gamma(\alpha - n)} \sum_{k=0}^{\infty} \frac{(\alpha)_k z^k}{(n+k)!k!}$$
$$\times [\psi(1 + k) - \psi(\alpha + k) + \psi(\alpha - n) - \log z]$$
$$+ \frac{1}{\Gamma(\alpha)} \sum_{k=0}^{n-1} \frac{(-1)^{n-k}(n-k-1)!(\alpha - n)_k}{k!} z^{k-n},$$

which leads to the following series expansion:

$$\Psi(\alpha, n+1; z)$$
$$= \frac{(-1)^{n+1}}{\Gamma(\alpha - n)} \sum_{k=0}^{\infty} \frac{(\alpha)_k z^k}{(n+k)!k!} [\psi(\alpha + k) - \psi(1 + k) - \psi(n + 1 + k) + \log z]$$
$$+ \frac{1}{\Gamma(\alpha)} \sum_{k=0}^{n-1} \frac{(-1)^k(n-k-1)!(\alpha - n)_k}{k!} z^{k-n}, \tag{9.10.6}$$
$$|\arg z| < \pi, \quad n = 0, 1, 2, \ldots, \quad \alpha \neq 0, -1, -2, \ldots$$

Here $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the logarithmic derivative of the gamma function, and the meaningless sum

$$\sum_{k=0}^{-1} \cdots,$$

which appears when $n = 0$, is set equal to zero.

If $\alpha = -m$ ($m = 0, 1, 2, \ldots$), passage to the limit $\gamma \to n + 1$ ($n = 0, 1, 2, \ldots$) in (9.10.3) leads to the expression[32]

$$\Psi(-m, n+1; z) = (-1)^m \frac{(m+n)!}{n!} \Phi(-m, n+1; z), \tag{9.10.7}$$
$$m = 0, 1, 2, \ldots, \quad n = 0, 1, 2, \ldots$$

---

[31] In differentiating g_2, we use the formula

$$\frac{d}{d\lambda}(\lambda)_k = (\lambda)_k[\psi(\lambda + k) - \psi(\lambda)].$$

From now on, we assume that $\alpha \neq 0, -1, -2, \ldots$

[32] Here we again use formula (1.2.2).

Moreover, it is an immediate consequence of (9.10.3) that the confluent hypergeometric function of the second kind satisfies the relation

$$\Psi(\alpha, \gamma; z) = z^{1-\gamma}\Psi(1 + \alpha - \gamma, 2 - \gamma; z), \qquad |\arg z| < \pi. \quad (9.10.8)$$

Using this formula, we can define the function $\Psi(\alpha, \gamma; z)$ for $\gamma = 0, -1, -2, \ldots$, obtaining

$$\Psi(\alpha, 1 - n; z) = \lim_{\gamma \to 1-n} \Psi(\alpha, \gamma; z) = z^n \Psi(\alpha + n, n + 1; z), \qquad |\arg z| < \pi, \quad n = 1, 2, \ldots \quad (9.10.9)$$

Thus we see that $\Psi(\alpha, \gamma; z)$ is meaningful for arbitrary values of the parameters α and γ. It follows from the definition (9.10.3) and the properties of $\Phi(\alpha, \gamma; z)$ that $\Psi(\alpha, \gamma; z)$ is an analytic function of z in the plane cut along $[-\infty, 0]$, and an entire function of α and γ.

Next we show that $\Psi(\alpha, \gamma; z)$ is a solution of the differential equation (9.10.1). For $\gamma \neq 0, \pm 1, \pm 2, \ldots$, this is an immediate consequence of (9.10.3), and for integral γ, the result follows from the principle of analytic continuation (cf. footnote 12, p. 167). For $\alpha \neq 0, -1, -2, \ldots$, the solutions $\Phi(\alpha, \gamma; z)$ and $\Psi(\alpha, \gamma; z)$ are linearly independent, as can easily be verified by calculating the Wronskian[33]

$$W\{\Phi(\alpha, \gamma, z), \Psi(\alpha, \gamma; z)\} = -\frac{\Gamma(\gamma)}{\Gamma(\alpha)} z^{-\gamma} e^z, \qquad |\arg z| < \pi, \quad \gamma \neq 0, -1, -2, \ldots, \quad (9.10.10)$$

and then the general solution of (9.10.1) can be written in the form

$$u = A\Phi(\alpha, \gamma; z) + B\Psi(\alpha, \gamma; z), \qquad |\arg z| < \pi, \quad \alpha, \gamma \neq 0, -1, -2, \ldots \quad (9.10.11)$$

The function $\Psi(\alpha, \gamma; z)$ has a number of properties analogous to those of $\Phi(\alpha, \gamma; z)$. For example, we have the differentiation formulas

$$\begin{aligned} \frac{d}{dz}\Psi(\alpha, \gamma; z) &= -\alpha\Psi(\alpha + 1, \gamma + 1; z), \\ \frac{d^m}{dz^m}\Psi(\alpha, \gamma; z) &= (-1)^m(\alpha)_m\Psi(\alpha + m, \gamma + m; z), \qquad m = 1, 2, \ldots, \end{aligned} \quad (9.10.12)$$

the recurrence relations

$$\Psi - \alpha\Psi(\alpha + 1) - \Psi(\gamma - 1) = 0, \quad (9.10.13)$$

$$(\gamma - \alpha)\Psi + \Psi(\alpha - 1) - z\Psi(\gamma + 1) = 0, \quad (9.10.14)$$

[33] Equation (9.10.1) implies

$$W\{\Phi, \Psi\} = Cz^{-\gamma}e^z.$$

Comparing both sides of this identity as $z \to 0$, we find that

$$C = -\frac{\Gamma(\gamma)}{\Gamma(\alpha)}.$$

$$(\alpha - 1 + z)\Psi - \Psi(\alpha - 1) + (\alpha - \gamma + 1)\Psi(\gamma - 1) = 0, \tag{9.10.15}$$

$$(\alpha + z)\Psi + \alpha(\gamma - \alpha - 1)\Psi(\alpha + 1) - z\Psi(\gamma + 1) = 0, \tag{9.10.16}$$

$$\Psi(\alpha - 1) - (2\alpha - \gamma + z)\Psi + \alpha(\alpha - \gamma + 1)\Psi(\alpha + 1) = 0, \tag{9.10.17}$$

$$(\gamma - \alpha - 1)\Psi(\gamma - 1) - (\gamma - 1 + z)\Psi + z\Psi(\gamma + 1) = 0, \tag{9.10.18}$$

$$\Psi \equiv \Psi(\alpha, \gamma; z), \quad \Psi(\alpha \pm 1) \equiv \Psi(\alpha \pm 1, \gamma; z), \quad \Psi(\gamma \pm 1) \equiv \Psi(\alpha, \gamma \pm 1; z)$$

and so on, whose validity follows from the definition of the Ψ function and the corresponding properties of the Φ function.

9.11. Integral Representations of the Confluent Hypergeometric Functions

The functions $\Phi(\alpha, \gamma; z)$ and $\Psi(\alpha, \gamma; z)$ have simple integral representations which play an important role in the theory and applications of confluent hypergeometric functions. Here we consider only the basic representations in terms of integrals evaluated along an interval of the real axis, referring the reader elsewhere for more general representations in terms of contour integrals.[34]

The simplest integral representation of the function $\Phi(\alpha, \gamma; z)$ can be obtained by summing the series (9.9.1) with the help of formula (9.1.2):

$$\frac{(\alpha)_k}{(\gamma)_k} = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma - \alpha)} \int_0^1 t^{\alpha - 1 + k}(1 - t)^{\gamma - \alpha - 1}\, dt,$$

$$\operatorname{Re}\gamma > \operatorname{Re}\alpha > 0, \quad k = 0, 1, 2, \ldots$$

This gives

$$\Phi(\alpha, \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma - \alpha)} \sum_{k=0}^{\infty} \frac{z^k}{k!} \int_0^1 t^{\alpha - 1 + k}(1 - t)^{\gamma - \alpha - 1}\, dt$$

$$= \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma - \alpha)} \int_0^1 t^{\alpha - 1}(1 - t)^{\gamma - \alpha - 1} dt \sum_{k=0}^{\infty} \frac{(zt)^k}{k!},$$

or

$$\Phi(\alpha, \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma - \alpha)} \int_0^1 e^{zt} t^{\alpha - 1}(1 - t)^{\gamma - \alpha - 1}\, dt, \qquad \operatorname{Re}\gamma > \operatorname{Re}\alpha > 0, \tag{9.11.1}$$

where reversing the order of integration and summation is justified by the usual absolute convergence argument (cf. footnote 2, p. 239).

---

[34] See the Bateman Manuscript Project, *Higher Transcendental Functions, Vol. 1*, pp. 256, 271 ff.

We can use the integral representation (9.11.1) to deduce an important relation satisfied by the function $\Phi(\alpha, \gamma; z)$. Assuming temporarily that $\operatorname{Re}\gamma > \operatorname{Re}\alpha > 0$, we make the change of variable $t = 1 - s$. Then (9.11.1) becomes

$$\Phi(\alpha, \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma - \alpha)} e^z \int_0^1 e^{-zs} s^{\gamma-\alpha-1}(1 - s)^{\alpha-1}\, ds,$$

which implies

$$\Phi(\alpha, \gamma; z) = e^z \Phi(\gamma - \alpha, \gamma; z), \tag{9.11.2}$$

since $\operatorname{Re}\gamma > \operatorname{Re}(\gamma - \alpha)$. The relation (9.11.2) was proved under the assumption that $\operatorname{Re}\gamma > \operatorname{Re}\alpha > 0$, but after dividing by $\Gamma(\gamma)$, both sides become entire functions of α and γ. Therefore, according to the principle of analytic continuation, (9.11.2) remains valid for arbitrary α and γ, provided that $\gamma \neq 0, -1, -2, \ldots$

To obtain an integral representation of $\Psi(\alpha, \gamma; z)$, we first note that the function u, defined by

$$u = \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-zt} t^{\alpha-1}(1 + t)^{\gamma-\alpha-1}\, dt, \qquad \operatorname{Re}\alpha > 0, \quad \operatorname{Re} z > 0, \tag{9.11.3}$$

is a solution of the differential equation (9.10.1). In fact, denoting the left-hand side of (9.11.3) by $l(u)$, we have[35]

$$\begin{aligned} l(u) &= \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-zt} t^{\alpha-1}(1 + t)^{\gamma-\alpha-1}[zt^2 - (\gamma - z)t - \alpha]\, dt \\ &= -\frac{1}{\Gamma(\alpha)} \int_0^\infty \frac{d}{dt}[e^{-zt} t^\alpha (1 + t)^{\gamma-\alpha}]\, dt = -\frac{1}{\Gamma(\alpha)} e^{-zt} t^\alpha (1 + t)^{\gamma-\alpha}\Big|_{t=0}^{t=\infty} \equiv 0. \end{aligned}$$

According to (9.10.2), the solution u can be written in the form

$$\begin{gathered} u = A\Phi(\alpha, \gamma; z) + Bz^{1-\gamma}\Phi(1 + \alpha - \gamma, 2 - \gamma; z), \\ |\arg z| < \pi, \quad \gamma \neq 0, \pm 1, \pm 2, \ldots \end{gathered} \tag{9.11.4}$$

Assuming temporarily that $0 < \operatorname{Re}\gamma < 1$ and $z > 0$, we take the limit of (9.11.3) as $z \to 0+$. This gives

$$A = \lim_{z \to 0+} u = \frac{1}{\Gamma(\alpha)} \int_0^\infty t^{\alpha-1}(1 + t)^{\gamma-\alpha-1}\, dt = \frac{\Gamma(1 - \gamma)}{\Gamma(1 + \alpha - \gamma)},$$

where we have used formulas (1.5.3) and (1.5.6) from the theory of the gamma function, and the passage to the limit behind the integral sign is easily

[35] With our restrictions on α and z, the differentiation behind the integral sign is justified.

justified. Moreover, differentiating (9.11.4) with respect to z, multiplying by z^γ and then taking the limit as $z \to 0+$, we obtain

$$\begin{aligned} B &= \frac{1}{1-\gamma} \lim_{z\to 0+} z^\gamma u' = \frac{1}{\gamma - 1}\frac{1}{\Gamma(\alpha)} \lim_{z\to 0+} z^\gamma \int_0^\infty e^{-zt} t^\alpha (1+t)^{\gamma-\alpha-1}\,dt \\ &= \frac{1}{(\gamma-1)\Gamma(\alpha)} \lim_{z\to 0+} \int_0^\infty e^{-s} s^\alpha (s+z)^{\gamma-\alpha-1}\,ds \\ &= \frac{1}{(\gamma-1)\Gamma(\alpha)} \int_0^\infty e^{-s} s^{\gamma-1}\,ds = \frac{\Gamma(\gamma-1)}{\Gamma(\alpha)}. \end{aligned}$$

It follows that

$$\begin{aligned} u = &\frac{\Gamma(1-\gamma)}{\Gamma(1+\alpha-\gamma)} \Phi(\alpha, \gamma; z) \\ &+ \frac{\Gamma(\gamma-1)}{\Gamma(\alpha)} z^{1-\gamma} \Phi(1+\alpha-\gamma, 2-\gamma; z) \equiv \Psi(\alpha, \gamma; z). \end{aligned} \tag{9.11.5}$$

Since both sides are entire functions of the parameter γ and analytic functions of the variable z in the half-plane $\operatorname{Re} z > 0$ (see Sec. 9.10), the temporary restrictions imposed on γ and z can be dropped, and we arrive at the integral representation

$$\Psi(\alpha, \gamma; z) = \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-zt} t^{\alpha-1} (1+t)^{\gamma-\alpha-1}\,dt, \quad \operatorname{Re}\alpha > 0, \quad \operatorname{Re} z > 0. \tag{9.11.6}$$

Some other integral representations of the functions $\Phi(\alpha, \gamma; z)$ and $\Psi(\alpha, \gamma; z)$ are given in Problems 11–13, p. 278.

9.12. Asymptotic Representations of the Confluent Hypergeometric Functions for Large $|z|$

We begin by deriving the asymptotic representation of $\Psi(\alpha, \gamma; z)$ for large $|z|$, which turns out to be simpler than the corresponding representation of $\Phi(\alpha, \gamma; z)$. Suppose that

$$\operatorname{Re}\alpha > 0, \qquad |\arg z| < \frac{\pi}{2} - \delta,$$

where $\delta > 0$ is arbitrarily small. According to (5.11.2),

$$\begin{aligned} (1+t)^{\gamma-\alpha-1} = &\sum_{k=0}^{n} \frac{(-1)^k (1+\alpha-\gamma)_k}{k!} t^k \\ &+ \frac{(-1)^{n+1}(1+\alpha-\gamma)_n}{n!} t^{n+1} \int_0^1 (1-s)^n (1+st)^{\gamma-\alpha-n-2}\,ds. \end{aligned}$$

Substituting this expansion into the integral representation (9.11.6) and integrating term by term, we obtain[36]

$$\Psi(\alpha, \gamma; z) = z^{-\alpha}\left[\sum_{k=0}^{n} \frac{(-1)^k(\alpha)_k(1+\alpha-\gamma)_k}{k!} z^{-k} + r_n(z)\right],$$

where

$$r_n(z) = \frac{(-1)^{n+1}(1+\alpha-\gamma)_n z^\alpha}{n!\Gamma(\alpha)} \int_0^\infty e^{-zt} t^{n+\alpha}\, dt \int_0^1 (1-s)^n (1+st)^{\gamma-\alpha-n-2}\, ds.$$

Estimating $|r_n(z)|$ we find that

$$|r_n(z)| \leqslant \left|\frac{(1+\alpha-\gamma)_n}{n!\Gamma(\alpha)} z^\alpha\right| \int_0^\infty e^{-|z|t \sin \delta} t^{n+\operatorname{Re}\alpha}\, dt \times \int_0^1 (1-s)^n (1+st)^{\operatorname{Re}(\gamma-\alpha)-n-2}\, ds.$$

If we choose n so large that $\operatorname{Re}(\gamma-\alpha) - n - 2 \leqslant 0$, then

$$(1+st)^{\operatorname{Re}(\gamma-\alpha)-n-2} \leqslant 1,$$

and hence[37]

$$|r_n(z)| \leqslant \left|\frac{(1+\alpha-\gamma)_n}{(n+1)!\Gamma(\alpha)}\right| \frac{\Gamma(n+\operatorname{Re}\alpha+1)|z|^{\operatorname{Re}\alpha} e^{\pi|\operatorname{Im}\alpha|}}{(|z|\sin\delta)^{n+\operatorname{Re}\alpha+1}} = O(|z|^{-n-1}).$$

It follows that

$$\Psi(\alpha, \gamma; z) = z^{-\alpha}\left[\sum_{k=0}^{n} \frac{(-1)^k(\alpha)_k(1+\alpha-\gamma)_k}{k!} z^{-k} + O(|z|^{-n-1})\right],$$

$$\operatorname{Re}\alpha > 0, \quad |\arg z| \leqslant \frac{\pi}{2} - \delta, \quad n \geqslant \operatorname{Re}(\gamma-\alpha) - 2 \qquad (9.12.1)$$

for large $|z|$.

We now show that the conditions under which this formula has been proved can be considerably weakened. First we note that even if $\operatorname{Re}(\gamma-\alpha) - n - 2 > 0$, an integer $m > n$ can always be found such that $\operatorname{Re}(\gamma-\alpha) - m - 2 \leqslant 0$. Since the expansion (9.12.1) certainly holds with n replaced by m, we have

$$\sum_{k=0}^{m} \cdots + O(|z|^{-m-1}) = \sum_{k=0}^{n} \cdots + \sum_{k=n+1}^{m} \cdots + O(|z|^{-m-1})$$
$$= \sum_{k=0}^{n} \cdots + O(|z|^{-n-1})$$

---

[36] According to (1.5.1),

$$\frac{1}{\Gamma(\alpha)}\int_0^\infty e^{-zt} t^{\alpha+k-1}\, dt = (\alpha)_k z^{-\alpha-k}. \qquad \operatorname{Re}\alpha > 0, \quad \operatorname{Re} z > 0, \qquad k = 0, 1, 2, \ldots$$

[37] For complex a and b we have

$$|a^b| = |a|^{\operatorname{Re} b} e^{-\operatorname{Im} b \cdot \arg a} \leqslant |a|^{\operatorname{Re} b} e^{\pi|\operatorname{Im} b|}.$$

which again gives (9.12.1). Therefore the condition imposed on n can be dropped, and (9.12.1) is valid for arbitrary n.

Next we get rid of the restriction imposed on the parameter α. Suppose α satisfies the weaker condition $\operatorname{Re} \alpha > -1$. Then $\operatorname{Re}(\alpha + 1) > 0$, and formula (9.12.1) can be applied to each of the hypergeometric functions in the right-hand side of the identity

$$\Psi(\alpha, \gamma; z) = z\Psi(\alpha + 1, \gamma + 1; z) + (1 + \alpha - \gamma)\Psi(\alpha + 1, \gamma; z), \quad (9.12.2)$$

obtained by replacing α by $\alpha + 1$ in (9.10.14). Carrying out the necessary calculations, we again arrive at the asymptotic representation (9.12.1), but this time with the condition $\operatorname{Re} \alpha > -1$. Repeating this argument, we see that (9.12.1) holds for arbitrary values of α. Moreover, by slightly modifying the method used to prove (9.12.1), we can replace the condition $|\arg z| \leqslant \frac{1}{2}\pi - \delta$ by the weaker condition $|\arg z| \leqslant \pi - \delta$.[38] Thus, finally, we arrive at the following asymptotic representation of $\Psi(\alpha, \gamma; z)$ for large $|z|$:

$$\Psi(\alpha, \gamma; z) = z^{-\alpha}\left[\sum_{k=0}^{n} \frac{(-1)^k(\alpha)_k(1 + \alpha - \gamma)_k}{k!} z^{-k} + O(|z|^{-n-1})\right],$$
$$|\arg z| \leqslant \pi - \delta. \quad (9.12.3)$$

The corresponding asymptotic representation of the function $\Phi(\alpha, \gamma; z)$ can be deduced from (9.12.3) and the relation

$$\Phi(\alpha, \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\gamma - \alpha)} e^{\pm \alpha\pi i}\Psi(\alpha, \gamma; z) + \frac{\Gamma(\gamma)}{\Gamma(\alpha)} e^{\pm(\alpha-\gamma)\pi i} e^{z}\Psi(\gamma - \alpha, \gamma; -z),$$
$$|\arg z| < \pi, \quad -z = ze^{\mp\pi i}, \quad \gamma \neq 0, -1, -2, \ldots, \quad (9.12.4)$$

which is the inverse of (9.10.3), where the plus sign is chosen if $\operatorname{Im} z > 0$ and the minus sign if $\operatorname{Im} z < 0$. To prove (9.12.4), we assume that $\gamma \neq 0, \pm 1, \pm 2, \ldots$ and use (9.10.3):

$$\Psi(\alpha, \gamma; z) = \frac{\Gamma(1 - \gamma)}{\Gamma(\alpha - \gamma + 1)} \Phi(\alpha, \gamma; z) + \frac{\Gamma(\gamma - 1)}{\Gamma(\alpha)} z^{1-\gamma}\Phi(1 + \alpha - \gamma, 2 - \gamma; z). \quad (9.12.5)$$

Replacing α by $\gamma - \alpha$ and z by $-z = ze^{\mp\pi i}$, we obtain

$$e^{z}\Psi(\gamma - \alpha, \gamma; -z) = \frac{\Gamma(1 - \gamma)}{\Gamma(1 - \alpha)} \Phi(\alpha, \gamma; z) - \frac{\Gamma(\gamma - 1)}{\Gamma(\gamma - \alpha)} z^{1-\gamma}e^{\pm\gamma\pi i}\Phi(1 + \alpha - \gamma, 2 - \gamma; z), \quad (9.12.6)$$

---

[38] Instead of (9.11.6), use the integral representation

$$\Psi(\alpha, \gamma; z) = \frac{1}{\Gamma(\alpha)} \int_0^{\infty \cdot e^{i\theta}} e^{-zt}t^{\alpha-1}(1 + t)^{\gamma-\alpha-1}\, dt, \qquad \operatorname{Re} \alpha > 0,$$

where

$$\theta = \begin{cases} \dfrac{\pi}{2} & \text{if } -(\pi - \delta) \leqslant \arg z \leqslant -\left(\dfrac{\pi}{2} - \delta\right), \\ -\dfrac{\pi}{2} & \text{if } \dfrac{\pi}{2} - \delta \leqslant \arg z \leqslant \pi - \delta. \end{cases}$$

where we have used (9.11.2). Eliminating $\Phi(1 + \alpha - \gamma, 2 - \gamma; z)$ from (9.12.5–6), we arrive at (9.12.4) after some simple calculations, where the validity of the result for positive integral values of γ follows from the principle of analytic continuation. Substituting (9.12.3) into (9.12.4), we find the desired asymptotic representation of $\Phi(\alpha, \gamma; z)$ for large $|z|$:

$$\begin{aligned}\Phi(\alpha, \gamma; z)& \\ &= \frac{\Gamma(\gamma)}{\Gamma(\gamma - \alpha)} e^{\pm \alpha\pi i} z^{-\alpha}\left[\sum_{k=0}^{n} \frac{(-1)^k(\alpha)_k(1 + \alpha - \gamma)_k}{k!} z^{-k} + O(|z|^{-n-1})\right] \\ &\quad + \frac{\Gamma(\gamma)}{\Gamma(\alpha)} e^z z^{-(\gamma-\alpha)}\left[\sum_{k=0}^{n} \frac{(\gamma - \alpha)_k(1 - \alpha)_k}{k!} z^{-k} + O(|z|^{-n-1})\right],\end{aligned}$$

$$|\arg z| \leqslant \pi - \delta, \qquad \gamma \neq 0, -1, -2, \ldots \tag{9.12.7}$$

As before, the plus sign corresponds to $\operatorname{Im} z > 0$ and the minus sign to $\operatorname{Im} z < 0$. If $|\arg z| \leqslant \frac{1}{2}\pi - \delta$, the first term is small compared to the second, and (9.12.7) takes the form

$$\Phi(\alpha, \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)} e^z z^{-(\gamma-\alpha)}\left[\sum_{k=0}^{n} \frac{(\gamma - \alpha)_k(1 - \alpha)_k}{k!} z^{-k} + O(|z|^{-n-1})\right],$$

$$|\arg z| \leqslant \frac{\pi}{2} - \delta, \quad \alpha, \gamma \neq 0, -1, -2, \ldots \tag{9.12.8}$$

9.13. Representation of Various Functions in Terms of the Confluent Hypergeometric Functions

As we now show, various familiar functions of mathematical analysis are special cases of the confluent hypergeometric functions $\Phi(\alpha, \gamma; z)$ and $\Psi(\alpha, \gamma; z)$, corresponding to suitable choices of the parameters α, γ and the variable z. Particular attention will be devoted to the special functions introduced in Chapters 2–5.

1. *Elementary functions.* Some typical relations involving elementary functions are

$$\Phi(\alpha, \alpha; z) = \sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z,$$

$$\Phi(1, 2; z) = \sum_{k=0}^{\infty} \frac{z^k}{(k + 1)!} = \frac{e^z - 1}{z},$$

$$\Phi(-2, 1; z) = 1 - 2z + \tfrac{1}{2}z^2.$$

2. *Error functions.* It follows from (2.1.5) and (2.1.2) that the error function has the expansion

$$\operatorname{Erf} z = \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k+1}}{k!(2k+1)} = z \sum_{k=0}^{\infty} \frac{(\frac{1}{2})_k}{(\frac{3}{2})_k} \frac{(-z^2)^k}{k!},$$

and hence

$$\operatorname{Erf} z = z\Phi(\tfrac{1}{2}, \tfrac{3}{2}; -z^2). \tag{9.13.1}$$

Similarly, the complementary error function (2.1.6) can be written in the form

$$\operatorname{Erfc} z = \int_z^{\infty} e^{-t^2}\, dt = \tfrac{1}{2} z e^{-z^2} \int_0^{\infty} \frac{e^{-z^2 s}}{\sqrt{1+s}}\, ds,$$

if we set $t = z\sqrt{1+s}$. Then, according to the integral representation (9.11.6),[39]

$$\operatorname{Erfc} z = \tfrac{1}{2} z e^{-z^2} \Psi(1, \tfrac{3}{2}; z^2),$$

or

$$\operatorname{Erfc} z = \tfrac{1}{2} e^{-z^2} \Psi(\tfrac{1}{2}, \tfrac{1}{2}; z^2), \qquad |\arg z| < \frac{\pi}{2}, \tag{9.13.2}$$

where we have used (9.10.8).

3. *The function $F(z)$.* Next we consider the function $F(z)$, related to the probability integral of imaginary argument (see Sec. 2.3). It follows from (2.3.4) that

$$F(z) = \sum_{k=0}^{\infty} \frac{(-1)^k 2^k z^{2k+1}}{1\cdot 3 \cdots (2k+1)} = z \sum_{k=0}^{\infty} \frac{(1)_k (-z^2)^k}{k!(\frac{3}{2})_k},$$

and hence

$$F(z) = z\Phi(1, \tfrac{3}{2}; -z^2). \tag{9.13.3}$$

4. *Fresnel integrals.* Combining (2.4.6), (2.1.5) and (9.13.1), we find that

$$\begin{aligned} C(z) &= \frac{z}{2}\left[\Phi\left(\frac{1}{2}, \frac{3}{2}; \frac{\pi i z^2}{2}\right) + \Phi\left(\frac{1}{2}, \frac{3}{2}; -\frac{\pi i z^2}{2}\right)\right], \\ S(z) &= \frac{z}{2i}\left[\Phi\left(\frac{1}{2}, \frac{3}{2}; \frac{\pi i z^2}{2}\right) - \Phi\left(\frac{1}{2}, \frac{3}{2}; -\frac{\pi i z^2}{2}\right)\right]. \end{aligned} \tag{9.13.4}$$

5. *The exponential integral.* By definition,

$$\operatorname{Ei}(-z) = -\int_z^{\infty} \frac{e^{-t}}{t}\, dt, \qquad |\arg z| < \pi$$

---

[39] In the derivation we assume that $z > 0$, and then use analytic continuation to extend (9.13.2) into the domain $|\arg z| < \pi/2$.

[cf. (3.1.2)], and hence, setting $t = z(1 + s)$ and using the integral representation (9.11.6), we have

$$\mathrm{Ei}(-z) = -e^{-z} \int_0^\infty \frac{e^{-zs}}{1 + s}\, ds = -e^{-z}\Psi(1, 1; z),$$

or

$$\mathrm{Ei}(z) = -e^{z}\Psi(1, 1; -z), \qquad |\arg(-z)| < \pi. \tag{9.13.5}$$

6. *The sine and cosine integrals.* Combining (3.3.6) and (9.13.5), we find that

$$\mathrm{Ci}(z) = -\frac{1}{2} e^{-iz}\Psi(1, 1; ze^{\pi i/2}) - \frac{1}{2} e^{iz}\Psi(1, 1; ze^{-\pi i/2}), \qquad |\arg z| < \frac{\pi}{2},$$

$$\mathrm{Si}(z) = \frac{\pi}{2} + \frac{1}{2i} e^{-iz}\Psi(1, 1; ze^{\pi i/2}) - \frac{1}{2i} e^{iz}\Psi(1, 1; ze^{-\pi i/2}), \ |\arg z| < \frac{\pi}{2}. \tag{9.13.6}$$

7. *The logarithmic integral.* It is an immediate consequence of (3.4.3) and (9.13.5) that

$$\mathrm{li}(z) = -z\Phi(1, 1; -\log z), \qquad |\arg z| < \pi, \qquad |\arg(1 - z)| < \pi. \tag{9.13.7}$$

8. *Hermite polynomials.* According to (4.9.2), the even Hermite polynomials can be written in the form

$$\begin{aligned} H_{2n}(z) &= \sum_{k=0}^{n} (-1)^k \frac{(2n)!}{k!(2n-2k)!}(2z)^{2n-2k} = (-1)^n (2n)! \sum_{k=0}^{n} \frac{(-1)^k (2z)^{2k}}{(n-k)!(2k)!} \\ &= (-1)^n \frac{(2n)!}{n!} \sum_{k=0}^{n} \frac{(-n)_k (2z)^{2k}}{(2k)!} = (-1)^n \frac{(2n)!}{n!} \sum_{k=0}^{n} \frac{(-n)_k (z^2)^k}{(\frac{1}{2})_k k!}, \end{aligned}$$

since

$$(2k)! = 2^{2k}(\tfrac{1}{2})_k k!,$$

and therefore

$$H_{2n}(z) = (-1)^n \frac{(2n)!}{n!}\Phi(-n, \tfrac{1}{2}; z^2). \tag{9.13.8}$$

For the odd Hermite polynomials we have the analogous formula

$$H_{2n+1}(z) = (-1)^n \frac{(2n+1)!}{n!} 2z\Phi(-n, \tfrac{3}{2}; z^2). \tag{9.13.9}$$

9. *Laguerre polynomials.* It follows from (4.17.2) that

$$L_n^\alpha(z) = \sum_{k=0}^{n} \frac{\Gamma(n + \alpha + 1)}{\Gamma(k + \alpha + 1)} \frac{(-z)^k}{k!(n-k)!} = \frac{(\alpha + 1)_n}{n!} \sum_{k=0}^{n} \frac{(-n)_k z^k}{(\alpha + 1)_k k!},$$

and hence

$$L_n^\alpha(z) = \frac{(\alpha + 1)_n}{n!}\Phi(-n, \alpha + 1; z). \tag{9.13.10}$$

10. *Cylinder functions.* Assuming temporarily that $\operatorname{Re}\nu > -\frac{1}{2}$, we set $s = \frac{1}{2}(1 + t)$ in the integral representation (5.10.3), obtaining

$$J_\nu(z) = \frac{2^{2\nu}(z/2)^\nu e^{-iz}}{\Gamma(\frac{1}{2})\Gamma(\nu + \frac{1}{2})} \int_0^1 e^{2izs} s^{\nu - \frac{1}{2}}(1 - s)^{\nu - \frac{1}{2}}\, ds.$$

Therefore, according to (9.11.1),

$$J_\nu(z) = \frac{2^{2\nu}(z/2)^\nu e^{-iz}\Gamma(\nu + \frac{1}{2})}{\Gamma(\frac{1}{2})\Gamma(2\nu + 1)}\, \Phi(\nu + \tfrac{1}{2}, 2\nu + 1; 2iz),$$

or

$$J_\nu(z) = \frac{(z/2)^\nu}{\Gamma(\nu + 1)}\, e^{-iz}\Phi(\nu + \tfrac{1}{2}, 2\nu + 1; 2iz), \qquad |\arg z| < \pi, \quad (9.13.11)$$

where we have used the duplication formula (1.2.3) for the gamma function. Then we use the principle of analytic continuation to show that (9.13.11) holds for arbitrary ν.

Similar representations can be obtained for the other cylinder functions. For example, it follows from (5.6.4), (9.13.11) and (9.10.3) that[40]

$$H_\nu^{(1)}(z) = -\frac{2i}{\sqrt{\pi}}\, e^{i(z - \nu\pi)}(2z)^\nu \Psi(\nu + \tfrac{1}{2}, 2\nu + 1; 2ze^{-\pi i/2}),$$
$$-\frac{\pi}{2} < \arg z < \pi, \quad (9.13.12)$$

$$H_\nu^{(2)}(z) = \frac{2i}{\sqrt{\pi}}\, e^{-i(z - \nu\pi)}(2z)^\nu \Psi(\nu + \tfrac{1}{2}, 2\nu + 1; 2ze^{\pi i/2}),$$
$$-\pi < \arg z < \frac{\pi}{2}. \quad (9.13.13)$$

Then, using (5.7.6), we obtain the following representations of the Bessel functions of imaginary argument:

$$I_\nu(z) = \frac{(z/2)^\nu}{\Gamma(\nu + 1)}\, e^{-z}\Phi(\nu + \tfrac{1}{2}, 2\nu + 1; 2z), \qquad |\arg z| < \pi, \quad (9.13.14)$$

$$K_\nu(z) = \sqrt{\pi}(2z)^\nu e^{-z}\, \Psi(\nu + \tfrac{1}{2}, 2\nu + 1; 2z), \qquad |\arg z| < \pi. \quad (9.13.15)$$

11. *Whittaker functions.* A class of functions related to the confluent hypergeometric functions, and often encountered in the applications, consists of the *Whittaker functions*, defined by the formulas[41]

$$\begin{aligned} M_{k,\mu}(z) &= z^{\mu + \frac{1}{2}} e^{-z/2}\Phi(\tfrac{1}{2} - k + \mu, 2\mu + 1; z), \quad |\arg z| < \pi, \\ W_{k,\mu}(z) &= z^{\mu + \frac{1}{2}} e^{-z/2}\Psi(\tfrac{1}{2} - k + \mu, 2\mu + 1; z), \quad |\arg z| < \pi. \end{aligned} \quad (9.13.16)$$

[40] We also use formulas (9.11.2) and (1.2.2–3).

[41] E. T. Whittaker and G. N. Watson, *op. cit.*, Chap. 16.

9.14. Generalized Hypergeometric Functions

Consider the power series

$$\sum_{k=0}^{\infty} \frac{\prod_{r=1}^{p} (\alpha_r)_k}{\prod_{s=1}^{q} (\gamma_s)_k} \frac{z^k}{k!} = \sum_{k=0}^{\infty} \frac{(\alpha_1)_k \cdots (\alpha_p)_k}{(\gamma_1)_k \cdots (\gamma_q)_k} \frac{z^k}{k!}, \tag{9.14.1}$$

where p and q are nonnegative integers ($p,q = 0, 1, 2, \ldots$) satisfying the condition $p \leqslant q + 1$, z is a complex variable, α_r and γ_s are arbitrary parameters (except that $\gamma_s \neq 0, -1, -2, \ldots$), and $(\lambda)_k = \Gamma(\lambda + k)/\Gamma(\lambda)$.[42] Using the ratio test, we see at once that the radius of convergence of the series (9.14.1) equals ∞ if $p \leqslant q$ and 1 if $p = q + 1$. The sum of the series (9.14.1) is called the *generalized hypergeometric function*, and is denoted by the symbol

$${}_pF_q\left(\begin{matrix}\alpha_1, \ldots, \alpha_p; z\\ \gamma_1, \ldots, \gamma_q\end{matrix}\right),$$

or more concisely, by ${}_pF_q(\alpha_r; \gamma_s; z)$, i.e.,

$${}_pF_q(\alpha_r; \gamma_s; z) = \sum_{k=0}^{\infty} \frac{\prod_{r=1}^{p} (\alpha_r)_k}{\prod_{s=1}^{q} (\gamma_s)_k} \frac{z^k}{k!}. \tag{9.14.2}$$

Clearly, ${}_pF_q(\alpha_r; \gamma_s; z)$ is an entire function of z if $p \leqslant q$. The function ${}_{q+1}F_q(\alpha_r; \gamma_s; z)$ is originally defined only in the disk $|z| < 1$, but can be extended outside this disk by using analytic continuation.

The following are the simplest generalized hypergeometric functions:

$${}_0F_0(\alpha_r; \gamma_s; z) = \sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z,$$

$${}_1F_0(\alpha_r; \gamma_s; z) = \sum_{k=0}^{\infty} \frac{(\alpha_1)_k}{k!} z^k = (1 - z)^{-\alpha_1},$$

$${}_0F_1(\alpha_r; \gamma_s; z) = \sum_{k=0}^{\infty} \frac{z^k}{(\gamma_1)_k k!} = \Gamma(\gamma_1) z^{-(\gamma_1 - 1)/2} I_{\gamma_1 - 1}(2z^{1/2}),$$

$${}_1F_1(\alpha_r; \gamma_s; z) = \sum_{k=0}^{\infty} \frac{(\alpha_1)_k}{(\gamma_1)_k} \frac{z^k}{k!} = \Phi(\alpha_1, \gamma_1; z),$$

$${}_2F_1(\alpha_r; \gamma_s; z) = \sum_{k=0}^{\infty} \frac{(\alpha_1)_k (\alpha_2)_k}{(\gamma_1)_k k!} = F(\alpha_1, \alpha_2; \gamma_1; z).$$

---

[42] As usual, the meaningless products

$$\prod_{r=1}^{0} \cdots, \quad \prod_{s=1}^{0} \cdots,$$

which appear when $p = 0$ or $q = 0$, are set equal to 1.

The last two examples show that the hypergeometric functions considered in this chapter are special cases of the more general function (9.14.2).

Some features of the theory of ordinary hypergeometric functions can be carried over to the case of generalized hypergeometric functions. For example, it is easily seen that the function $u = {}_pF_q(\alpha_r; \gamma_s; z)$ is a particular solution of the linear differential equation

$$\left[\delta \prod_{s=1}^{q} (\delta + \gamma_s - 1) - z \prod_{r=1}^{p} (\delta + \alpha_r)\right] u = 0 \tag{9.14.3}$$

of order $q + 1$, where δ denotes the operator $z(d/dz)$.[43] This equation reduces to (9.10.1) if $p = q = 1$, and to the hypergeometric equation (9.2.16) of $p = 2, q = 1$. There is a well-developed theory of generalized hypergeometric functions, with appropriate recurrence relations, integral representations, etc.[44]

PROBLEMS

1. Starting from the integral representation (9.1.6), prove that

$$F(\alpha, \beta; \gamma; x + i0) - F(\alpha, \beta; \gamma; x - i0)$$
$$= \frac{2\pi i \Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(1 + \gamma - \alpha - \beta)} (x - 1)^{\gamma - \alpha - \beta} F(\gamma - \alpha, \gamma - \beta; 1 + \gamma - \alpha - \beta; 1 - x),$$
$$x > 1, \quad \gamma \neq 0, -1, -2, \ldots$$

Hint. During the proof, assume that $\operatorname{Re} \alpha < 1$, $\operatorname{Re} \gamma > \operatorname{Re} \beta > 0$, and then use analytic continuation.

Comment. This formula shows why the cut $[1, \infty]$ is necessary in defining $F(\alpha, \beta; \gamma; z)$ for $\alpha, \beta \neq 0, -1, -2, \ldots$

2. Derive the formulas

$$\frac{d}{dz}(z^\alpha F) = \alpha z^{\alpha - 1} F(\alpha + 1), \quad \frac{d}{dz}(z^{\gamma - 1} F) = (\gamma - 1) z^{\gamma - 2} F(\gamma - 1),$$

where the notation is the same as in Sec. 9.2.

3. Prove the following identities:

$$F(2\alpha, 2\beta; \alpha + \beta + \tfrac{1}{2}; \tfrac{1}{2}) = \frac{\Gamma(\alpha + \beta + \frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(\alpha + \frac{1}{2})\Gamma(\beta + \frac{1}{2})}, \quad \alpha + \beta + \tfrac{1}{2} \neq 0, -1, -2, \ldots,$$

$$F(\alpha, \beta; 1 + \alpha - \beta; -1) = 2^{-\alpha} \frac{\Gamma(1 + \alpha - \beta)\Gamma(\frac{1}{2})}{\Gamma\left(1 - \beta + \dfrac{\alpha}{2}\right)\Gamma\left(\dfrac{1}{2} + \dfrac{\alpha}{2}\right)},$$
$$1 + \alpha - \beta \neq 0, -1, -2, \ldots$$

---

[43] Note that applying δ to u corresponds to multiplying u by k.

[44] For a summary of the theory and references for further reading, see the Bateman Manuscript Project, *Higher Transcendental Functions, Vol. 1*, Chap. 4. Some new results are given by N. E. Nörlund, *Sur les fonctions hypergéométriques d'ordre supérieur*, Mat.-Fys. Skr. Danske Vid. Selsk., **1**, no. 2 (1956).

4. Show that the *hypergeometric polynomials* $F(-n, \beta; \gamma; z)$ $(n = 0, 1, 2, \ldots, \gamma \neq 0, -1, -2, \ldots)$ can be defined as the expansion coefficients of the generating function

$$w(z, t) = (1 - t)^{\beta-\gamma}(1 - t + zt)^{-\beta} = \sum_{n=0}^{\infty} \frac{(\gamma)_n}{n!} F(-n, \beta; \gamma; z)t^n,$$
$$|t| < \min\{1, |z - 1|^{-1}\}.$$

5. Derive the integral representation

$$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\gamma)} F(\alpha, \beta; \gamma; z) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{\Gamma(\alpha + s)\Gamma(\beta + s)\Gamma(-s)}{\Gamma(\gamma + s)} (-z)^s\, ds,$$
$$\operatorname{Re}\alpha > 0, \quad \operatorname{Re}\beta > 0, \quad |\arg(-z)| < \pi, \quad \gamma \neq 0, -1, -2, \ldots,$$

where $\min\{\operatorname{Re}\alpha, \operatorname{Re}\beta\} < c < 0$.

Hint. Complete the contour of integration on the right with the arc of a circle of radius $R_n = n + \frac{1}{2}$ $(n \to \infty)$, and then use residue theory.

Comment. The restrictions imposed on the parameters can be eliminated by suitably deforming the contour of integration.[45]

6. Using term-by-term integration, verify the following formulas:

$$F(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(c)\Gamma(\gamma - c)}\int_0^1 t^{c-1}(1 - t)^{\gamma-c-1}F(\alpha, \beta; c; zt)\, dt,$$
$$\operatorname{Re}\gamma > \operatorname{Re} c > 0, \quad |\arg(1 - z)| < \pi,$$

$$F(\alpha, \beta; \gamma + 1; z) = \gamma\int_0^1 F(\alpha, \beta; \gamma; zt)t^{\gamma-1}\, dt, \quad \operatorname{Re}\gamma > 0, \quad |\arg(1 - z)| < \pi.$$

7. By analogy with Sec. 9.10, the *hypergeometric function of the second kind* $G(\alpha, \beta; \gamma; z)$ can be defined as

$$G(\alpha, \beta; \gamma; z) = \frac{\Gamma(1 - \gamma)}{\Gamma(\alpha - \gamma + 1)\Gamma(\beta - \gamma + 1)} F(\alpha, \beta; \gamma; z)$$
$$+ \frac{\Gamma(\gamma - 1)}{\Gamma(\alpha)\Gamma(\beta)} z^{1-\gamma}F(1 + \alpha - \gamma, 1 + \beta - \gamma; 2 - \gamma; z),$$
$$|\arg z| < \pi, \quad |\arg(1 - z)| < \pi, \quad \gamma \neq 0, \pm 1, \pm 2, \ldots$$

Prove that $G(\alpha, \beta; \gamma; z)$ satisfies the relation

$$G(\alpha, \beta; \gamma; z) = z^{1-\gamma}G(\alpha - \gamma + 1, \beta - \gamma + 1; 2 - \gamma; z).$$

8. Repeating the considerations of Sec. 9.10, show that $G(\alpha, \beta; \gamma; z)$ is an entire function of α, β, γ, and derive the formula

$$G(\alpha, \beta; n + 1; z) = \frac{(-1)^{n+1}}{\Gamma(\alpha - n)\Gamma(\beta - n)}\sum_{k=0}^{\infty} \frac{(\alpha)_k(\beta)_k}{(n + k)!k!} z^k$$
$$\times [\psi(\alpha + k) + \psi(\beta + k) - \psi(1 + k) - \psi(n + 1 + k) + \log z]$$
$$+ \frac{1}{\Gamma(\alpha)\Gamma(\beta)}\sum_{k=0}^{n-1} \frac{(-1)^k(n - k - 1)!(\alpha - n)_k(\beta - n)_k}{k!} z^{k-n},$$
$$|\arg z| < \pi, \quad |z| < 1, \quad n = 0, 1, 2, \ldots, \quad \alpha, \beta \neq 0, -1, -2, \ldots$$

[45] E. T. Whittaker and G. N. Watson, *op. cit.*, p. 286.

9. Prove that the functions $F(\alpha, \beta; \gamma; z)$ and $G(\alpha, \beta; \gamma; z)$ are a pair of solutions of the hypergeometric equation (9.2.16) with Wronskian

$$W\{F(\alpha, \beta; \gamma; z), G(\alpha, \beta; \gamma; z)\} = -\frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)} z^{-\gamma}(1-z)^{\gamma-\alpha-\beta-1},$$

$$|\arg(1-z)| < \pi, \quad |\arg z| < \pi, \quad \gamma \neq 0, -1, -2, \ldots$$

Comment. It follows that the two solutions are linearly independent if $\alpha, \beta \neq 0, -1, -2, \ldots$

10. Find differentiation formulas and recurrence relations for the function $G(\alpha, \beta; \gamma; z)$.

Hint. Use the corresponding relations for the function $F(\alpha, \beta; \gamma; z)$.

11. Derive the integral representation

$$\frac{\Gamma(\alpha)}{\Gamma(\gamma)} \Phi(\alpha, \gamma; z) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\Gamma(\alpha+s)\Gamma(-s)}{\Gamma(\gamma+s)} (-z)^s \, ds,$$

$$\operatorname{Re} \alpha > 0, \quad -\operatorname{Re} \alpha < c < 0, \quad \gamma \neq 0, -1, -2, \ldots \quad |\arg(-z)| < \frac{\pi}{2}.$$

Hint. Use residue theory.

12. Derive the integral representation

$$\Phi(\alpha, \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\gamma-\alpha)} e^z z^{(1-\gamma)/2} \int_0^\infty e^{-t} t^{\frac{1}{2}(\gamma-1)-\alpha} J_{\gamma-1}(2\sqrt{zt}) \, dt.$$

$$\operatorname{Re}(\gamma-\alpha) > 0, \quad |\arg z| < \pi, \quad \gamma \neq 0, -1, -2, \ldots$$

Hint. Expand the Bessel function in power series, and then integrate term by term.

13. Derive the integral representation

$$\Psi(\alpha, \gamma; z) = \frac{2z^{(1-\gamma)/2}}{\Gamma(\alpha)\Gamma(\alpha-\gamma+1)} \int_0^\infty e^{-t} t^{\alpha-\frac{1}{2}(1+\gamma)} K_{\gamma-1}(2\sqrt{zt}) \, dt,$$

$$\operatorname{Re} \alpha > 0, \quad \operatorname{Re}(\alpha-\gamma) > -1, \quad |\arg z| < \pi,$$

where $K_\nu(z)$ is Macdonald's function.

14. Prove the formulas

$$\Phi(\alpha, \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(c)\Gamma(\gamma-c)} \int_0^1 t^{c-1}(1-t)^{\gamma-c-1} \Phi(\alpha, c; zt) \, dt,$$

$$\operatorname{Re} \gamma > \operatorname{Re} c > 0,$$

$$\Phi(\alpha, \gamma+1; z) = \gamma \int_0^1 \Phi(\alpha, \gamma; zt) t^{\gamma-1} \, dt, \qquad \operatorname{Re} \gamma > 0.$$

15. Show that the Laplace transform of $\Phi(\alpha, \gamma; x)$ is

$$\overline{\Phi}(\alpha, \gamma; x) = \frac{1}{p} F\left(\alpha, 1; \gamma; \frac{1}{p}\right).$$

16. Verify that the Whittaker functions $M_{k,\mu}(z)$ and $W_{k,\mu}(z)$ are a pair of solutions of *Whittaker's equation*

$$u'' + \left(-\frac{1}{4} + \frac{k}{z} + \frac{\frac{1}{4} - \mu^2}{z^2}\right)u = 0,$$

with Wronskian

$$W\{M_{k,\mu}(z),\ W_{k,\mu}(z)\} = -\frac{\Gamma(2\mu + 1)}{\Gamma(\frac{1}{2} - k + \mu)}, \qquad 2\mu + 1 \neq 0, -1, -2, \ldots$$

Hint. Use the definitions (9.13.16).

17. Derive the integral representation[46]

$$W_{k,\mu}(z) = \frac{z^k e^{-z/2}}{\Gamma(\mu - k + \frac{1}{2})} \int_0^\infty e^{-t} t^{\mu - k - \frac{1}{2}} \left(1 + \frac{t}{z}\right)^{\mu + k - \frac{1}{2}} dt,$$
$$\operatorname{Re}(\mu - k + \tfrac{1}{2}) > 0, \qquad |\arg z| < \pi.$$

18. Using the result of the preceding problem, prove the asymptotic formula

$$W_{k,\mu}(z) \approx e^{-z/2} z^k, \qquad |z| \to \infty, \quad |\arg z| \leqslant \pi - \delta.$$

19. Using the results of Sec. 9.13, derive the following representations of various special functions in terms of $W_{k,\mu}(z)$:

$$\operatorname{Erfc} z = \frac{1}{2\sqrt{z}} e^{-z^2/2} W_{-\frac{1}{4},\frac{1}{4}}(z^2), \qquad |\arg z| < \frac{\pi}{2},$$

$$\operatorname{Ei}(z) = -\frac{1}{\sqrt{-z}} e^{z/2} W_{-\frac{1}{2},0}(-z), \qquad |\arg(-z)| < \pi,$$

$$\operatorname{li}(z) = -\sqrt{\frac{z}{-\log z}}\, W_{-\frac{1}{2},0}(-\log z), \qquad |\arg z| < \pi, \quad |\arg(1 - z)| < \pi,$$

$$K_\nu(z) = \sqrt{\frac{\pi}{2z}}\, W_{0,\nu}(2z), \qquad |\arg z| < \pi.$$

20. Prove that

$$\frac{d}{dz}\, {}_pF_q(\alpha_r; \gamma_s; z) = \frac{\prod_{r=1}^{p} \alpha_r}{\prod_{s=1}^{q} \gamma_s}\, {}_pF_q(\alpha_r + 1; \gamma_s + 1; z).$$

21. Prove that

$${}_{p+1}F_{q+1}(\alpha_r; \gamma_s; z)$$
$$= \frac{\Gamma(\gamma_{q+1})}{\Gamma(\alpha_{p+1})\Gamma(\gamma_{q+1} - \alpha_{p+1})} \int_0^1 t^{\alpha_{p+1} - 1}(1 - t)^{\gamma_{q+1} - \alpha_{p+1} - 1}\, {}_pF_q(\alpha_r; \gamma_s; zt)\, dt,$$
$$\operatorname{Re} \gamma_{q+1} > \operatorname{Re} \alpha_{p+1} > 0,$$

where $|\arg(1 - z)| < \pi$ if $p = q + 1$.

22. Derive the formula

$$[F(\alpha, \beta; \alpha + \beta + \tfrac{1}{2}; z)]^2 = {}_3F_2\left(\begin{matrix} 2\alpha, 2\beta, \alpha + \beta; z \\ \alpha + \beta + \frac{1}{2}, 2\alpha + 2\beta \end{matrix}\right).$$

---

[46] E. T. Whittaker and G. N. Watson, *op. cit.*, p. 340.

Hint. Find a third-order linear differential equation satisfied by the square of the function $F(\alpha, \beta; \alpha + \beta + \frac{1}{2}; z)$,[47] and show that the function

$$ {}_3F_2\left(\begin{matrix} 2\alpha, 2\beta; \alpha + \beta; z \\ \alpha + \beta + \frac{1}{2}; 2\alpha + 2\beta \end{matrix}\right) $$

is the solution of this equation which is analytic in a neighborhood of the point $z = 0$.

---

[47] E. T. Whittaker and G. N. Watson, *op. cit.*, Problems 10–11, p. 298.

2.11 Asymptotic Expansion and Mellin-Barnes transform

Barnes' integral representation

The Mellin transform of a function is given by

$$F(s) := \int_0^\infty x^{s-1} f(x)\,dx.$$

The inversion formula is given by

$$f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-s} F(s)\,ds, \quad c > 0.$$

Note that the definition of the gamma function,

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x}\,dx, \quad \operatorname{Re} s > 0,$$

implies that $\Gamma(s)$ is the Mellin transform of e^{-x}. The inversion formula reads

$$e^{-x} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-s} \Gamma(s)\,ds, \quad c > 0. \tag{1}$$

This can be proved directly by using Cauchy's residue theorem. Consider the rectangular contour $\mathcal{C}$ with vertices $c \pm iR$ $(c > 0)$ and $-(N + 1/2) \pm iR$, where N is a positive integer. Then the poles of $\Gamma(s)$ inside this contour are at $0, -1, -2, \ldots, -N$ with residues $(-1)^j/j!$ for $j = 0, 1, 2, \ldots, N$ respectively. Hence, Cauchy's residue theorem implies that

$$\frac{1}{2\pi i} \int\limits_{\mathcal{C}} x^{-s} \Gamma(s)\,ds = \sum_{j=0}^{N} \frac{(-1)^j}{j!} x^j.$$

Now we let R and N tend to infinity. Then Stirling's asymptotic formula for the gamma function implies that the integral on $\mathcal{C}$ minus the line joining $c - iR$ and $c + iR$ tends to zero. Hence we have

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-s} \Gamma(s)\,ds = \lim_{N\to\infty} \sum_{j=0}^{N} \frac{(-1)^j}{j!} x^j = \sum_{j=0}^{\infty} \frac{(-x)^j}{j!} = e^{-x}.$$

This proves (1).

Another example of a Mellin transform is:

$$B(s,t) = \int_0^1 x^{s-1} (1-x)^{t-1}\,dx, \quad \operatorname{Re} s > 0, \quad \operatorname{Re} t > 0.$$

This implies that the beta function $B(s,t)$ is de Mellin transform of the function

$$f(x) := \begin{cases} (1-x)^{t-1}, & 0 < x < 1 \\ 0, & x \geq 1. \end{cases}$$

The inversion formula then reads

$$f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-s} B(s,t)\, ds = \frac{\Gamma(t)}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-s}\frac{\Gamma(s)}{\Gamma(s+t)}\, ds, \quad \operatorname{Re} t > 0, \quad c > 0.$$

Another representation of the beta function is:

$$B(s,t) = \int_0^\infty \frac{x^{s-1}}{(1+x)^{s+t}}\, dx, \quad \operatorname{Re} s > 0, \quad \operatorname{Re} t > 0.$$

This implies that the beta function $B(s, t-s)$ is the Mellin transform of the function

$$g(x) = \frac{1}{(1+x)^t}.$$

The inversion formula then reads

$$\begin{aligned}\frac{1}{(1+x)^t} &= g(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-s} B(s,t-s)\, ds \\ &= \frac{1}{2\pi i}\frac{1}{\Gamma(t)}\int_{c-i\infty}^{c+i\infty} x^{-s}\Gamma(s)\Gamma(t-s)\, ds, \quad 0 < c < \operatorname{Re} t.\end{aligned}$$

These examples suggest that we might obtain a complex integral representation for the hypergeometric function ${}_2F_1$ by finding its Mellin transform:

$$\begin{aligned}\int_0^\infty x^{s-1}{}_2F_1\left(\begin{matrix}a,\, b\\ c\end{matrix}; -x\right) dx &= \int_0^\infty x^{s-1}\frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1+xt)^{-a}\, dt\, dx \\ &= \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}\int_0^\infty \frac{x^{s-1}}{(1+xt)^a}\, dx\, dt.\end{aligned}$$

Here we used Euler's integral representation for the ${}_2F_1$. The substitution $xt = u$ now gives

$$\int_0^\infty \frac{x^{s-1}}{(1+xt)^a}\, dx = t^{-s}\int_0^\infty \frac{u^{s-1}}{(1+u)^a}\, du = t^{-s}\, B(s, a-s) = t^{-s}\,\frac{\Gamma(s)\Gamma(a-s)}{\Gamma(a)}.$$

Hence we have

$$\begin{aligned}\int_0^\infty x^{s-1}{}_2F_1\left(\begin{matrix}a,\, b\\ c\end{matrix}; -x\right) dx &= \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\,\frac{\Gamma(s)\Gamma(a-s)}{\Gamma(a)}\int_0^1 t^{b-s-1}(1-t)^{c-b-1}\, dt \\ &= \frac{\Gamma(c)\Gamma(s)\Gamma(a-s)}{\Gamma(b)\Gamma(c-b)\Gamma(a)}\, B(b-s, c-b) \\ &= \frac{\Gamma(c)\Gamma(s)\Gamma(a-s)}{\Gamma(b)\Gamma(c-b)\Gamma(a)}\,\frac{\Gamma(b-s)\Gamma(c-b)}{\Gamma(c-s)} \\ &= \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)}\,\frac{\Gamma(s)\Gamma(a-s)\Gamma(b-s)}{\Gamma(c-s)}.\end{aligned}$$

Here we assumed that $\min(\operatorname{Re} a, \operatorname{Re} b) > \operatorname{Re} s > 0$. Applying the inversion formula for the Mellin transform we might expect that

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(c)}\,{}_2F_1\left(\begin{matrix}a,\, b\\ c\end{matrix};x\right) = \frac{1}{2\pi i}\int_{k-i\infty}^{k+i\infty}\frac{\Gamma(s)\Gamma(a-s)\Gamma(b-s)}{\Gamma(c-s)}(-x)^{-s}\,ds$$

for $\min(\operatorname{Re} a, \operatorname{Re} b) > k > 0$ and $c \neq 0, -1, -2, \ldots$.
In fact, this is Barnes' integral representation for the ${}_2F_1$ which is usually written as:

Theorem 1.

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(c)}\,{}_2F_1\left(\begin{matrix}a,\, b\\ c\end{matrix};z\right) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{\Gamma(a+s)\Gamma(b+s)\Gamma(-s)}{\Gamma(c+s)}(-z)^{s}\,ds, \quad |\arg(-z)| < \pi. \quad (2)$$

The path of integration is curved, if necessary, to separate the poles $s = -a-n$ and $s = -b-n$ from the poles $s = n$ with $n \in \{0, 1, 2, \ldots\}$. Such a contour always exists if a and b are not negative integers.

Proof. Let $\mathcal{C}$ be the closed contour formed by a part of the curve used in the theorem from $-(N+1/2)i$ to $(N+1/2)i$ together with the semicircle of radius $N+1/2$ to the right of the imaginary axis with 0 as center. We first show that the above integral defines an analytic function for $|\arg(-z)| \leq \pi - \delta$ with $\delta > 0$. By using Euler's reflection formula and Stirling's asymptotic formula for the gamma function we find for the integrand:

$$\frac{\Gamma(a+s)\Gamma(b+s)\Gamma(-s)}{\Gamma(c+s)}(-z)^s = -\frac{\Gamma(a+s)\Gamma(b+s)}{\Gamma(c+s)\Gamma(1+s)}\,\frac{\pi}{\sin\pi s}\,(-z)^s \sim -s^{a+b-c-1}\frac{\pi}{\sin\pi s}\,(-z)^s.$$

For $s = it$ we have

$$-s^{a+b-c-1}\frac{\pi}{\sin\pi s}\,(-z)^s = -(it)^{a+b-c-1}\,2\pi i\,\frac{e^{it(\ln|z|+i\arg(-z))}}{e^{-\pi t}-e^{\pi t}} = \mathcal{O}(|t|^{a+b-c-1}e^{-|t|\delta})$$

for $|\arg(-z)| \leq \pi - \delta$. This shows that the integral represents an analytic function in $|\arg(-z)| \leq \pi - \delta$ for every $\delta > 0$, which implies that it is analytic for $|\arg(-z)| < \pi$.

On the semicircular part of the contour $\mathcal{C}$ the integrand is

$$\mathcal{O}(N^{a+b-c-1})\frac{(-z)^s}{\sin\pi s}$$

for large N. For $s = (N+1/2)e^{i\theta}$ and $|z| < 1$ we have

$$\frac{(-z)^s}{\sin\pi s} = \mathcal{O}\left[e^{(N+1/2)(\cos\theta\,\ln|z| - \sin\theta\,\arg(-z) - \pi|\sin\theta|)}\right].$$

Since $-\pi+\delta \leq \arg(-z) \leq \pi - \delta$, the last expression is

$$\mathcal{O}\left[e^{(N+1/2)(\cos\theta\,\ln|z| - \delta|\sin\theta|)}\right].$$

Now we have $\cos\theta \geq \frac{1}{2}\sqrt{2}$ for $0 \leq |\theta| \leq \pi/4$ and $|\sin\theta| \geq \frac{1}{2}\sqrt{2}$ for $\pi/4 \leq |\theta| \leq \pi/2$. Hence, since $\ln|z| < 0$, the integrand is

$$\begin{cases} \mathcal{O}\left(N^{a+b-c-1}e^{\frac{1}{2}\sqrt{2}(N+1/2)\ln|z|}\right), & 0 \leq |\theta| \leq \pi/4 \\ \mathcal{O}\left(N^{a+b-c-1}e^{-\frac{1}{2}\sqrt{2}(N+1/2)\delta}\right), & \pi/4 \leq |\theta| \leq \pi/2. \end{cases}$$

This implies that the integral on the semicircle tends to zero for $N \to \infty$. The residue theorem then implies that the integral tends to the limit of the sums of the residues at $s = n$ with $n \in \{0, 1, 2, \ldots\}$, id est

$$\sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)\, n!} z^n = \frac{\Gamma(a)\Gamma(b)}{\Gamma(c)} \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n} \frac{z^n}{n!} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(c)}\, {}_2F_1\left(\begin{matrix} a,\, b \\ c \end{matrix}\, ;\, z\right).$$

This proves Barnes' integral representation (2).

The Mellin transform has a convolution property, which can be obtained as follows. Assume that

$$F(s) = \int_0^\infty x^{s-1} f(x)\, dx \quad \text{and} \quad G(s) = \int_0^\infty x^{s-1} g(x)\, dx,$$

then we have

$$f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} F(s) x^{-s}\, ds \quad \text{and} \quad g(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} G(s) x^{-s}\, ds.$$

This implies that

$$\begin{aligned}
\int_0^\infty x^{s-1} f(x) g(x)\, dx &= \frac{1}{2\pi i} \int_0^\infty x^{s-1} g(x) \left(\int_{c-i\infty}^{c+i\infty} F(t) x^{-t}\, dt \right) dx \\
&= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} F(t) \left(\int_0^\infty x^{s-t-1} g(x)\, dx \right) dt \\
&= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} F(t) G(s-t)\, dt.
\end{aligned}$$

As an application we consider

$$f(x) = \frac{x^b}{(1+x)^a} \qquad \longleftrightarrow \qquad F(s) = \frac{\Gamma(b+s)\Gamma(a-b-s)}{\Gamma(a)}$$

and

$$g(x) = \frac{x^d}{(1+x)^c} \qquad \longleftrightarrow \qquad G(s) = \frac{\Gamma(d+s)\Gamma(c-d-s)}{\Gamma(c)},$$

which leads to

$$\begin{aligned}
&\frac{1}{2\pi i} \int_{k-i\infty}^{k+i\infty} \frac{\Gamma(b+s)\Gamma(a-b-s)\Gamma(d+1-s)\Gamma(c-d-1+s)}{\Gamma(a)\Gamma(c)}\, ds \\
&\quad = \frac{1}{2\pi i} \int_{k-i\infty}^{k+i\infty} F(s) G(1-s)\, ds = \int_0^\infty f(x) g(x)\, dx = \int_0^\infty \frac{x^{b+d}}{(1+x)^{a+c}}\, dx \\
&\quad = B(b+d+1, a+c-b-d-1) = \frac{\Gamma(b+d+1)\Gamma(a+c-b-d-1)}{\Gamma(a+c)}
\end{aligned}$$

for a suitable k. By renaming the parameters, this can be written as

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\Gamma(a+s)\Gamma(b+s)\Gamma(c-s)\Gamma(d-s)\,ds=\frac{\Gamma(a+c)\Gamma(a+d)\Gamma(b+c)\Gamma(b+d)}{\Gamma(a+b+c+d)},$$

which holds for $\operatorname{Re} a>0$, $\operatorname{Re} b>0$, $\operatorname{Re} c>0$ and $\operatorname{Re} d>0$, shortly for $\operatorname{Re}(a,b,c,d)>0$. This integral is called the Mellin-Barnes integral. More general we have

Theorem 2. *If the path of integration is curved to separate the poles of* $\Gamma(a+s)\Gamma(b+s)$ *from the poles of* $\Gamma(c-s)\Gamma(d-s)$, *then we have*

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\Gamma(a+s)\Gamma(b+s)\Gamma(c-s)\Gamma(d-s)\,ds=\frac{\Gamma(a+c)\Gamma(a+d)\Gamma(b+c)\Gamma(b+d)}{\Gamma(a+b+c+d)}.$$

Proof. Again we use Euler's reflection formula to write the integrand as

$$\frac{\Gamma(a+s)\Gamma(b+s)}{\Gamma(1-c+s)\Gamma(1-d+s)}\cdot\frac{\pi^2}{\sin\pi(c-s)\,\sin\pi(d-s)}.$$

Let $\mathcal{C}$ be the closed contour formed by a part of the curve used in the theorem from $-iR$ to iR together with a semicircle of radius R to the right of the imaginary axis with 0 as center. By Stirling's asymptotic formula the integrand is

$$\mathcal{O}\left(s^{a+b+c+d-2}\,e^{-2\pi|\operatorname{Im} s|}\right)\quad\text{for}\quad |s|\to\infty\quad\text{on}\quad\mathcal{C}.$$

Since $\operatorname{Im} s$ can be arbitrarily small when $|s|$ is large, we assume that $\operatorname{Re}(a+b+c+d)<1$ to ensure that the integral on the semicircle tends to zero for $R\to\infty$. Then, by Cauchy's residue theorem (poles in $s=c+n$ and $s=d+n$ with $n\in\{0,1,2,\ldots\}$ with residues $(-1)^n/n!$), the integral equals

$$\sum_{n=0}^{\infty}\Gamma(a+c+n)\Gamma(b+c+n)\Gamma(d-c-n)\frac{(-1)^n}{n!}+\sum_{n=0}^{\infty}\Gamma(a+d+n)\Gamma(b+d+n)\Gamma(c-d-n)\frac{(-1)^n}{n!}.$$

Note that for $n\in\{0,1,2,\ldots\}$ we have

$$\begin{aligned}\Gamma(d-c-n)&=\frac{\Gamma(d-c)}{(d-c-1)(d-c-2)\cdots(d-c-n)}\\&=\frac{\Gamma(d-c)}{(-1)^n(1+c-d)(2+c-d)\cdots(n+c-d)}=\frac{\Gamma(d-c)}{(-1)^n(1+c-d)_n}.\end{aligned}$$

This implies that the integral equals

$$\Gamma(a+c)\Gamma(b+c)\Gamma(d-c)\sum_{n=0}^{\infty}\frac{(a+c)_n(b+c)_n}{(1+c-d)_n\,n!}+\Gamma(a+d)\Gamma(b+d)\Gamma(c-d)\sum_{n=0}^{\infty}\frac{(a+d)_n(b+d)_n}{(1+d-c)_n\,n!}.$$

Finally we use Gauss's summation formula and Euler's reflection formula to find that the integral equals

$$\begin{aligned}
& \Gamma(a+c)\Gamma(b+c)\Gamma(d-c)\,{}_2F_1\left(\begin{matrix} a+c,\, b+c \\ 1+c-d \end{matrix};1\right) \\
& \qquad + \Gamma(a+d)\Gamma(b+d)\Gamma(c-d)\,{}_2F_1\left(\begin{matrix} a+d,\, b+d \\ 1+d-c \end{matrix};1\right) \\
= \; & \Gamma(a+c)\Gamma(b+c)\Gamma(d-c)\frac{\Gamma(1+c-d)\Gamma(1-a-b-c-d)}{\Gamma(1-a-d)\Gamma(1-b-d)} \\
& \qquad + \Gamma(a+d)\Gamma(b+d)\Gamma(c-d)\frac{\Gamma(1+d-c)\Gamma(1-a-b-c-d)}{\Gamma(1-a-c)\Gamma(1-b-c)} \\
= \; & \frac{\Gamma(a+c)\Gamma(b+c)\Gamma(a+d)\Gamma(b+d)}{\Gamma(a+b+c+d)} \\
& \qquad\qquad \times \frac{\pi}{\sin\pi(d-c)}\frac{\sin\pi(a+d)}{\pi}\frac{\sin\pi(b+d)}{\pi}\frac{\pi}{\sin\pi(a+b+c+d)} \\
& \qquad + \frac{\Gamma(a+d)\Gamma(b+d)\Gamma(a+c)\Gamma(b+c)}{\Gamma(a+b+c+d)} \\
& \qquad\qquad \times \frac{\pi}{\sin\pi(c-d)}\frac{\sin\pi(a+c)}{\pi}\frac{\sin\pi(b+c)}{\pi}\frac{\pi}{\sin\pi(a+b+c+d)} \\
= \; & \frac{\Gamma(a+c)\Gamma(a+d)\Gamma(b+c)\Gamma(b+d)}{\Gamma(a+b+c+d)} \\
& \qquad\qquad \times \frac{\sin\pi(a+c)\sin\pi(b+c)-\sin\pi(a+d)\sin\pi(b+d)}{\sin\pi(c-d)\sin\pi(a+b+c+d)}.
\end{aligned}$$

By using trigonometric relations it can be shown that

$$\frac{\sin\pi(a+c)\sin\pi(b+c)-\sin\pi(a+d)\sin\pi(b+d)}{\sin\pi(c-d)\sin\pi(a+b+c+d)} = 1.$$

This proves the theorem for $\mathrm{Re}(a+b+c+d) < 1$. This condition can be removed by using analytic continuation of the parameters.

We thus find that*

$$(1)\ F(a,\ b;\ c;\ z)=\frac{\Gamma(c)\,\Gamma(c-a-b)}{\Gamma(c-a)\,\Gamma(c-b)}F(a,\ b;\ a+b+1-c;\ 1-z)$$
$$+\frac{\Gamma(c)\,\Gamma(a+b-c)}{\Gamma(a)\,\Gamma(b)}(1-z)^{c-a-b}F(c-a,\ c-b;\ 1+c-a-b;\ 1-z).$$

1.5. A definite integral for *F(a, b; c; z)*. Consider the tegral

$$I=\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\,dt,$$

here, for convergence, $R(c)>R(b)>0$, and $|z|<1$. It is sup- osed that the branch of $(1-tz)^{-a}$ is chosen so that $(1-tz)^{-a}\to 1$ $t\to 0$. Then

$$\begin{aligned} I&=\int_0^1 \sum_{n=0}^{\infty}\frac{(a)_n}{n!}z^n t^{b+n-1}(1-t)^{c-b-1}\,dt\\ &=\sum_{n=0}^{\infty}\frac{(a)_n}{n!}z^n\frac{\Gamma(b+n)\,\Gamma(c-b)}{\Gamma(c+n)}\\ &=\frac{\Gamma(b)\,\Gamma(c-b)}{\Gamma(c)}\sum_{n=0}^{\infty}\frac{(a)_n\,(b)_n}{n!\,(c)_n}z^n,\end{aligned}$$

e change in the order of integration and summation being easily astified. We therefore have, under the given conditions,

$$(1)\ F(a,b;\ c;\ z)=\frac{\Gamma(c)}{\Gamma(b)\,\Gamma(c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\,dt.$$

When $z=1$, the integral on the right reduces to a beta function nd we are led again to Gauss's theorem.

Again, if $z=-1$, $a=1+b-c$, the integral in (1) becomes

$$\int_0^1 t^{b-1}(1-t^2)^{c-b-1}\,dt,$$

hich can be evaluated in terms of gamma functions. This sug- ests that probably the sum of the series $F(b,\ 1+b-c;\ c;\ -1)$ can e found.

Finally, if $b=1-a$, $z=\frac{1}{2}$, we are led to the integral

$$\int_0^1 (2t-t^2)^{-a}(1-t)^{c-b-1}\,dt,$$

* See also Barnes 1 where another method is used to obtain this formula. The ethod is reproduced in Whittaker and Watson, *Modern Analysis* (ed. 4, 1927), 14.53.

and, taking $(1-t)^2$ as the new variable, this becomes a beta function. We can thus evaluate $F(a,1-a;c;\frac{1}{2})$ in terms of gamma functions. The actual formulae will be given in Chapter II.

1.6. Barnes' contour integral for *F(a, b; c; z)*.* Consider the contour integral

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{\Gamma(a+s)\,\Gamma(b+s)\,\Gamma(-s)}{\Gamma(c+s)}(-z)^s\,ds,$$

where $|\arg(-z)|<\pi$, and the path of integration is curved, if necessary, to separate the poles $s=-a-n$, $s=-b-n$, $(n=0,\ 1,\ 2,\ \ldots)$ from the poles $s=0,\ 1,\ 2,\ \ldots$. This contour can always be drawn if a and b are not negative integers, as then none of the decreasing sequences of poles coincides with one of the increasing sequence.

Now,† if $|\arg(s+a)|\leqslant\pi-\delta$, $|\arg s|\leqslant\pi-\delta$, then

$$\log\Gamma(s+a)=(s+a-\tfrac{1}{2})\log s-s+\tfrac{1}{2}\log(2\pi)+o(1),$$

when $|s|\to\infty$.

Thus, the integrand, which can be written

$$-\frac{\Gamma(a+s)\,\Gamma(b+s)}{\Gamma(c+s)\,\Gamma(1+s)}\frac{\pi(-z)^s}{\sin s\pi},$$

is asymptotically equal to

$$-\frac{\pi(-z)^s}{\sin s\pi}\exp\,[(a+b-c-1)\log s].$$

Putting $s=iv$ on the contour, we see that, for large values of v, the integrand is

$$O\,[v^{a+b-c-1}\exp\{-v\arg(-z)-\pi\,|\,v\,|\}].$$

Thus the integral is an analytic function of z throughout the domain $|\arg(-z)|\leqslant\pi-\delta$, where δ is any positive number.

Now let C denote the semi-circle of radius $N+\frac{1}{2}$ on the right of the imaginary axis with centre at the origin, N being an integer. As before the integrand is

$$O\,(N^{a+b-c-1})\frac{(-z)^s}{\sin s\pi}$$

for large values of N, the implied constant being independent of $\arg s$ when s is on the semi-circle.

* Barnes 1. † Whittaker and Watson, *Modern Analysis*, § 13.6.

If $s=(N+\frac{1}{2})e^{i\theta}$ and $|z|<1$, we have

$$\frac{(-z)^s}{\sin s\pi}=O\left[\exp\{(N+\tfrac{1}{2})\cos\theta\log|z|-(N+\tfrac{1}{2})\sin\theta\arg(-z)\right.$$
$$\left.-(N+\tfrac{1}{2})\pi\,|\sin\theta\,|\}\right]$$
$$=O\left[\exp\{(N+\tfrac{1}{2})\cos\theta\log|z|-(N+\tfrac{1}{2})\delta\,|\sin\theta\,|\}\right],$$

nd this is

$$O\left[\exp\{2^{-\frac{1}{2}}(N+\tfrac{1}{2})\log|z|\}\right]\quad\text{if}\quad 0\leqslant|\theta|\leqslant\tfrac{1}{4}\pi,$$

nd $\quad O\left[\exp\{-2^{-\frac{1}{2}}\delta(N+\frac{1}{2})\}\right]\quad$ if $\quad\frac{1}{4}\pi\leqslant|\theta|\leqslant\frac{1}{2}\pi.$

Hence, if $\log|z|$ is negative, that is if $|z|<1$, the integrand nds to zero sufficiently rapidly to ensure that $\int_C\to 0$ as $N\to\infty$.

By using Cauchy's theorem for the contour formed by C and e part of the imaginary axis from $i(N+\frac{1}{2})$ to $-i(N+\frac{1}{2})$, and en making $N\to\infty$, we see that, when $|\arg(-z)|\leqslant\pi-\delta$ and $|<1$,

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{\Gamma(a+s)\,\Gamma(b+s)\,\Gamma(-s)}{\Gamma(c+s)}(-z)^s\,ds$$
$$=\lim_{N\to\infty}\sum_{n=0}^{N}\frac{\Gamma(a+n)\,\Gamma(b+n)}{n!\,\Gamma(c+n)}z^n,$$

nce $\Gamma(-s)$ has a simple pole at $s=n$, $(n=0,1,2,\ldots)$ with residue $-1)^{n-1}/n!$. Thus the integral represents an analytic function in e region $|\arg(-z)|<\pi$, and when $|z|<1$ this analytic function ay be represented by the series

$$\frac{\Gamma(a)\,\Gamma(b)}{\Gamma(c)}F(a,b;c;z).$$

The symbol $F(a,b;c;z)$ may therefore be used to denote the ore general function defined by the integral when divided by $(a)\,\Gamma(b)/\Gamma(c)$.

1.7. Barnes' lemma.* *If the path of integration is curved so as separate the increasing and decreasing sequences of poles, then*

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\Gamma(\alpha+s)\,\Gamma(\beta+s)\,\Gamma(\gamma-s)\,\Gamma(\delta-s)\,ds$$
$$=\frac{\Gamma(\alpha+\gamma)\,\Gamma(\alpha+\delta)\,\Gamma(\beta+\gamma)\,\Gamma(\beta+\delta)}{\Gamma(\alpha+\beta+\gamma+\delta)}.$$

* Barnes 1.

Write I for the expression on the left. Let C be the semi-circle of radius ρ on the right of the imaginary axis with its centre at the origin, and suppose that $\rho\to\infty$ in such a way that the lower bound of the distance of C from the poles of $\Gamma(\gamma-s)\,\Gamma(\delta-s)$ is definitely positive. Then

$$\Gamma(\alpha+s)\,\Gamma(\beta+s)\,\Gamma(\gamma-s)\,\Gamma(\delta-s)$$
$$=\frac{\Gamma(\alpha+s)\,\Gamma(\beta+s)}{\Gamma(1-\gamma+s)\,\Gamma(1-\delta+s)}\cdot\frac{\pi^2}{\sin(\gamma-s)\,\pi\sin(\delta-s)\,\pi}$$
$$=O\left[s^{\alpha+\beta+\gamma+\delta-2}\exp\{-2\pi\,|\,I(s)\,|\}\right]$$

as $|s|\to\infty$ on the imaginary axis or on C. Thus the original integral converges, and the integral round C tends to zero as $\rho\to\infty$ when $R(\alpha+\beta+\gamma+\delta-1)<0$. The integral is therefore equal to minus $2\pi i$ times the sum of the residues of the integrand at the poles on the right of the contour. Thus

$$I=\sum_{n=0}^{\infty}\Gamma(\alpha+\gamma+n)\,\Gamma(\beta+\gamma+n)\,\Gamma(\delta-\gamma-n)\,(-1)^n/n!$$
$$+\sum_{n=0}^{\infty}\Gamma(\alpha+\delta+n)\,\Gamma(\beta+\delta+n)\,\Gamma(\gamma-\delta-n)\,(-1)^n/n!$$
$$=\Gamma(\alpha+\gamma)\,\Gamma(\beta+\gamma)\,\Gamma(\delta-\gamma)\,F(\alpha+\gamma,\,\beta+\gamma;\,1+\gamma-\delta;\,1)$$

+ a similar expression with γ and δ interchanged.

Using Gauss's theorem we obtain the required result after a little reduction. The formula has been proved only when

$$R(\alpha+\beta+\gamma+\delta-1)<0,$$

but by the theory of analytic continuation it is true for all values of α, β, γ, δ for which none of the poles of $\Gamma(\alpha+s)\,\Gamma(\beta+s)$ coincide with any of the poles of $\Gamma(\gamma-s)\,\Gamma(\delta-s)$.

By writing $s-k$, $\alpha+k$, $\beta+k$, $\gamma-k$, $\delta-k$ for s, α, β, γ, δ, we see that the result is still true when the limits of integration are $k\pm i\infty$, where k is any real constant.

CHAPTER VI

THODS OF OBTAINING TRANSFORMATIONS OF HYPERGEOMETRIC SERIES; (3) BY BARNES' CONTOUR INTEGRALS

. **Introductory remarks.** In Chapter V we saw how sformations of non-terminating series can sometimes be ved by a use of Carlson's theorem, and in § 4.4 some trans- ations of such series were obtained by a limiting process from sformations connecting terminating series of higher orders. is chapter a direct method* is given in which free use is made ntour integrals of Barnes' type.

. **Barnes' second lemma.** We now prove the formula†

$$\frac{1}{2\pi i}\int \frac{\Gamma(\alpha_1+s)\,\Gamma(\alpha_2+s)\,\Gamma(\alpha_3+s)\,\Gamma(1-\beta_1-s)\,\Gamma(-s)\,ds}{\Gamma(\beta_2+s)}$$
$$=\frac{\Gamma(\alpha_1)\,\Gamma(\alpha_2)\,\Gamma(\alpha_3)\,\Gamma(1-\beta_1+\alpha_1)\,\Gamma(1-\beta_1+\alpha_2)\,\Gamma(1-\beta_1+\alpha_3)}{\Gamma(\beta_2-\alpha_1)\,\Gamma(\beta_2-\alpha_2)\,\Gamma(\beta_2-\alpha_3)}$$

ided that $\beta_1+\beta_2=\alpha_1+\alpha_2+\alpha_3+1$. The path of integration is e parallel to the imaginary axis except that it is curved, if ssary, so that the decreasing sequences of poles lie to the left, the increasing sequences of poles to the right of the contour.‡ he integrals in this chapter are of this type.

y Barnes' lemma (§ 1.7) we have

$$\frac{1}{2\pi i}\int \Gamma(\alpha_1+s)\,\Gamma(\alpha_2+s)\,\Gamma(n-s)\,\Gamma(\beta_1-\alpha_1-\alpha_2-s)\,ds$$
$$=\frac{\Gamma(\alpha_1+n)\,\Gamma(\alpha_2+n)\,\Gamma(\beta_1-\alpha_1)\,\Gamma(\beta_1-\alpha_2)}{\Gamma(\beta_1+n)}.$$

Bailey **8**.
Barnes **2**.
The integral is taken from $c-i\infty$ to $c+i\infty$. In some papers the integrals are in the opposite direction, and so variations in sign occur.

Thus

$${}_3F_2\left[\begin{matrix}\alpha_1,\alpha_2,\alpha_3;\\ \beta_1,\beta_2\end{matrix}\right]=\frac{\Gamma(\beta_1)}{\Gamma(\beta_1-\alpha_1)\,\Gamma(\beta_1-\alpha_2)\,\Gamma(\alpha_1)\,\Gamma(\alpha_2)}$$
$$\times\sum_{n=0}^{\infty}\frac{1}{2\pi i}\int\frac{(\alpha_3)_n}{n!\,(\beta_2)_n}\Gamma(\alpha_1+s)\,\Gamma(\alpha_2+s)\,\Gamma(n-s)\,\Gamma(\beta_1-\alpha_1-\alpha_2-s)\,ds$$
$$=\frac{\Gamma(\beta_1)}{\Gamma(\beta_1-\alpha_1)\,\Gamma(\beta_1-\alpha_2)\,\Gamma(\alpha_1)\,\Gamma(\alpha_2)}$$
$$\times\frac{1}{2\pi i}\int\Gamma(\alpha_1+s)\,\Gamma(\alpha_2+s)\,\Gamma(\beta_1-\alpha_1-\alpha_2-s)\,\Gamma(-s)\,{}_2F_1\left[\begin{matrix}\alpha_3,-s;\\ \beta_2\end{matrix}\right]ds$$

and so

$$(2)\quad {}_3F_2\left[\begin{matrix}\alpha_1,\alpha_2,\alpha_3;\\ \beta_1,\beta_2\end{matrix}\right]=\frac{\Gamma(\beta_1)\,\Gamma(\beta_2)}{\Gamma(\beta_1-\alpha_1)\,\Gamma(\beta_1-\alpha_2)\,\Gamma(\beta_2-\alpha_3)\,\Gamma(\alpha_1)\,\Gamma(\alpha_2)}$$
$$\times\frac{1}{2\pi i}\int\frac{\Gamma(\alpha_1+s)\,\Gamma(\alpha_2+s)\,\Gamma(\beta_1-\alpha_1-\alpha_2-s)\,\Gamma(\beta_2-\alpha_3+s)\,\Gamma(-s)\,ds}{\Gamma(\beta_2+s)}.$$

The interchange in the order of summation and integration can easily be justified if $R(\beta_2-\alpha_3+s)>0$. Now take $\beta_1=\alpha_3$; the series on the left can be summed by Gauss's theorem, and the lemma is proved.

If the integral in (1) is evaluated in terms of hypergeometric series by considering the residues at poles on the right of the contour, we obtain a relation which reduces to Saalschütz's theorem when one of the parameters α_1, α_2, α_3 is a negative integer.*

6.3. Integrals representing well-poised series. From Barnes' second lemma it is easily verified that

$$\frac{\Gamma(\alpha_1+n)\,\Gamma(\alpha_2+n)\,\Gamma(\alpha_3+n)}{\Gamma(\kappa-\alpha_1+n)\,\Gamma(\kappa-\alpha_2+n)\,\Gamma(\kappa-\alpha_3+n)}$$
$$=\frac{1}{\Gamma(\kappa-\alpha_2-\alpha_3)\,\Gamma(\kappa-\alpha_3-\alpha_1)\,\Gamma(\kappa-\alpha_1-\alpha_2)}$$
$$\times\frac{1}{2\pi i}\int\frac{\Gamma(\alpha_1+s)\Gamma(\alpha_2+s)\Gamma(\alpha_3+s)\Gamma(\kappa-\alpha_1-\alpha_2-\alpha_3-s)\Gamma(n-s)\,ds}{\Gamma(\kappa+n+s)}.$$

* The relation similarly obtained from (2) is equivalent to § 3.8 (1).

Asymptotic expansions and analytic continuations for a class of Barnes-integrals

by

B. L. J. Braaksma

§ 1. Introduction

§ 1.1. Asymptotic expansions for $|z| \to \infty$ and analytic continuations will be derived for the function $H(z)$ defined by

$$H(z) = \frac{1}{2\pi i}\int_C h(s) z^s ds \tag{1.1}$$

where z is not equal to zero and

$$z^s = \exp\{s(\text{Log}\,|z| + i\arg z)\} \tag{1.2}$$

in which $\text{Log}\,|z|$ denotes the natural logarithm of $|z|$ and $\arg z$ is not necessarily the principal value. Further

$$h(s) = \frac{\prod_1^n \Gamma(1-a_j+\alpha_j s) \prod_1^m \Gamma(b_j-\beta_j s)}{\prod_{m+1}^q \Gamma(1-b_j+\beta_j s) \prod_{n+1}^p \Gamma(a_j-\alpha_j s)}, \tag{1.3}$$

where p, q, n, m are integers satisfying

$$0 \leqq n \leqq p,\ 1 \leqq m \leqq q, \tag{1.4}$$

$\alpha_j (j = 1, \ldots, p)$, $\beta_j (j = 1, \ldots, q)$ are positive numbers and $a_j (j = 1, \ldots, p)$, $b_j (j = 1, \ldots, q)$ are complex numbers such that

$$\alpha_j(b_h+\nu) \neq \beta_h(a_j-1-\lambda) \text{ for } \nu, \lambda = 0, 1, \ldots; \quad h = 1, \ldots, m;\ j = 1, \ldots, n. \tag{1.5}$$

C is a contour in the complex s-plane such that the points

$$s = (b_j+\nu)/\beta_j \qquad (j = 1, \ldots, m;\ \nu = 0, 1, \ldots) \tag{1.6}$$

resp.

$$s = (a_j-1-\nu)/\alpha_j \qquad (j = 1, \ldots, n;\ \nu = 0, 1, \ldots) \tag{1.7}$$

lie to the right resp. left of C, while further C runs from $s = \infty - ik$ to $s = \infty + ik$. Here k is a constant with $k > |\text{Im}\, b_j|/\beta_j$ $(j = 1, \ldots, m)$. The conditions for the contour C can be fulfilled on account of (1.5). Contours like C and also contours like C but with endpoints $s = -i\,\infty+\sigma$ and $s = i\infty+\sigma$ (σ real) instead of $s = \infty-ik$ and $s = \infty+ik$ are called Barnes-contours and the corresponding integrals are called Barnes-integrals.

In the following we always assume (1.4) and (1.5).

In § 6.1, theorem 1, we show that $H(z)$ makes sense in the following cases:

I. for every $z \neq 0$ if μ is positive where

$$\mu = \sum_1^q \beta_j - \sum_1^p \alpha_j, \tag{1.8}$$

II. if $\mu = 0$ and

$$0 < |z| < \beta^{-1} \tag{1.9}$$

where

$$\beta = \prod_1^p \alpha_j^{\alpha_j} \prod_1^q \beta_j^{-\beta_j}. \tag{1.10}$$

In general $H(z)$ is a multiple-valued function of z.

§ **1.2.** The function $H(z)$ and special cases of it occur at various places in the literature. A first systematic study of the function $H(z)$ has been made in a recent paper by C. Fox [18]. [1]) In the case that some special relations between the constants α_j, β_j, a_j, b_j are satisfied Fox derives theorems about $H(z)$ as a symmetrical Fourier kernel and a theorem about the asymptotic behaviour of $H(z)$ for $z \to \infty$ and $z > 0$.

The function defined by (1.1) but with the contour C replaced by another contour C' has been considered by A. L. Dixon and W. L. Ferrar [12]. C' is a contour like C but with endpoints $s = -\infty\, i+\sigma$ and $s = \infty\, i+\sigma$ (σ real). Their investigation concerns the convergence of the integrals, discontinuities and analytic continuations (not for all values of z) and integrals in whose integrands the function defined by (1.1) with $C \,||\, C'$ ($||$ means: replaced by) occurs.

Special cases of the function $H(z)$ occur in papers on functional equations with multiple gamma-factors and on the average order of arithmetical functions by S. Bochner [5], [5a], [6] and K. Chandrasekharan and Raghavan Narasimhan [9]. In these papers in some cases the analytic continuation resp. an estimation for $H(z)$ has been derived.

A large number of special functions are special cases of the

[1]) Numbers between brackets refer to the bibliography. In each paragraph the footnotes are numbered anew.

function $H(z)$. In the first place the G-function and all special cases of it as for instance Bessel-, Legendre-, Whittaker-, Struve-functions, the ordinary generalized hypergeometric functions (cf. [15] pp. 216-222) and a function considered by J. Boersma [7]. The G-function is the special case of the function $H(z)$ in (1.1) with $\alpha_j = 1$ $(j = 1, \ldots, p)$, $\beta_j = 1$ $(j = 1, \ldots, q)$. The ordinary generalized hypergeometric function is a special case of the G-function with $m = 1$, $n = p$ among others.

Further $H(z)$ contains as special cases the function of G. Mittag-Leffler (cf. G. Sansone-J. C. H. Gerretsen [28], p. 345), the generalized Bessel-function considered by E. M. Wright [31], [35] and the generalization of the hypergeometric function considered by C. Fox [17] and E. M. Wright [32], [34].

The results about the function $H(z)$ which will be derived here contain the asymptotic expansions for $|z| \to \infty$ and the analytic continuations of the functions mentioned above. The latter expansions and continuations have been derived by various methods among others by E. W. Barnes [2], [3], [4] (cf. a correction in F. W. J. Olver [25]), G. N. Watson [29], D. Wrinch [38], [39], [40], C. Fox [17], [18], W. B. Ford [16], E. M. Wright [31]–[36], C. V. Newsom [23], [24], H. K. Hughes [19], [20], T. M. MacRobert [21], C. S. Meijer [22] and J. Boersma [7]. The most important papers in this connection are those of Barnes, Wright and Meijer.

In [3] Barnes considered the asymptotic expansion of a number of G-functions. In the first place he derived algebraic asymptotic expansions (cf. § 4.6) for a class of G-functions. These expansions are derived by means of a simple method involving Barnes-integrals and the theorem of residues. In the second place he derived exponentially small and exponentially infinite asymptotic expansions (cf. § 4.4 for a definition) for another G-function. The derivation of these expansions is difficult and complicated. The G-function is written as a suitable exponential function multiplied by a contour integral. The integrand in this integral is a series of which the analytic continuation and the residues in the singular points are derived by means of an ingenious, complicated method involving among others zeta-functions and other functions considered previously by Barnes. The contour integral mentioned before has an algebraic asymptotic expansion which can be deduced by means of the theorem of residues. The investigation in [3] yields among others the asymptotic expansions of the ordinary generalized hypergeometric functions. Barnes also obtained the analytic continuation of a special case of the G-function by means of Barnes-integrals (cf. [4]).

In [22] Meijer has derived all asymptotic expansions and analytic continuations of the G-function. The method depends upon the fact that the G-function satisfies a homogeneous linear differential equation of which the G-functions considered by Barnes constitute a fundamental system of solutions. So every G-function can be expressed linearly in these G-functions of Barnes and from the asymptotic behaviour of these functions the asymptotic behaviour of arbitrary G-functions can be derived.

In [31], [32], [34] and [35] Wright considered the asymptotic expansions of generalizations of Bessel- and hypergeometric functions. The majority of his results are derived by a method which is based on the theorem of Cauchy and an adapted and simplified version of the method of steepest descents. In [33] and [36] these methods are applied to a class of more general integral functions. However, these methods do not yield all asymptotic expansions: any exponentially small asymptotic expansion has to be established by different methods (cf. [32], [34], [35]). The results of Wright have as an advantage over the results of the other authors mentioned before that his asymptotic expansions hold uniformly on sectors which cover the entire z-plane. Further the results of Wright — and also those of H. K. Hughes [19] — contain more information about the coefficients occurring in the asymptotic expansions.

§ 1.3. A description of the methods which we use to obtain the asymptotic expansions and analytic continuations of the function $H(z)$ is given in § 2. The results cannot be derived in the same manner as in the case of the G-function in [22] because in general the functions $H(z)$ do not satisfy expansion-formulae which express $H(z)$ in terms of some special functions $H(z)$ (if this would be the case then we should have to consider in detail only these latter functions as in the case of the G-function).

The analytic continuations of $H(z)$ in the case $\mu = 0$ can be found by bending parallel to the imaginary axis the contour in the integral (1.1) in which from the integrand some suitable analytic functions have been subtracted. This method is an extension of the method of Barnes in [4].

This method can be applied also in a number of cases to the determination of the asymptotic expansion of $H(z)$ for $|z| \to \infty$ if $\mu > 0$. Then algebraic asymptotic expansions are obtained.

However, in some cases all coefficients in these expansions are equal to zero ("dummy" expansions) and in these cases $H(z)$ has an exponentially small asymptotic expansion. This expansion will be derived by approximating the integrand in (1.1) by means of lemma 3 in § 3.3. In this way the difficulties in the researches of Barnes and Wright (cf. [3], [34], [35]) about special cases of these expansions are avoided. Contrary to their proofs here the derivation of the exponentially small expansions is easier than the derivation of the exponentially infinite expansions.

The remaining asymptotic expansions of $H(z)$ in the case $\mu > 0$ are derived by splitting in (1.1) the integrand into parts so that in the integral of some of these parts the contour can be bended parallel to the imaginary axis while the integrals of the other parts can be estimated by a method similar to the method which yields the exponentially small expansions. Some aspects of this method have been borrowed from Wright [33].

In the derivation of the asymptotic expansions of $H(z)$ the estimation of the remainder-terms is the most difficult part. The method used here depends upon the lemmas in § 5 which contain analytic continuations and estimates for a class of integrals related to Barnes-integrals. This method is related to the indirect Abelian asymptotics of Laplace transforms.

The remainder-terms can also be estimated by a direct method viz. the method of steepest descents. This will be sketched in § 10. In the case of the exponentially infinite expansions of $H(z)$ this method is analogous to the method of Wright in [33].

The asymptotic expansions of $H(z)$ are given in such a way that given a certain closed sector in the z-plane this sector can be divided into a finite number of closed subsectors on each of which the expansion of $H(z)$ for $|z| \to \infty$ holds uniformly in arg z. Moreover it is indicated how the coefficients in the asymptotic expansions can be found.

§ **1.4.** The results concerning $H(z)$ are contained in theorem 1 in § 6.1 (behaviour near $z = 0$), theorem 2 in § 6.2 (analytic continuations and behaviour near $z = \infty$ in the case $\mu = 0$), theorem 3 in § 6.3 (algebraic behaviour near $z = \infty$ in the case $\mu > 0$), theorem 4 in § 7.3 (exponentially small expansions in the case $\mu > 0$), theorems 5 and 6 in § 9.1 (exponentially infinite expansions in the case $\mu > 0$) and theorems 7–9 (expansions in the remaining barrier-regions for $\mu > 0$). In these theorems the notations introduced in (1.8), (1.10) and the definitions I–IV in § 4 are used. The terminology of asymptotic expansions is given in § 4.4 and § 4.6. In § 9.3 we have given a survey from which one may deduce which theorem contains the asymptotic expansion for $|z| \to \infty$ of a given function $H(z)$ on a given sector. In § 10.2 and § 10.3 some supplements on the theorems in § 6 and § 9 are given.

In § 11 the results about the function $H(z)$ are applied to the G-function: see the theorems 10–17. The asymptotic expansions and analytic continuations given in [22] are derived again. An advantage is that the asymptotic expansions are formulated in such a way that they hold uniformly on closed sectors — also in transitional regions — while moreover the coefficients of the expansions can be found by means of recurrence formulae. The notations used in the theorems and a survey of the theorems are given in § 11.3.

In § 12.1 and § 12.2 the results concerning $H(z)$ are applied to the generalized hypergeometric functions considered by Wright (cf. theorems 18–22). A survey of the theorems and the notations are given at the end of § 12.1 and in § 12.2. In § 12.3 a general class of series which possess exponentially small asymptotic expansions is considered. In § 12.4 the generalized Bessel-function is considered. The results are formulated in the theorems 24–26. The notations used in these theorems are given in (12.45).

§ 2. Description of the methods

§ 2.1. In this section we sketch the method by which the algebraic asymptotic expansions for $|z| \to \infty$ resp. the analytic continuation of the function $H(z)$ in case I resp. II of § 1 will be derived. First we consider the simplest cases which are analogous to the simplest cases considered by Barnes in [3] and [4].

To that end we replace the contour C in (1.1) by two other paths L and L_1. L resp. L_1 runs from $s = w$ to $s = w+il$ resp. $w-il$ and then to $s = \infty+il$ resp. $\infty-il$, while both parts of L resp. L_1 are rectilinear. Here w and l are real numbers so that

$$(2.1) \quad w \neq \operatorname{Re}(a_j-1-\nu)/\alpha_j \qquad (j = 1, \ldots, p;\ \nu = 0, 1, 2, \ldots)$$

$$(2.2) \quad w < \operatorname{Re} b_j/\beta_j \qquad (j = 1, \ldots, m)$$

$$(2.3) \quad \begin{cases} l = 1+\max\{|\operatorname{Im} a_j/\alpha_j|\,(j=1, \ldots, p),\ |\operatorname{Im} b_j/\beta_j|\,(j=1, \ldots, q), \\ |\operatorname{Im} \alpha/\mu|\}\ \text{(cf. (3.24) for } \alpha) \text{ if } \mu \text{ is positive, while for } \mu = 0 \\ l = 1+\max\{|\operatorname{Im} a_j/\alpha_j|\,(j=1, \ldots, p),\ |\operatorname{Im} b_j/\beta_j|\,(j=1, \ldots, q)\}. \end{cases}$$

Then we easily deduce from (1.1) and the theorem of residues that

$$H(z) = Q_w(z) + \frac{1}{2\pi i}\int_L h(s)z^s ds - \frac{1}{2\pi i}\int_{L_1} h(s)z^s ds \tag{2.4}$$

where

$$Q_w(z) = \sum \text{residues of } h(s)z^s \text{ in those points (1.7) where } \operatorname{Re} s > w. \tag{2.5}$$

Formula (2.4) holds in case I and also in case II of § 1. For the consideration of the integrals in (2.4) we need approximations for the function $h(s)$. Therefore we write $h(s)$ as a product of two other functions. Using

$$\Gamma(s) = \pi/\{\sin \pi s \Gamma(1-s)\} \tag{2.6}$$

we see that

$$h(s) = h_0(s)h_1(s) \tag{2.7}$$

where if

$$s \neq (a_j-1-\nu)/\alpha_j \qquad (j = 1, \ldots, p;\ \nu = 0, 1, 2, \ldots) \tag{2.8}$$

resp.

$$s \neq (b_j+\nu)/\beta_j \qquad (j = 1, \ldots, m;\ \nu = 0, \pm 1, \pm 2, \ldots) \tag{2.9}$$

we define

$$h_0(s) = \prod_1^p \Gamma(1-a_j+\alpha_j s)/\prod_1^q \Gamma(1-b_j+\beta_j s) \tag{2.10}$$

resp.

$$h_1(s) = \pi^{m+n-p} \prod_{n+1}^p \sin \pi(a_j-\alpha_j s)/\prod_1^m \sin \pi(b_j-\beta_j s). \tag{2.11}$$

For $h_0(s)$ resp. $h_1(s)$ approximation formulae are formulated in the lemmas 2 and 2a in § 3.2 resp. 4a in § 4.3. From these formulae estimates for $h(s)$ can be obtained.

Now define δ_0 by

$$\delta_0 = (\sum_1^m \beta_j - \sum_{n+1}^p \alpha_j)\pi. \tag{2.12}$$

Consider the case that

$$\delta_0 > \tfrac{1}{2}\mu\pi. \tag{2.13}$$

Then we may derive from the estimates for $h(s)$ mentioned above that the lemmas 6–7a from § 5 can be applied to the integrals in (2.4) for certain values of z; the path of integration L resp. L_1 in

(2.4) may be replaced by the half-line from $s = w$ to $s = w+i\infty$ resp. $w-i\infty$:

$$H(z) = Q_w(z) + \frac{1}{2\pi i}\int_{w-i\infty}^{w+i\infty} h(s)z^s ds \tag{2.14}$$

if $\mu = 0$, (1.9) and

$$|\arg z| < \delta_0 - \tfrac{1}{2}\mu\pi \tag{2.15}$$

holds and also if μ is positive and (2.15) is satisfied. The integral in (2.14) is absolutely convergent if $\mu \geqq 0$ and (2.15) holds. If $\mu = 0$ then $H(z)$ can be continued analytically by means of (2.14) into the domain (2.15). If μ is positive then

$$H(z) = Q_w(z) + O(z^w)$$

for $|z| \to \infty$ uniformly on every closed subsector of (2.15) with the vertex in $z = 0$. Hence by definition IV in § 4.6

$$H(z) \sim Q(z) \tag{2.16}$$

for $|z| \to \infty$ uniformly on every closed subsector of (2.15) with the vertex in $z = 0$, if μ is positive. The asymptotic expansion (2.16) is algebraic.

In the case that $\mu = 0$ another application of the lemmas 6 and 6a shows that the integral in (2.14) — and so $H(z)$ — can be continued analytically for $|z| > \beta^{-1}$.

Next we drop the assumption (2.13) and we extend the method used above to obtain the analytic continuation of $H(z)$ if $\mu = 0$ and the algebraic asymptotic expansion of $H(z)$ for $|z| \to \infty$ if $\mu > 0$ in the general case. Therefore we define (cf. (2.10) for $h_0(s)$):

$$\left\{\begin{array}{l} P_w(z) = \sum \text{ residues of } h_0(s)z^s \text{ in those points } s \text{ for which} \\ \operatorname{Re} s > w \quad \text{as well as } s = (a_j - 1 - \nu)/\alpha_j \ (j = 1, \ldots, p; \\ \nu = 0, 1, 2, \ldots). \end{array}\right. \tag{2.17}$$

Let r be an arbitrary integer and let δ_j, κ, C_j and D_j be given by the definitions I and II in § 4.2. Then it easily follows from (1.1), the theorem of residues, the definition of L, L_1, $Q_w(z)$ and $P_w(z)$ (cf. (2.5) and (2.17)) and (2.7) that

$$\begin{aligned} H(z) = Q_w(z) &+ \sum_r^{\kappa} D_j P_w(ze^{i\delta_j}) - \sum_0^{r-1} C_j P_w(ze^{i\delta_j}) \\ &+ \frac{1}{2\pi i}\left(\int_L - \int_{L_1}\right) h_0(s)\left\{h_1(s) + \sum_r^{\kappa} D_j e^{i\delta_j s} - \sum_0^{r-1} C_j e^{i\delta_j s}\right\} z^s ds \end{aligned} \tag{2.18}$$

in case I and also in case II of § 1. Like in (2.4) we want to stretch the path of integration L and L_1 in (2.18) to the straight line $\operatorname{Re} s = w$. This is possible if $\mu > 0$,

$$(2.19) \qquad \delta_r - \delta_{r-1} > \mu\pi$$

and

$$(2.20) \qquad -\delta_r + \tfrac{1}{2}\mu\pi < \arg z < -\delta_{r-1} - \tfrac{1}{2}\mu\pi$$

hold and also if $\mu = 0$, (1.9), (2.19) and (2.20) hold. The proof depends on the lemmas 6–7a from § 5. The assumptions concerning the integrands in these lemmas can be verified with the help of the estimates in the lemmas 2, 2a and 4a from § 3.2 and § 4.3 for the factors of the integrand in (2.18). Moreover the lemmas 6–7a applied to the integrals in (2.18) furnish the analytic continuation of $H(z)$ into (2.20), if $\mu = 0$ and (2.19) holds, and the algebraic asymptotic expansion of $H(z)$ on subsectors of (2.20) if (2.19) is satisfied and μ is positive. The results are formulated in theorem 2 and theorem 3 in § 6 where also the complete proof is given. The case with (2.13) appears to be contained in theorem 2 (cf. remark 1 after theorem 2) and theorem 3.

§ 2.2. In this section we consider the exponentially small asymptotic expansions of $H(z)$. A condition for the occurrence of these expansions is that $n = 0$. If $n = 0$ then $Q_w(z) \equiv 0$ by (2.5) and $Q(z)$ represents a formal series of zeros (cf. (4.26)). So if $n = 0$, $\mu > 0$ and (2.13) are fulfilled then by (2.14)

$$(2.21) \qquad H(z) = \frac{1}{2\pi i} \int_{w-i\infty}^{w+i\infty} h(s) z^s \, ds$$

on (2.15) where the integral in (2.21) converges absolutely on (2.15), and moreover by (2.16) and the definitions in § 4.6 we have $H(z) = O(z^w)$ for $|z| \to \infty$ uniformly on closed subsectors of (2.15) with the vertex in $z = 0$ and where w is an arbitrary negative number. Hence in this case better estimates for $H(z)$ have to be obtained. It appears that $H(z)$ has an exponentially small asymptotic expansion in this case.

To derive this expansion we first treat the special case that $n = 0$, $m = q$, $\mu > 0$. Then $\delta_0 = \mu\pi$ by (1.8) and (2.12). So the sector (2.15) can be written as

$$(2.22) \qquad |\arg z| < \tfrac{1}{2}\mu\pi.$$

We temporarily denote the function $H(z)$ for which the assumptions above are satisfied by $H_0(z)$. Further we denote $h(s)$ by $h_2(s)$ in this case. So by (1.3), if

$$s \neq (b_j + \nu)/\beta_j \qquad (j = 1, \ldots, q;\ \nu = 0, 1, 2, \ldots) \tag{2.23}$$

then

$$h_2(s) = \prod_1^q \Gamma(b_j - \beta_j s) / \prod_1^p \Gamma(a_j - \alpha_j s). \tag{2.24}$$

Hence if (2.22) is satisfied then by (2.21)

$$H_0(z) = \frac{1}{2\pi i} \int_{w-i\infty}^{w+i\infty} h_2(s) z^s \, ds \tag{2.25}$$

where the integral is absolutely convergent.

To the factor $h_2(s)$ of the integrand in (2.25) we want to apply lemma 3 in § 3.3. Therefore we choose an arbitrary non-negative integer N and next the real number w so that besides (2.1) and (2.2) also

$$w < (1 - \mathrm{Re}\,\alpha - N)/\mu \tag{2.26}$$

is satisfied. Then we derive from lemma 3 and (2.25):

$$\begin{aligned} &H_0(\beta^{-1}\mu^{-\mu} z) \\ &= \sum_0^{N-1} (-1)^j A_j (2\pi)^{q-p-1} \frac{1}{2\pi i} \int_{w-i\infty}^{w+i\infty} \Gamma(1 - \mu s - \alpha - j) z^s \, ds \\ &\quad - i(2\pi)^{q-p-2} \int_{w-i\infty}^{w+i\infty} \rho_N(s) \Gamma(1 - \mu s - \alpha - N) z^s \, ds \end{aligned} \tag{2.27}$$

on (2.22); all integrals in (2.27) converge absolutely (cf. § 7.1 for details of the proof). To the first N integrals in (2.27) we apply (cf. § 7.1)

$$\frac{1}{2\pi i} \int_{w-i\infty}^{w+i\infty} \Gamma(1 - \mu s - \alpha - j) z^s \, ds = \frac{1}{\mu} z^{(1-\alpha-j)/\mu} \exp\left(-z^{1/\mu}\right) \tag{2.28}$$

for $j = 0, \ldots, N-1$ and (2.22). So the first N terms at the righthand side of (2.27) vanish exponentially on closed subsectors of (2.22) with the vertex in $z = 0$.

Next we have to estimate the last integral in (2.27) which we denote by $\sigma(z)$. So if (2.22) is fulfilled

$$\begin{aligned} \sigma(z) &= \int_{w-i\infty}^{w+i\infty} \rho_N(s) \Gamma(1 - \mu s - \alpha - N) z^s \, ds \\ &= \int_{w-i\infty}^{w+i\infty} \rho_N(s) \frac{\Gamma(3 - \mu s - \alpha - N)}{(1 - \mu s - \alpha - N)_2} z^s \, ds \end{aligned} \tag{2.29}$$

(cf. (3.11) for the notation $(\lambda)_2$). These integrals converge absolutely as this is also the case with the integrals in (2.27). We estimate $\sigma(z)$ in a crude manner with the method of indirect Abelian asymptotics (cf. G. Doetsch [13] II p. 41). An alternative approach will be sketched in § 10.1; there we use the method of steepest descents.

Here we start with the formula

$$\Gamma(3-\mu s-\alpha-N)z^s = z^{(3-\alpha-N)/\mu}\int_0^\infty t^{2-\mu s-\alpha-N}\exp(-z^{1/\mu}t)\,dt \tag{2.30}$$

for $\operatorname{Re} s = w$ and (2.22); on account of (2.26) the integral is absolutely convergent. In the last integrand in (2.29) we replace the lefthand side of (2.30) by the righthand side of (2.30) and next we revert the order of integration (justification in § 7.2); then we obtain

$$\sigma(z) = z^{(3-\alpha-N)/\mu}\int_0^\infty \rho(t)\exp(-z^{1/\mu}t)\,dt \tag{2.31}$$

for (2.22) where for $t > 0$:

$$\rho(t) = \int_{w-i\infty}^{w+i\infty} \rho_N(s)t^{2-\mu s-\alpha-N}\,\frac{ds}{(1-\mu s-\alpha-N)_2}. \tag{2.32}$$

So $\sigma(z^\mu)$ and $\rho(t)$ are related to each other by the Laplace transformation. By (3.33)

$$\rho_N(s)/(1-\mu s-\alpha-N)_2 = O(s^{-2}) \tag{2.33}$$

for $|s| \to \infty$ uniformly on $\operatorname{Re} s \leqq w$ (cf. § 7.2 for all details of the proofs of (2.33)–(2.36)). Then it is easy to deduce that

$$\begin{aligned} |\rho(t)| &\leqq \int_{w-i\infty}^{w+i\infty} |\rho_N(s)t^{2-\mu s-\alpha-N}/(1-\mu s-\alpha-N)_2|\cdot|ds| \\ &\leqq Kt^{2-\mu w-\operatorname{Re}\alpha-N} \end{aligned} \tag{2.34}$$

for $t > 0$ and some constant K independent of t. Further it appears that $\rho(t) = 0$ for $0 < t \leqq 1$. From this, (2.34) and (2.31) we derive

$$\sigma(z) = z^{(3-\alpha-N)/\mu}\exp(-z^{1/\mu})O(1) \tag{2.35}$$

for $|z| \to \infty$ uniformly on every closed sector with vertex $z = 0$ which is contained in (2.22). From (2.27), (2.28), (2.29) and (2.35) we may derive

$$H_0(z) = (2\pi)^{q-p}e^{(\pm\alpha-\frac{1}{2})\pi i}E_N(ze^{\pm\mu\pi i}) \tag{2.36}$$

for $|z| \to \infty$ uniformly on every closed sector with vertex in $z = 0$ and which is contained in (2.22). Here N is an arbitrary nonnegative integer, the lower resp. upper signs belong together and $E_N(z)$ is defined in definition III in § 4.4. From (2.36) we immediately derive the exponentially small asymptotic expansions of $H(z)$ (or $H_0(z)$) in the case considered above. This will be formulated in theorem 4 in § 7.3.

Now we consider the case that $\mu > 0$, $n = 0$, $0 < m < q$ and (2.13) hold. Then by (2.6), (2.24) and (1.3)

$$h(s) = h_2(s)\pi^{m-q}\prod_{m+1}^{q}\sin\pi(b_j-\beta_j s) \tag{2.37}$$

if (2.23) is fulfilled. The factor of $h_2(s)$ in (2.37) satisfies

$$\pi^{m-q}\prod_{m+1}^{q}\sin\pi(b_j-\beta_j s) = \sum_{0}^{M}\tau_j e^{i\omega_j s} \tag{2.38}$$

where M is a positive integer, $\omega_0, \ldots, \omega_M$ are real and independent of s with

$$\omega_0 < \omega_1 < \ldots < \omega_M,\ \ \omega_M = \pi\sum_{m+1}^{q}\beta_j = \mu\pi-\delta_0 = -\omega_0 \tag{2.39}$$

(cf. (1.8) and (2.12)) while $\tau_0, \ldots, \tau_M$ are complex and independent of s with

$$\begin{aligned}\tau_0 &= (2\pi i)^{m-q}\exp\,(\pi i\sum_{m+1}^{q}b_j),\\ \tau_M &= (-2\pi i)^{m-q}\exp\,(-\pi i\sum_{m+1}^{q}b_j).\end{aligned} \tag{2.40}$$

By (2.39) we have if (2.15) holds:

$$-\tfrac{1}{2}\mu\pi < \arg z+\delta_0-\mu\pi \leqq \arg z+\omega_j \leqq \arg z+\mu\pi-\delta_0 < \tfrac{1}{2}\mu\pi \tag{2.41}$$

for $j = 0, \ldots, M$. Since further (2.25) holds for (2.22), now also (2.25) with $z \,||\, ze^{i\omega_j}$ is valid on (2.15) by (2.41). From this, (2.21), (2.37) and (2.38) we deduce

$$H(z) = \sum_{0}^{M}\tau_j\, H_0(ze^{i\omega_j}). \tag{2.42}$$

This implies on account of (2.36) and (2.41)

$$H(z) = (2\pi)^{q-p}\sum_{0}^{M}\tau_j e^{(\pm\alpha-\frac{1}{2})\pi i}E_N(ze^{i(\omega_j\pm\mu\pi)}) \tag{2.43}$$

for $|z| \to \infty$ uniformly on closed subsectors of (2.15) with vertex $z = 0$ and for every non-negative integer N. In (2.43) the upper resp. lower signs in the products belong together but for different values of j the terms may be taken with different signs. With the help of lemma 5 from § 4.5 we can now derive the asymptotic expansion of $H(z)$ in the case we consider here. These expansions which are again exponentially small are given in theorem 4 in § 7.3.

§ 2.3. We have to consider now the methods which can be used to obtain the asymptotic expansions of $H(z)$ which are not algebraic and not exponentially small in the case $\mu > 0$. Therefore we consider a sector

$$\varepsilon_0 - \tfrac{1}{2}(\delta_r + \delta_{r+1}) \leqq \arg z \leqq \varepsilon_0 - \tfrac{1}{2}(\delta_{r-1} + \delta_r) \tag{2.44}$$

where r is an integer, δ_j is defined in definition I in § 4.2 and ε_0 is a positive number independent of z.

Let N be a non-negative integer and w a real number satisfying (2.1), (2.2), (2.26) and

$$w \neq -(\nu + \mathrm{Re}\,\alpha)/\mu \qquad (\nu = 0, \pm 1, \pm 2, \ldots) \tag{2.45}$$

while l is defined by (2.3). Then we have to approximate the integrals in (2.4) on the sector (2.44). This will be done by using (2.7) for $h(s)$ and approximating $h_1(s)$. However, in the general case it appears that we have to use different approximations for $h_1(s)$ on L and on L_1 contrary to the case where (2.19) holds and where we could use the same approximation on L and on L_1 (cf. § 2.1: (2.18)). Here we introduce integers λ and ν so that

$$\begin{cases} \delta_{\nu+1} \geqq \tfrac{1}{2}(\mu\pi + \delta_r + \delta_{r+1}),\ \delta_{\lambda-1} \leqq \tfrac{1}{2}(-\mu\pi + \delta_r + \delta_{r-1}) - 2\varepsilon_0, \\ \lambda \leqq 0 \leqq \nu,\ \lambda \leqq \kappa \leqq \nu,\ \lambda < r \leqq \nu. \end{cases} \tag{2.46}$$

Here κ is given by definition II in § 4.2 and r and ε_0 are the same as in (2.44). Then we may deduce from lemma 7 and lemma 7a from § 5, (4.8), (4.9) and lemma 2 in § 3.2 that

$$\begin{cases} \displaystyle\int_L h_0(s)\{h_1(s) + \sum_{\nu+1}^{\kappa} D_j e^{i\delta_j s} - \sum_0^{\nu} C_j e^{i\delta_j s}\} z^s ds = O(z^w) \\ \displaystyle\int_{L_1} h_0(s)\{h_1(s) + \sum_{\lambda}^{\kappa} D_j e^{i\delta_j s} - \sum_0^{\lambda-1} C_j e^{i\delta_j s}\} z^s ds = O(z^w) \end{cases} \tag{2.47}$$

for $|z| \to \infty$ uniformly on (2.44).

Now define for $z \neq 0$ and $\mu > 0$:

$$F(z) = \frac{1}{2\pi i}\int_L h_0(s) z^s ds. \tag{2.48}$$

Then by (2.17) and the definition of L and L_1:

$$F(z) = -P_w(z) + \frac{1}{2\pi i}\int_{L_1} h_0(s)z^s ds. \tag{2.49}$$

From (2.4) and (2.47)–(2.49) we may deduce

$$H(z) = Q_w(z) + \sum_{\lambda}^{\kappa} D_j P_w(ze^{i\vartheta_j}) + \sum_{\lambda}^{\nu} (C_j+D_j)F(ze^{i\vartheta_j}) + O(z^w) \tag{2.50}$$

for $|z| \to \infty$ uniformly on (2.44). Hence it is sufficient to derive estimates for $F(z)$ for then we deduce by means of (2.50) estimates for $H(z)$.

To derive the estimates for $F(z)$ we choose a constant ε such that $0 < \varepsilon < \frac{1}{2}\mu\pi$. Then by (2.48), lemma 7 in § 5.2 and lemma 2 in § 3.2 we have

$$F(z) = O(z^w) \tag{2.51}$$

for $|z| \to \infty$ uniformly on (5.14). In the same way using (2.49) and lemma 7a in § 5.3 we obtain

$$F(z) = -P_w(z) + O(z^w) \tag{2.52}$$

for $|z| \to \infty$ uniformly on (5.29).

For the consideration of $F(z)$ on the sector

$$|\arg z| \leqq \tfrac{1}{2}\mu\pi + \varepsilon \tag{2.53}$$

we use the property

(2.54) $|e^{\mp\mu\pi i s}/\sin \pi(\mu s+\alpha)|$ is bounded for $\pm$ Im $s \geqq l$,

where the upper resp. lower signs belong together. Using lemma 7 from § 5.2, the property (2.54) and lemma 2 from § 3.2 we may deduce

$$\int_L h_0(s)z^s e^{\mu\pi i s} \frac{ds}{\sin \pi(\mu s+\alpha)} = O(z^w) \tag{2.55}$$

and

$$\int_{L_1} h_0(s)z^s e^{-\mu\pi i s} \frac{ds}{\sin \pi(\mu s+\alpha)} = O(z^w) \tag{2.56}$$

for $|z| \to \infty$ uniformly on (2.53). In view of

$$\frac{1}{2i}\,(e^{\pi i(\mu s+\alpha)} - e^{-\pi i(\mu s+\alpha)})/\sin \pi(\mu s+\alpha) = 1, \tag{2.57}$$

the definition of $F(z)$ in (2.48) and (2.55), (2.56) imply

$$(2.58)\quad F(z)=\frac{1}{4\pi}e^{-\pi i\alpha}\left(\int_L-\int_{L_1}\right)h_0(s)(ze^{-\mu\pi i})^s\frac{ds}{\sin\pi(\mu s+\alpha)}+O(z^w)$$

for $|z|\to\infty$ uniformly on (2.53). By (3.25) and (2.6) we may write instead of (2.58)

$$(2.59)\quad F(\beta^{-1}\mu^{-\mu}z)$$
$$=\frac{1}{4\pi^2}e^{-\pi i\alpha}\sum_0^{N-1}(-1)^jA_j\left(\int_L-\int_{L_1}\right)\Gamma(1-\mu s-\alpha-j)(ze^{-\mu\pi i})^s ds$$
$$+\frac{1}{4\pi^2}e^{-\pi i\alpha}\tau(z)+O(z^w)$$

for $|z|\to\infty$ uniformly on (2.53) where

$$(2.60)\quad \tau(z)=\pi\left(\int_L-\int_{L_1}\right)r_N(s)(ze^{-\mu\pi i})^s\frac{ds}{\sin\pi(\mu s+\alpha)\Gamma(\mu s+\alpha+N)}$$
$$=(-1)^N\left(\int_L-\int_{L_1}\right)r_N(s)\Gamma(1-\mu s-\alpha-N)(ze^{-\mu\pi i})^s ds.$$

Using (2.59) and (7.1) we infer

$$(2.61)\qquad F(\beta^{-1}\mu^{-\mu}z)=\frac{1}{2\pi i\mu}\sum_0^{N-1}A_jz^{(1-\alpha-j)/\mu}\exp(z^{1/\mu})$$
$$+\frac{1}{4\pi^2}e^{-\pi i\alpha}\tau(z)+O(z^w)$$

for $|z|\to\infty$ uniformly on (2.53). So we have to estimate the analytic function $\tau(z)$ on (2.53).

From (2.60), (3.27), (2.54) and the lemmas 7 and 7a from § 5 we deduce that if $\arg z=\frac{1}{2}\mu\pi+\varepsilon$

$$(2.62)\quad \tau(z)=(-1)^N\int_{w-i\infty}^{w+i\infty}r_N(s)\frac{\Gamma(3-\mu s-\alpha-N)}{(1-\mu s-\alpha-N)_2}(ze^{-\mu\pi i})^s ds.$$

The last integral is of the same type as that in (2.29); the only difference is that almost all poles of $r_N(s)$ are lying to the left of $\mathrm{Re}\, s=w$ while all poles of $\rho_N(s)$ are lying to the right of $\mathrm{Re}\, s=w$. The integral in (2.62) can be rewritten using (2.30) as a multiple integral; in this multiple integral we revert the order of integration like at (2.29) and (2.31). Then we obtain for $\arg z=\frac{1}{2}\mu\pi+\varepsilon$

$$(2.63)\qquad \tau(z)=(ze^{-\mu\pi i})^{(3-\alpha-N)/\mu}\int_0^\infty r(t)\exp(z^{1/\mu}t)\,dt$$

where for $t>0$

$$(2.64)\quad r(t) = (-1)^N \int_{w-i\infty}^{w+i\infty} r_N(s) t^{2-\mu s-\alpha-N} \frac{ds}{(1-\mu s-\alpha-N)_2}.$$

In view of (3.27) we have

$$(2.65)\qquad r_N(s)/(1-\mu s-\alpha-N)_2 = O(s^{-2})$$

for $|s| \to \infty$ uniformly on $\operatorname{Re} s \geqq w$. So the integral in (2.64) converges absolutely for $t > 0$ and

$$(2.66)\qquad |r(t)| \leqq K|t|^{2-\mu w-\operatorname{Re}\alpha-N}$$

for $t > 0$; here K is independent of t. From lemma 2 and (2.66) we derive that for $t > 1$ the function $r(t)$ is equal to the sum of the residues of the integrand in (2.64) in the poles s with $\operatorname{Re} s > w$ multiplied by $2\pi i(-1)^{N+1}$. The number of these poles is finite and it follows that the function $r(t)$ for $t > 1$ can be continued analytically for $t \neq 0$. It is easy now to estimate the integral in (2.63) with the help of the lemmas 6 and 6a from § 5 and the properties of $r(t)$. The results are formulated in lemma 8 in § 8.

From the properties of $F(z)$ mentioned in lemma 8 and (2.50) we deduce in § 9 the asymptotic expansions of $H(z)$ for $|z| \to \infty$ in the case $\mu > 0$ which are not contained in the theorems 3 and 4 (though theorem 3 can also be deduced from lemma 8 and (2.50) again). In § 8 the details of the proofs of the assertions in § 2.3 are presented.

§ 3. Approximations for quotients of gamma-functions

In this paragraph approximation formulae for the functions $h_0(s)$ defined by (2.10) and $h_2(s)$ defined by (2.24) will be derived. Here and in the following paragraphs we use the notation of § 1.1 and § 2.1. Further Log z always denotes the principal value of $\log z$ and $\sum_{j=k}^{l} \ldots$ is interpreted as zero if $k > l$.

§ 3.1. In this section we derive lemma 1 on which the approximations for $h_0(s)$ and $h_2(s)$ will be based. Lemma 1 will be derived from the formula of Stirling in the following form:

Let a be a complex number, ε a constant satisfying $0 < \varepsilon < \pi$ and M a non-negative integer. Then

$$(3.1)\quad \operatorname{Log} \Gamma(s+a) = (s+a-\tfrac{1}{2}) \operatorname{Log} s - s + \tfrac{1}{2} \operatorname{Log}(2\pi) + \sum_{1}^{M-1} B_{j+1}(a) \cdot \frac{(-1)^{j+1} \cdot s^{-j}}{j(j+1)} + O(s^{-M})$$

for $|s| \to \infty$ uniformly on the sector

2.12 Green's function

11
Green's function

This chapter marks the beginning of a series of chapters dealing with the solution to differential equations of theoretical physics. Very often, these differential equations are *linear*; that is, the "sought after" function $\Psi(x), y(x), \phi(t)$ *et cetera* occur only as a polynomial of degree zero and one, and *not* of any higher degree, such as, for instance, $[y(x)]^2$.

11.1 Elegant way to solve linear differential equations

Green's function present a very elegant way of solving linear differential equations of the form

$$
\begin{aligned}
&\mathscr{L}_x y(x) = f(x), \text{ with the differential operator}\\
\mathscr{L}_x &= a_n(x)\frac{d^n}{dx^n} + a_{n-1}(x)\frac{d^{n-1}}{dx^{n-1}} + \ldots + a_1(x)\frac{d}{dx} + a_0(x)\\
&= \sum_{j=0}^{n} a_j(x)\frac{d^j}{dx^j},
\end{aligned}
\tag{11.1}
$$

where $a_i(x)$, $0 \le i \le n$ are functions of x. The idea is quite straightforward: if we are able to obtain the "inverse" G of the differential operator $\mathscr{L}$ defined by

$$
\mathscr{L}_x G(x,x') = \delta(x-x'), \tag{11.2}
$$

with δ representing Dirac's delta function, then the solution to the inhomogenuous differential equation (11.1) can be obtained by integrating $G(x-x')$ alongside with the inhomogenuous term $f(x')$; that is,

$$
y(x) = \int_{-\infty}^{\infty} G(x,x')f(x')dx'. \tag{11.3}
$$

This claim, as posted in Eq. (11.3), can be verified by explicitly applying the differential operator $\mathscr{L}_x$ to the solution $y(x)$,

$$\begin{aligned} &\mathscr{L}_x y(x) \\ &= \mathscr{L}_x \int_{-\infty}^{\infty} G(x,x') f(x') dx' \\ &= \int_{-\infty}^{\infty} \mathscr{L}_x G(x,x') f(x') dx' \\ &= \int_{-\infty}^{\infty} \delta(x-x') f(x') dx' \\ &= f(x). \end{aligned} \tag{11.4}$$

Let us check whether $G(x,x') = H(x-x')\sinh(x-x')$ is a Green's function of the differential operator $\mathscr{L}_x = \frac{d^2}{dx^2} - 1$. In this case, all we have to do is to verify that $\mathscr{L}_x$, applied to $G(x,x')$, actually renders $\delta(x-x')$, as required by Eq. (11.2).

$$\begin{aligned} \mathscr{L}_x G(x,x') &= \delta(x-x') \\ \left(\frac{d^2}{dx^2} - 1\right) H(x-x')\sinh(x-x') &\stackrel{?}{=} \delta(x-x') \end{aligned}$$

Note that $\frac{d}{dx}\sinh x = \cosh x$, $\frac{d}{dx}\cosh x = \sinh x$; and hence

$$\frac{d}{dx}\left(\underbrace{\delta(x-x')\sinh(x-x')}_{=0} + H(x-x')\cosh(x-x')\right) - H(x-x')\sinh(x-x') =$$

$$\delta(x-x')\cosh(x-x') + H(x-x')\sinh(x-x') - H(x-x')\sinh(x-x') = \delta(x-x').$$

The solution (11.4) so obtained is *not unique*, as it is only a special solution to the inhomogenuous equation (11.1). The general solution to (11.1) can be found by adding the general solution $y_0(x)$ of the corresponding *homogenuous* differential equation

$$\mathscr{L}_x y(x) = 0 \tag{11.5}$$

to one special solution – say, the one obtained in Eq. (11.4) through Green's function techniques.

Indeed, the most general solution

$$Y(x) = y(x) + y_0(x) \tag{11.6}$$

clearly is a solution of the inhomogenuous differential equation (11.4), as

$$\mathscr{L}_x Y(x) = \mathscr{L}_x y(x) + \mathscr{L}_x y_0(x) = f(x) + 0 = f(x). \tag{11.7}$$

Conversely, any two distinct special solutions $y_1(x)$ and $y_2(x)$ of the inhomogenuous differential equation (11.4) differ only by a function which is a solution tho the homogenuous differential equation (11.5), because due to linearity of $\mathscr{L}_x$, their difference $y_d(x) = y_1(x) - y_2(x)$ can be parameterized by some function in y_0

$$\mathscr{L}_x[y_1(x) - y_2(x)] = \mathscr{L}_x y_1(x) + \mathscr{L}_x y_2(x) = f(x) - f(x) = 0. \tag{11.8}$$

From now on, we assume that the coefficients $a_j(x) = a_j$ in Eq. (11.1) are constants, and thus *translational invariant*. Then the entire *Ansatz* (11.2) for $G(x, x')$ is translation invariant, because derivatives are defined only by relative distances, and $\delta(x - x')$ is translation invariant for the same reason. Hence,

$$G(x, x') = G(x - x'). \tag{11.9}$$

For such translation invariant systems, the Fourier analysis represents an excellent way of analyzing the situation.

Let us see why translanslation invariance of the coefficients $a_j(x) = a_j(x + \xi) = a_j$ under the translation $x \to x + \xi$ with arbitrary ξ – that is, independence of the coefficients a_j on the "coordinate" or "parameter" x – and thus of the Green's function, implies a simple form of the latter. Translanslation invariance of the Green's function really means

$$G(x + \xi, x' + \xi) = G(x, x'). \tag{11.10}$$

Now set $\xi = -x'$; then we can define a new green's functions which just depends on one argument (instead of previously two), which is the difference of the old arguments

$$G(x - x', x' - x') = G(x - x', 0) \to G(x - x'). \tag{11.11}$$

What is important for applications is the possibility to adapt the solutions of some inhomogenuous differential equation to boundary and initial value problems. In particular, a properly chosen $G(x - x')$, in its dependence on the parameter x, "inherits" some behaviour of the solution $y(x)$. Suppose, for instance, we would like to find solutions with $y(x_i) = 0$ for some parameter values x_i, $i = 1, \ldots, k$. Then, the Green's function G must vanish there also

$$G(x_i - x') = 0 \text{ for } i = 1, \ldots, k. \tag{11.12}$$

11.2 Finding Green's functions by spectral decompositions

It has been mentioned earlier (cf. Section 10.6.5 on page 160) that the δ-function can be expressed in terms of various *eigenfunction expansions*. We shall make use of these expansions here [1].

[1] Dean G. Duffy. *Green's Functions with Applications*. Chapman and Hall/CRC, Boca Raton, 2001

Suppose $\psi_i(x)$ are *eigenfunctions* of the differential operator $\mathscr{L}_x$, and λ_i are the associated *eigenvalues*; that is,

$$\mathscr{L}_x \psi_i(x) = \lambda_i \psi_i(x). \tag{11.13}$$

Suppose further that $\mathscr{L}_x$ is of degree n, and therefore (we assume without proof) that we know all (a complete set of) the n eigenfunctions $\psi_1(x), \psi_2(x), \ldots, \psi_n(x)$ of $\mathscr{L}_x$. In this case, orthogonality of the system of eigenfunctions holds, such that

$$\int_{-\infty}^{\infty} \psi_i(x)\overline{\psi_j(x)}dx = \delta_{ij}, \tag{11.14}$$

as well as completeness, such that,

$$\sum_{i=1}^{n} \psi_i(x)\overline{\psi_i(x')} = \delta(x-x'). \tag{11.15}$$

$\overline{\psi_i(x')}$ stands for the complex conjugate of $\psi_i(x')$. The sum in Eq. (11.15) stands for an integral in the case of continuous spectrum of $\mathscr{L}_x$. In this case, the Kronecker δ_{ij} in (11.14) is replaced by the Dirac delta function $\delta(k-k')$. It has been mentioned earlier that the δ-function can be expressed in terms of various *eigenfunction expansions*.

The Green's function of $\mathscr{L}_x$ can be written as the spectral sum of the absolute squares of the eigenfunctions, divided by the eigenvalues λ_j; that is,

$$G(x-x') = \sum_{j=1}^{n} \frac{\psi_j(x)\overline{\psi_j(x')}}{\lambda_j}. \tag{11.16}$$

For the sake of proof, apply the differential operator $\mathscr{L}_x$ to the Green's function *Ansatz* G of Eq. (11.16) and verify that it satisfies Eq. (11.2):

$$\begin{aligned}
&\mathscr{L}_x G(x-x') \\
&= \mathscr{L}_x \sum_{j=1}^{n} \frac{\psi_j(x)\overline{\psi_j(x')}}{\lambda_j} \\
&= \sum_{j=1}^{n} \frac{[\mathscr{L}_x\psi_j(x)]\overline{\psi_j(x')}}{\lambda_j} \\
&= \sum_{j=1}^{n} \frac{[\lambda_j\psi_j(x)]\overline{\psi_j(x')}}{\lambda_j} \\
&= \sum_{j=1}^{n} \psi_j(x)\overline{\psi_j(x')} \\
&= \delta(x-x').
\end{aligned} \tag{11.17}$$

For a demonstration of completeness of systems of eigenfunctions, consider, for instance, the differential equation corresponding to the harmonic vibration [please do not confuse this with the harmonic oscillator (9.29)]

$$\mathscr{L}_t\psi = \frac{d^2}{dt^2}\psi = k^2, \tag{11.18}$$

with $k \in \mathbb{R}$.

Without any boundary conditions the associated eigenfunctions are

$$\psi_\omega(t) = e^{-i\omega t}, \tag{11.19}$$

with $0 \le \omega \le \infty$, and with eigenvalue $-\omega^2$. Taking the complex conjugate and integrating over ω yields [modulo a constant factor which depends on the choice of Fourier transform parameters; see also Eq. (10.76)]

$$\begin{aligned}
&\int_{-\infty}^{\infty} \psi_\omega(t)\overline{\psi_\omega(t')}d\omega \\
&= \int_{-\infty}^{\infty} e^{-i\omega t} e^{i\omega t'} d\omega \\
&= \int_{-\infty}^{\infty} e^{-i\omega(t-t')} d\omega \\
&= \delta(t-t').
\end{aligned} \tag{11.20}$$

The associated Green's function is

$$G(t-t') = \int_{-\infty}^{\infty} \frac{e^{-i\omega(t-t')}}{(-\omega^2)} d\omega. \tag{11.21}$$

And the solution is obtained by integrating over the constant k^2; that is,

$$\psi(t) = \int_{-\infty}^{\infty} G(t-t')k^2 dt' = -\int_{-\infty}^{\infty} \left(\frac{k}{\omega}\right)^2 e^{-i\omega(t-t')} d\omega dt'. \tag{11.22}$$

Note that if we are imposing boundary conditions; e.g., $\psi(0) = \psi(L) = 0$, representing a string "fastened" at positions 0 and L, the eigenfunctions change to

$$\psi_k(t) = \sin(\omega_n t) = \sin\left(\frac{n\pi}{L} t\right), \tag{11.23}$$

with $\omega_n = \frac{n\pi}{L}$ and $n \in \mathbb{Z}$. We can deduce orthogonality and completeness by listening to the orthogonality relations for sines (9.11).

For the sake of another example suppose, from the Euler-Bernoulli bending theory, we know (no proof is given here) that the equation for the quasistatic bending of slender, isotropic, homogeneous beams of constant cross-section under an applied transverse load $q(x)$ is given by

$$\mathcal{L}_x y(x) = \frac{d^4}{dx^4} y(x) = q(x) \approx c, \tag{11.24}$$

with constant $c \in \mathbb{R}$. Let us further assume the boundary conditions

$$y(0) = y(L) = \frac{d^2}{dx^2} y(0) = \frac{d^2}{dx^2} y(L) = 0. \tag{11.25}$$

Also, we require that y(x) vanishes everywhere except inbetween 0 and L; that is, $y(x) = 0$ for $x = (-\infty, 0)$ and for $x = (l, \infty)$. Then in accordance with these boundary conditions, the system of eigenfunctions $\{\psi_j(x)\}$ of $\mathcal{L}_x$ can be written as

$$\psi_j(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{\pi j x}{L}\right) \tag{11.26}$$

for $j = 1, 2, \ldots$. The associated eigenvalues

$$\lambda_j = \left(\frac{\pi j}{L}\right)^4$$

can be verified through explicit differentiation

$$
\begin{aligned}
\mathscr{L}_x \psi_j(x) &= \mathscr{L}_x \sqrt{\frac{2}{L}} \sin\left(\frac{\pi j x}{L}\right) \\
&= \mathscr{L}_x \sqrt{\frac{2}{L}} \sin\left(\frac{\pi j x}{L}\right) \\
&= \left(\frac{\pi j}{L}\right)^4 \sqrt{\frac{2}{L}} \sin\left(\frac{\pi j x}{L}\right) \\
&= \left(\frac{\pi j}{L}\right)^4 \psi_j(x).
\end{aligned}
\tag{11.27}
$$

The cosine functions which are also solutions of the Euler-Bernoulli equations (11.24) do not vanish at the origin $x = 0$.

Hence,

$$
\begin{aligned}
G(x-x')(x) &= \frac{2}{L} \sum_{j=1}^{\infty} \frac{\sin\left(\frac{\pi j x}{L}\right) \sin\left(\frac{\pi j x'}{L}\right)}{\left(\frac{\pi j}{L}\right)^4} \\
&= \frac{2L^3}{\pi^4} \sum_{j=1}^{\infty} \frac{1}{j^4} \sin\left(\frac{\pi j x}{L}\right) \sin\left(\frac{\pi j x'}{L}\right)
\end{aligned}
\tag{11.28}
$$

Finally we are in a good shape to calculate the solution explicitly by

$$
\begin{aligned}
y(x) &= \int_0^L G(x-x') g(x') dx' \\
&\approx \int_0^L c \left[\frac{2L^3}{\pi^4} \sum_{j=1}^{\infty} \frac{1}{j^4} \sin\left(\frac{\pi j x}{L}\right) \sin\left(\frac{\pi j x'}{L}\right) \right] dx' \\
&\approx \frac{2cL^3}{\pi^4} \sum_{j=1}^{\infty} \frac{1}{j^4} \sin\left(\frac{\pi j x}{L}\right) \left[\int_0^L \sin\left(\frac{\pi j x'}{L}\right) dx' \right] \\
&\approx \frac{4cL^4}{\pi^5} \sum_{j=1}^{\infty} \frac{1}{j^5} \sin\left(\frac{\pi j x}{L}\right) \sin^2\left(\frac{\pi j}{2}\right)
\end{aligned}
\tag{11.29}
$$

11.3 *Finding Green's functions by Fourier analysis*

If one is dealing with translation invariant systems of the form

$$
\begin{aligned}
&\mathscr{L}_x y(x) = f(x), \text{ with the differential operator} \\
&\mathscr{L}_x = a_n \frac{d^n}{dx^n} + a_{n-1} \frac{d^{n-1}}{dx^{n-1}} + \ldots + a_1 \frac{d}{dx} + a_0 \\
&\qquad = \sum_{j=0}^{n} a_j(x) \frac{d^j}{dx^j},
\end{aligned}
\tag{11.30}
$$

with constant coefficients a_j, then we can apply the following strategy using Fourier analysis to obtain the Green's function.

First, recall that, by Eq. (10.75) on page 159 the Fourier transform of the delta function $\widetilde{\delta}(k) = 1$ is just a constant; with our definition unity. Then, δ can be written as

$$\delta(x - x') = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik(x-x')}dk \tag{11.31}$$

Next, consider the Fourier transform of the Green's function

$$\widetilde{G}(k) = \int_{-\infty}^{\infty} G(x)e^{-ikx}dx \tag{11.32}$$

and its back trasform

$$G(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{G}(k)e^{ikx}dk. \tag{11.33}$$

Insertion of Eq. (11.33) into the *Ansatz* $\mathcal{L}_x G(x - x') = \delta(x - x')$ yields

$$\begin{aligned} &\mathcal{L}_x G(x) \\ &= \mathcal{L}_x \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{G}(k)e^{ikx}dk \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{G}(k)\left(\mathcal{L}_x e^{ikx}\right)dk = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}dk. \end{aligned} \tag{11.34}$$

and thus, through comparison of the integral kernels,

$$\begin{aligned} \frac{1}{2\pi}\int_{-\infty}^{\infty} \left[\widetilde{G}(k)\mathcal{L}_x - 1\right] e^{ikx}dk = 0, \\ \widetilde{G}(k)\mathcal{L}_k - 1 = 0, \\ \widetilde{G}(k) = (\mathcal{L}_k)^{-1}, \end{aligned} \tag{11.35}$$

where $\mathcal{L}_k$ is obtained from $\mathcal{L}_x$ by substituting every derivative $\frac{d}{dx}$ in the latter by ik in the former. in that way, the Fourier transform $\widetilde{G}(k)$ is obtained as a polynomial of degree n, the same degree as the highest order of derivative in $\mathcal{L}_x$.

In order to obtain the Green's function $G(x)$, and to be able to integrate over it with the inhomogenuous term $f(x)$, we have to Fourier transform $\widetilde{G}(k)$ back to $G(x)$.

Then we have to make sure that the solution obeys the initial conditions, and, if necessary, we have to add solutions of the homogenuos equation $\mathcal{L}_x G(x - x') = 0$. That is all.

Let us consider a few examples for this procedure.

1. First, let us solve the differential operator $y' - y = t$ on the intervall $[0, \infty)$ with the boundary conditions $y(0) = 0$.

   We observe that the associated differential operator is given by

   $$\mathcal{L}_t = \frac{d}{dt} - 1,$$

   and the inhomogenuous term can be identified with $f(t) = t$.

We use the *Ansatz* $G_1(t,t') = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \tilde{G}_1(k) e^{ik(t-t')}dk$; hence

$$
\begin{aligned}
\mathscr{L}_t G_1(t,t') &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \tilde{G}_1(k) \underbrace{\left(\frac{d}{dt}-1\right) e^{ik(t-t')}}_{=(ik-1)e^{ik(t-t')}} dk = \\
&= \delta(t-t') = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{ik(t-t')}dk
\end{aligned}
$$

Now compare the kernels of the Fourier integrals of $\mathscr{L}_t G_1$ and δ:

$$
\tilde{G}_1(k)(ik-1) = 1 \Longrightarrow \tilde{G}_1(k) = \frac{1}{ik-1} = \frac{1}{i(k+i)}
$$

$$
G_1(t,t') = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{ik(t-t')}}{i(k+i)}dk
$$

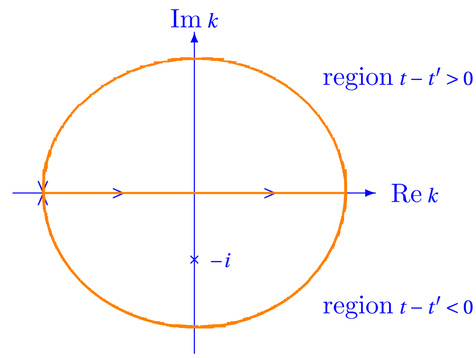


Figure 11.1: Plot of the two paths reqired for solving the Fourier integral.

The paths in the upper and lower integration plain are drawn in Frig. 11.1.

The "closures" throught the respective half-circle paths vanish.

$$
\begin{aligned}
\text{residuum theorem: } G_1(t,t') &= 0 \quad \text{for} \quad t>t' \\
G_1(t,t') &= -2\pi i \,\mathrm{Res}\left(\frac{1}{2\pi i}\frac{e^{ik(t-t')}}{(k+i)};-i\right) = \\
&= -e^{t-t'} \quad \text{for } t<t'.
\end{aligned}
$$

Hence we obtain a Green's function for the inhomogenuous differential equation

$$
G_1(t,t') = -H(t'-t)e^{t-t'}
$$

However, this Green's function and its associated (special) solution does not obey the boundary conditions $G_1(0,t') = -H(t')e^{-t'} \neq 0$ for $t' \in [0,\infty)$.

Therefore, we have to fit the Green's function by adding an appropriately weighted solution to the homogenuos differential equation. The homogenuous Green's function is found by

$$
\mathscr{L}_t G_0(t,t') = 0,
$$

and thus, in particular,

$$
\frac{d}{dt}G_0 = G_0 \Longrightarrow G_0 = ae^{t-t'}.
$$

with the *Ansatz*

$$
G(0,t') = G_1(0,t') + G_0(0,t';a) = -H(t')e^{-t'} + ae^{-t'}
$$

for the general solution we can choose the constant coefficient a so that

$$G(0,t') = G_1(0,t') + G_0(0,t';a) = -H(t')e^{-t'} + ae^{-t'} = 0$$

For $a = 1$, the Green's function and thus the solution obeys the boundary value conditions; that is,

$$G(t,t') = \left[1 - H(t'-t)\right]e^{t-t'}.$$

Since $H(-x) = 1 - H(x)$, $G(t,t')$ can be rewritten as

$$G(t,t') = H(t-t')e^{t-t'}.$$

In the final step we obtain the solution through integration of G over the inhomogenuous term t:

$$\begin{aligned}
y(t) &= \int_0^\infty \underbrace{H(t-t')}_{=1 \text{ for } t'<t} e^{t-t'} t' dt' = \\
&= \int_0^t e^{t-t'} t' dt' = e^t \int_0^t t' e^{-t'} dt' = \\
&= e^t \left(-t'e^{-t'}\Big|_0^t - \int_0^t (-e^{-t'}) dt' \right) = \\
&= e^t \left[(-te^{-t}) - e^{-t'}\Big|_0^t \right] = e^t \left(-te^{-t} - e^{-t} + 1 \right) = e^t - 1 - t.
\end{aligned}$$

2. Next, let us solve the differential equation $\frac{d^2y}{dt^2} + y = \cos t$ on the intervall $t \in [0,\infty)$ with the boundary conditions $y(0) = y'(0) = 0$.

   First, observe that

$$\mathscr{L} = \frac{d^2}{dt^2} + 1.$$

   The Fourier *Ansatz* for the Green's function is

$$\begin{aligned}
G_1(t,t') &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \tilde{G}(k)e^{ik(t-t')}dk \\
\mathscr{L}G_1 &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \tilde{G}(k)\left(\frac{d^2}{dt^2}+1\right)e^{ik(t-t')}dk = \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \tilde{G}(k)((ik)^2+1)e^{ik(t-t')}dk = \\
&= \delta(t-t') = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{ik(t-t')}dk =
\end{aligned}$$

   Hence

$$\tilde{G}(k)(1-k^2) = 1$$

ant thus

$$\tilde{G}(k) = \frac{1}{(1-k^2)} = \frac{-1}{(k+1)(k-1)}$$

The Fourier transformation is

$$\begin{aligned}
G_1(t,t') &= -\frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{ik(t-t')}}{(k+1)(k-1)}dk = \\
&= -\frac{1}{2\pi}2\pi i\left[\operatorname{Res}\left(\frac{e^{ik(t-t')}}{(k+1)(k-1)};k=1\right)+\right.\\
&\qquad \left.\operatorname{Res}\left(\frac{e^{ik(t-t')}}{(k+1)(k-1)};k=-1\right)\right]H(t-t')
\end{aligned}$$

The path in the upper integration plain is drawn in Fig. 11.2.

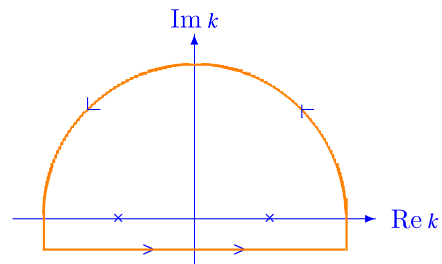


Figure 11.2: Plot of the path reqired for solving the Fourier integral.

$$\begin{aligned}
G_1(t,t') &= -\frac{i}{2}\left(e^{i(t-t')}-e^{-i(t-t')}\right)H(t-t') = \\
&= \frac{e^{i(t-t')}-e^{-i(t-t')}}{2i}H(t-t') = \sin(t-t')H(t-t') \\
G_1(0,t') &= \sin(-t')H(-t') = 0 \qquad \text{since} \quad t'>0 \\
G_1'(t,t') &= \cos(t-t')H(t-t') + \underbrace{\sin(t-t')\delta(t-t')}_{=0} \\
G_1'(0,t') &= \cos(-t')H(-t') = 0 \qquad \text{as before.}
\end{aligned}$$

G_1 already satisfies the boundary conditions; hence we do not need to find the Green's function G_0 of the homogenuous equation.

$$\begin{aligned}
y(t) &= \int_0^\infty G(t,t')f(t')dt' = \int_0^\infty \sin(t-t')\underbrace{H(t-t')}_{=1\text{ for } t>t'}\cos t'dt' = \\
&= \int_0^t \sin(t-t')\cos t'dt' = \int_0^t (\sin t\cos t' - \cos t\sin t')\cos t'dt' = \\
&= \int_0^t \left[\sin t(\cos t')^2 - \cos t\sin t'\cos t'\right]dt' = \\
&= \sin t\int_0^t (\cos t')^2 dt' - \cos t\int_0^t sint'\cos t'dt' = \\
&= \sin t\left[\frac{1}{2}(t'+\sin t'\cos t')\right]\Bigg|_0^t - \cos t\left[\frac{\sin^2 t'}{2}\right]\Bigg|_0^t = \\
&= \frac{t\sin t}{2} + \frac{\sin^2 t\cos t}{2} - \frac{\cos t\sin^2 t}{2} = \frac{t\sin t}{2}.
\end{aligned}$$

Part IV:
Differential equations

2.13 References

References:

- Morse and Feshbach: Methods of Theoretical Physics. This book used to be hard to find, but can now be bought at feshbachpublishing.com.
- Mathews and Walker: Mathematical Methods of Physics. An absolute joy for those who love math, and very informative even for those who don't.
- Arfken: Mathematical Methods for Physicists Academic Press Good introduction at graduate level. Not comprehensive in any area, but covers many areas widely.
- Gradshteyn and Ryzhik: Table of Integrals, Series, and Products Academic THE book of integrals. Huge, but useful when you need an integral.
- F.W. Byron and R. Fuller: Mathematics of Classical and Quantum Physics (2 vols) is a really terrific text for self-study; it is like a baby version of Morse & Feshbach.
- Luke, Y. L. The Special Functions and Their Approximations, 2 vols. New York: Academic Press, 1969.
- Mathai, A. M. A Handbook of Generalized Special Functions for Statistical and Physical Sciences. New York: Oxford University Press, 1993.
- Knopp, K. Theory of Functions Parts I and II, Two Volumes Bound as One, Part II. New York: Dover, pp. 99-118, 1996.
- Abramowitz, M. and Stegun, I. A. (Eds.). "Gamma (Factorial) Function" and "Incomplete Gamma Function." Pharagraph 6.1 and 6.5 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, pp. 255-258 and 260-263, 1972.
- Abramowitz, M. and Stegun, I. A. (Eds.). "Riemann Zeta Function and Other Sums of Reciprocal Powers." Paragraph 23.2 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, pp. 807-808, 1972.
- Aizenberg, L.; Adamchik, V.; and Levit, V. E. "Approaching the Riemann Hypothesis with Mathematica." http://library.wolfram.com/infocenter/Articles/3268/.
- Arfken, G. Mathematical Methods for Physicists, 3rd ed. Orlando, FL: Academic Press, pp. 332-335, 1985.
- Bailey, D. H.; Borwein, J. M.; Calkin, N. J.; Girgensohn, R.; Luke, D. R.; and Moll, V. H. Experimental Mathematics in Action. Wellesley, MA: A K Peters, 2007.
- Watson, G. N. A Treatise on the Theory of Bessel Functions, 2nd ed. Cambridge, England: Cambridge University Press, 1966.
- Abramowitz, M. and Stegun, I. A. (Eds.). "Hypergeometric Functions." Ch. 15 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, pp. 555-566, 1972.

- Appell, P. and Kampé de Fériet, J. Fonctions hypergómétriques et hypersphériques: polynomes d'Hermite. Paris: Gauthier-Villars, 1926.
- Arfken, G. "Hypergeometric Functions." Paragraph 13.5 in Mathematical Methods for Physicists, 3rd ed. Orlando, FL: Academic Press, pp. 748-752, 1985.
- Bailey, W. N. Generalised Hypergeometric Series. Cambridge, England: University Press, 1935.
- Borwein, J.; Bailey, D.; and Girgensohn, R. Experimentation in Mathematics: Computational Paths to Discovery. Wellesley, MA: A K Peters, 2004.
- Erdélyi, A.; Magnus, W.; Oberhettinger, F.; and Tricomi, F. G. Higher Transcendental Functions, Vol. 1. New York: Krieger, 1981.

Chapter 3

Lie Groups

1. Introduction

Many systems studied in physics show some form of symmetry. In physics, this means the following: we can consider some *transformation rule*, like a rotation, a displacement, or the reflection by a mirror, and we compare the original system with the transformed system. If they show some resemblance, we have a symmetry. A snow flake looks like itself when we rotate it by 60° or when we perform a mirror reflection. We say that the snow flake has a symmetry. If we replace a proton by a neutron, and *vice versa*, the replaced particles behave very much like the originals; this is also a symmetry. Many laws of Nature have symmetries in this sense. Sometimes the symmetry is perfect, but often it is not exact; the transformed system is then slightly different from the original; the symmetry is broken.

If system A resembles system B, and system B resembles C, then A resembles C. Therefore, the product of two symmetry transformations is again a symmetry transformation. Thus, the set of all symmetry transformations that characterize the symmetry of a system, are elements of a *group*. For example, the reflections with respect to a plane form a group that contains just two elements: the reflection operation and the identity — the identity being the one operation that leaves everything the same. The rotations in three-dimensional space, the set of all Lorentz transformations, and the set of all parallel displacements also form groups, which have an unlimited number of elements. For obvious reasons, groups with a finite (or denumerable) number of elements are called *discrete* groups; groups of transformations that continuously depend on a number of parameters, such as the rotations, which can be defined in terms of a few angular parameters, are called *continuous* groups.

The symmetry of a system implies certain relations among observable quantities, which may be obeyed with great precision, independently of the nature of the forces acting in the system. In the hydrogen atom, for example, one finds that the energies of different states of the atom, are exactly equal, as a consequence of the rotational invariance of the system. However, one also often finds that the symmetry of a physical system is only approximately realized. An infinite crystal, for example, is invariant under those translations for which the displacement is an integral multiple of the distance between two adjacent atoms. In reality, however, the crystal has a definite size, and its surface perturbs the translational symmetry. Nevertheless, if the crystal contains a sufficiently large number of atoms, the disturbance due to the surface has little effects on the properties at the interior.

An other example of a symmetry that is only approximately realized, is encountered in elementary particle physics. The so-called Δ^+ particle, which is one of the excited states of the nucleons, decays into a nucleon and an other particle, the π-meson, also called *pion*. There exist two kinds of nucleons, neutrons and protons, and there are three types of pions, the electrically charged pions π^+ and π^-, and the neutral one, π^0. Since the total electric charge of the Δ^+ must be preserved during its decay, one distinguishes

[1] This lecture course was originally set up by M. Veltman, and subsequently modified and extended by B. de Wit and G. 't Hooft.

nucleons	pions	Δ particles
$m_{\text{proton}} \approx 938\ \text{MeV}/c^2$	$m_{\pi^+} \approx 140\ \text{MeV}/c^2$	$m_{\Delta^{++}} \approx 1231\ \text{MeV}/c^2$
$m_{\text{neutron}} \approx 939\ \text{MeV}/c^2$	$m_{\pi^0} \approx 135\ \text{MeV}/c^2$	$m_{\Delta^+} \approx 1232\ \text{MeV}/c^2$
	$m_{\pi^-} \approx 140\ \text{MeV}/c^2$	$m_{\Delta^0} \approx 1233\ \text{MeV}/c^2$
		$m_{\Delta^-} \approx 1235\ \text{MeV}/c^2$

Table 1: Masses of nucleons, pions and Δ particles, expressed in MeV/c^2.

two possible decay modes:

$$\Delta^+ \to n\ \pi^+ \qquad \text{and} \qquad \Delta^+ \to p\ \pi^0\,. \tag{1.1}$$

Remarkably, the second decay occurs twice as often as the the first one, a fact that seems to be difficult to explain as being due to the differences in the charges of the decay products. A natural explanation of this factor 2 could follow from symmetry considerations. This is not as strange as it might seem, because protons and neutrons have nearly identical masses, just as the three species of pions and the four Δ particles that are found in Nature (see table).

It will be demonstrated that the near equality of the masses, and also the factor 2 in the two decay modes (1.1), can be explained by assuming nature to be invariant under so-called *isospin* transformations. The notion of 'isobaric spin', or 'isospin' for short, was introduced by Heisenberg in 1932. He was puzzled by the fact that protons and neutrons have nearly equal masses, while, apart from the obvious differences in electrical charge, also other properties are much alike. Thus, the nucleons form a *doublet*, just like electrons that show a doublet structure as a consequence of the fact that there are two possible spin orientations for the electron states — hence the term isobaric *spin*. Later, it turned out that elementary particles with nearly equal masses can always be arranged in so-called isospin multiplets. The nucleons form an isospin doublet, the pions an isospin triplet, and the Δ particles an isospin quadruplet. Particles inside a single multiplet all have approximately identical masses, but different electric charges. The charge arrangement is as indicated in the table: no two particles in one multiplet have the same charge, and the particles can always be arranged in such a way that the charge difference between two successive particles is exactly one elementary charge unit.

However, it will be clear that isospin invariance can only be an approximation, since the masses of the nucleons, pions and Δ particles turn out to depend somewhat on their electric charges. The mass differences within a multiplet are only of the order of a few percent, and this is the degree of accuracy that one can expect for theoretical predictions based upon isospin invariance.

The above example is an application of group theory in the physics of elementary particles, but invariance principles play an important role in nearly all branches of physics. In atomic physics we frequently notice the consequences of rotation invariance, in nuclear physics we have rotation and isospin invariance, in solid state physics also invariance under discrete translations and rotations. Also in (quantum) field theory, symmetry transformations are important. A very special kind of transformations are encountered for example in electrodynamics. Here, electric and magnetic fields can be expressed in terms of the so-called *vector potential* $A_\mu(x)$, for which we use a relativistic four-vector notation ($\mu = 0,1,2,3$):

$$A_\mu(x) = (\,-\,c^{-1}\,\phi(x),\,\mathbf{A}(x))\ , \qquad x^\mu = (ct,\,\mathbf{x})\ , \tag{1.2}$$

where ϕ denotes the potential, and $\mathbf{A}$ the three-dimensional vector potential field; c is the velocity of light. The electric and magnetic fields are defined by

$$\mathbf{E} \;=\; -\nabla\phi - c^{-1}\,\frac{\partial \mathbf{A}}{\partial t}\ , \tag{1.3}$$

$$\mathbf{B} \;=\; \nabla \times \mathbf{A}\ . \tag{1.4}$$

An electrically charged particle is described by a complex wave function $\psi(\vec{x},\,t)$. The Schrödinger equation obeyed by this wave function remains valid when one performs a rotation in the complex plane:

$$\psi(\vec{x},\,t) \to e^{i\Lambda}\,\psi(\vec{x},\,t)\ . \tag{1.5}$$

Is the phase factor Λ allowed to vary in space and time?

The answer to this is yes, however only if the Schrödinger equation depends on the vector potential in a very special way. Wherever a derivative ∂_μ occurs, it must be in the combination

$$D_\mu = \partial_\mu - ieA_\mu\ , \tag{1.6}$$

where e is the electric charge of the particle in question. If $\Lambda(\vec{x},\,t)$ depends on $\vec{x}$ and t, then (1.5) must be associated with the following transformation rules for the potential fields:

$$\mathbf{A}(x) \;\to\; \mathbf{A}(x) + e^{-1}\nabla\Lambda(x)\ , \tag{1.7}$$

$$\phi(x) \;\to\; \phi(x) - (c\,e)^{-1}\frac{\partial}{\partial t}\Lambda(x)\ , \tag{1.8}$$

or, in four-vector notation,

$$A_\mu(x) \to A_\mu(x) + e^{-1}\partial_\mu\Lambda(x)\ . \tag{1.9}$$

It can now easily be established that $\mathbf{E}$ en $\mathbf{B}$ will not be affected by this so-called *gauge transformation.* Furthermore, we derive:

$$D_\mu\psi(x) \to e^{i\Lambda(x)}D_\mu\psi(x)\ . \tag{1.10}$$

Notice that the substitution (1.6) in the Schrödinger equation is all that is needed to include the interaction of a charged particle with the fields $\mathbf{E}$ en $\mathbf{B}$.

These phase factors define a group, called the group of 1×1 unitary matrices, $U(1)$. In this case, the group is quite a simple one, but it so happens that similar theories exist that are based on other (continuous) groups that are quite a bit more complicated such as the group $SU(2)$ that will be considered in these lectures. Theories of this type are known as *gauge theories*, or Yang-Mills theories, and the field A_μ is called a *gauge field*. The fact that $\mathbf{E}$ en $\mathbf{B}$ are invariant under gauge transformations implies that electromagnetic phenomena are *gauge-invariant*. For more general groups it turns out that several of these gauge fields are needed: they form multiplets.

Surprisingly, the theory of gravitation, Einstein's general relativity theory, turns out to be a gauge theory as well, be it of a somewhat different type. This theory can be considered to be the gauge theory of the general coordinate transformations, the most general reparametrizations of points in space and time,

$$x^\mu \to x^\mu + \xi^\mu(x) \ . \tag{1.11}$$

The gauge field here is the gravitational field, taking the form of a metric, which is to be used in the definitions of distances and angles in four-dimensional space and time. All of this is the subject of an entire lecture course, *Introduction to General Relativity*.

The fact that gauge transformations are associated to an abstract group, and can depend on space and time as well, can give rise to interesting phenomena of a topological nature. Examples of this are flux quantization in super conductors, the Aharonov-Bohm effect in quantum mechanics, and magnetic monopoles. To illustrate the relevance of topology, we consider again the group of the $U(1)$ gauge transformations, but now in two-dimensional space (or equivalently, in a situation where the fields only depend on two of the three space coordinates). Let $\psi(x,y)$ be a complex function, such as a wave function in quantum mechanics, transforming under these gauge transformations, *i.e.*

$$\psi(x,y) \to \mathrm{e}^{i\Lambda(x,y)}\psi(x,y) \ . \tag{1.12}$$

From the fact that the phase of ψ can be modified everywhere by applying different gauge transformations, one might conclude that the phase of ψ is actually irrelevant for the description of the system. This however is not quite the case. Consider for instance a function that vanishes at the origin. Now take a closed curve in the x - y plane, and check how the phase of $\psi(x,y)$ changes along the curve. After a complete run along the curve the phase might not necessarily take the same value as at the beginning, but if we assume that $\psi(x,y)$ is single-valued on the plane, then the phase difference will be equal to $2\pi n$, where n is an arbitrary *integral* number. This number is called the *winding number*. An example of a situation with winding number $n=1$ is pictured in Fig. 1; the phase angle makes a full turn over 2π when we follow the function $\psi(x,y)$ along a curve winding once around the origin. One can easily imagine situations with other winding numbers. The case $n=0$ for instance occurs when the phase of $\psi(x,y)$ is constant.

If we change the function $\psi(x,y)$ continuously, the winding number will not change. This is why the winding number is called a *topological invariant*. This also implies that the winding number will not change under the gauge transformations (1.12), provided that we limit ourselves to gauge transformations that are well-defined in the entire plane. Note also that the winding number does not depend on the choice of the closed curve around the origin, as long as it is not pulled across the origin or any other zero of the function

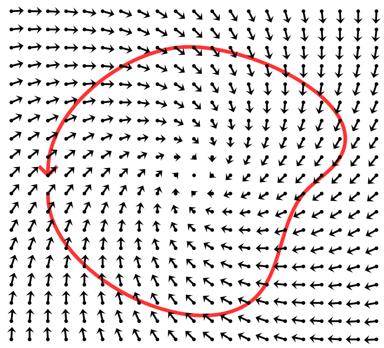


Figure 1: De phase angle of $\psi(x,y)$ indicated by an arrow (whose length is immaterial, but could be given for instance by $|\psi(x,y)|$) at various spots in the x - y plane. This function has a zero at the origin.

$\psi(x,y)$. All this implies that although locally, that is, at one point and its immediate neighborhood, the phase of ψ can be made to vanish, this can be realized globally, that is, on the entire plane, only if the winding number for any closed curve equals zero.

A similar situation can be imagined for the vector potential. Once more consider the two-dimensional plane, and assume that we are dealing with a magnetic field that is everywhere equal to zero, except for a small region surrounding the origin. In this region, $\mathbf{A}$ cannot be equal to zero, because of the relation (1.4). However, in the surrounding region, where $\mathbf{B}$ vanishes, there may seem to be no reason why not also $\mathbf{A}$ should vanish. Indeed, one can show that, at every given point and its neighborhood, a suitably chosen gauge transformation can ensure $\mathbf{A}(x)$ to vanish there. This result, however, can only hold locally, as we can verify by considering the following loop integral:

$$\Phi[C]=\oint_C A_i\,\mathrm{d}x^i, \tag{1.13}$$

where C is a given closed curve. It is easy to check that $\Phi[C]$ does not change under a gauge transformation (1.5). Indeed, we know from the theory of magnetism that $\Phi[C]$ must be proportional to the total magnetic flux through the surface enclosed by the curve C .

Applying this to the given situation, we take the curve C to surround the origin and the region where $\mathbf{B}\neq\mathbf{0}$, so that the $\mathbf{B}$ field vanishes on the curve itself. The quantity $\Phi[C]$ equals the total flux through C , which may well be different from zero. If this is the case, we cannot transform $\mathbf{A}$ away in the entire outside region, even if it can be transformed away locally[2]. Note that the magnetic flux here plays the same role as the winding number of the previous example. Indeed, in superconducting material, the gauge phases can be chosen such that $\mathbf{A}$ vanishes, and consequently, magnetic flux going through a superconducting coil is limited to integral values: the flux is *quantized*.

[2]This causes an interesting quantum mechanical effect in electrons outside a magnetic field, to wit, the *Aharonov-Bohm effect*.

Under some circumstances, magnetic field lines can penetrate superconducting materials in the form of *vortices*. These vortices again are quantized. In the case of more complicated groups, such as $SU(2)$, other situations of a similar nature can occur: magnetic monopoles are topologically stable objects in three dimensions; even in four dimensions one can have such phenomena, referred to as "instantons".

Clearly, group theory plays an essential role in physics. In these lectures we will primarily limit ourselves to the group of three-dimensional rotations, mostly in the context of quantum mechanics. Many of the essentials can be clarified this way, and the treatment can be made reasonably transparent, physically and mathematically. The course does not intend to give a complete mathematical analysis; rather, we wish to illustrate as clearly as possible the relevance of group theory for physics. Therefore, some physical applications will be displayed extensively. The rotation group is an example of a so-called *compact Lie group*. In most applications, we consider the *representations* of this group. Representation theory for such groups is completely known in mathematics. Some advance knowledge of linear algebra (matrices, inner products, traces, functions and derivatives of matrices, etc.) will be necessary. For completeness, some of the most important properties of matrices are summarized in a couple of appendices.

2. Quantum mechanics and rotation invariance

Quantum mechanics tells us that any physical system can be described by a (usually complex) wave function. This wave function is a solution of a differential equation (for instance the Schrödinger equation, if a non-relativistic limit is applicable) with boundary conditions determined by the physical situation. We will not indulge in the problems of determining this wave function in all sorts of cases, but we are interested in the properties of wave functions that follow from the fact that Nature shows certain symmetries. By making use of these symmetries we can save ourselves a lot of hard work doing calculations.

One of the most obvious symmetries that we observe in nature around us, is invariance of the laws of nature under rotations in three-dimensional space. An observer expects that the results of measurements should be independent of the orientation of his or her apparatus in space, assuming that the experimental setup is not interacting with its environment, or with the Earth's gravitational field. For instance, one does not expect that the time shown by a watch will depend on its orientation in space, or that the way a calculator works changes if we rotate it. Rotational symmetry can be found in many fundamental equations of physics: Newton's laws, Maxwell's laws, and Schrödinger's equation for example do not depend on orientation in space. To state things more precisely: Nature's laws are invariant under rotations in three-dimensional space.

We now intend to find out what the consequences are of this invariance under rotation for wave functions. From classical mechanics it is known that rotational invariance of a system with no interaction with its environment, gives rise to conservation of angular momentum: in such a system, the total angular momentum is a constant of the motion. This conservation law turns out to be independent of the details of the dynamical laws; it simply follows from more general considerations. It can be deduced in quantum mechanics as well. There turns out to be a connection between the behavior of a wave function under rotations and the conservation of angular momentum.

The equations may be hard to solve explicitly. But consider a wave function ψ depending on all sorts of variables, being the solution of some linear differential equation:

$$\mathcal{D}\psi = 0 \; . \tag{2.1}$$

The essential thing is that the exact form of $\mathcal{D}$ does not matter; the only thing that matters is that $\mathcal{D}$ be invariant under rotations. An example is Schrödinger's equation for a particle moving in a spherically symmetric potential $V(r)$,

$$\left[\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}+\frac{\partial^2}{\partial x_3^2}\right) - V(r) + i\hbar\frac{\partial}{\partial t}\right]\psi(\vec{x},t) = 0 \; , \qquad r \stackrel{\text{def}}{=} \sqrt{\vec{x}^{\,2}} \; . \tag{2.2}$$

Consider now the behavior of this differential equation under rotations. When we rotate, the position vector $\vec{x}$ turns into an other vector with coordinates x_i' :

$$x_i' = \sum_j R_{ij}\, x_j \; . \tag{2.3}$$

Here, we should characterize the rotation using a 3×3 matrix R, that is orthogonal and has determinant equal to 1 (orthogonal matrices with determinant -1 correspond to mirror reflections). The orthogonality condition for R implies that

$$\widetilde{R}\,R = R\,\widetilde{R} = \mathbf{1}\ , \qquad \text{or} \qquad \sum_i R_{ij}\;R_{ik} = \delta_{jk}\ ; \qquad \sum_j R_{ij}\;R_{kj} = \delta_{ik}\ , \tag{2.4}$$

where $\widetilde{R}$ is the transpose of R (defined by $\tilde{R}_{ij} = R_{ji}$).

It is not difficult now to check that equation (2.2) is rotationally invariant. To see this, consider[3] the function $\psi'(\vec{x},\,t) \stackrel{\mathrm{def}}{=} \psi(\vec{x}\,',\,t)$,

$$\frac{\partial}{\partial x_i}\psi'(\vec{x},\,t) = \frac{\partial}{\partial x_i}\psi(\vec{x}\,',t) = \sum_j \frac{\partial x'_j}{\partial x_i}\frac{\partial}{\partial x'_j}\psi(\vec{x}\,',t) = \sum_j R_{ji}\frac{\partial}{\partial x'_j}\psi(\vec{x}\,',t)\ , \tag{2.5}$$

where use was made of Eq. (2.3). Subsequently, we observe that

$$\begin{aligned}\sum_i \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_i}\,\psi(\vec{x}\,',t) \quad &= \quad \sum_{i,j,k} R_{ji}\;R_{ki}\,\frac{\partial}{\partial x'_j}\frac{\partial}{\partial x'_k}\,\psi(\vec{x}\,',t) \\ &= \quad \sum_i \frac{\partial}{\partial x'_i}\frac{\partial}{\partial x'_i}\,\psi(\vec{x}\,',t)\ ,\end{aligned} \tag{2.6}$$

where we made use of Eq. (2.4). Since $\vec{x}\,'^2 = \vec{x}^{\,2}$, the potential $V(r)$ also remains the same after a rotation. From the above, it follows that Equation (2.2) is invariant under rotations: if $\psi(\vec{x},\,t)$ is a solution of Eq. (2.2), then also $\psi'(\vec{x},\,t)$ must be a solution of the same equation.

In the above, use was made of the fact that rotations can be represented by real 3×3 matrices R. Their determinant must be $+1$, and they must obey the orthogonality condition $R\,\widetilde{R} = \mathbf{1}$. Every rotation in 3 dimensions can be represented by three angles (this will be made more precise in Chapter 3) Let R_1 and R_2 both be matrices belonging to some rotations; then their product $R_3 = R_1 R_2$ will also be a rotation. This statement is proven as follows: assume that R_1 and R_2 are orthogonal matrices with determinant 1. From the fact that

$$\widetilde{R}_1 = R_1^{-1}\ , \qquad \widetilde{R}_2 = R_2^{-1}\ , \tag{2.7}$$

it follows that also $R_3 = R_1 R_2$ is orthogonal:

$$\widetilde{R}_3 = \widetilde{R_1 R_2} = \widetilde{R}_2\widetilde{R}_1 = R_2^{-1}R_1^{-1} = (R_1R_2)^{-1} = R_3^{-1}\ . \tag{2.8}$$

Furthermore, we derive that

$$\det R_3 = \det(R_1R_2) = \det R_1\ \det R_2 = 1\ , \tag{2.9}$$

[3]By rotating $\vec{x}$ first, and taking the old function at the new point $\vec{x}\,'$ afterwards, we actually rotate the wave function into the opposite direction. This is a question of notation that, rather from being objectionable, avoids unnecessary complications in the calculations.

so that R_3 is a rotation as well. Note that also the product $R_4 = R_2 R_1$ is a rotation, but that R_3 en R_4 need not be the same. In other words, rotations are not *commutative*; when applied in a different order, the result will be different, in general.

We observe that the rotations form what is known as a *group*. A set of elements (here the set of real 3×3 matrices R with determinant 1 and $\widetilde{R}R = \mathbf{1}$) is called a group if an operation exists that we call 'multiplication' (here the ordinary matrix multiplication), in such a way that the following demands are obeyed:

1. If R_1 en R_2 are elements of the group, then also the product $R_1 R_2$ is an element of the group.

2. The multiplication is associative: $R_1(R_2R_3) = (R_1R_2)R_3$. So, one may either first multiply R_2 with R_3 , and then multiply the result with R_1 , or perform these two operations in the opposite order. Note that the order in which the matrices appear in this expression *does* have to stay the same.

3. There exists a unity element $\mathbf{1}$, such that $\mathbf{1}R = R$ for all elements R of the group. This unity element is also an element of the group.

4. For all elements R of the group, there exists in the group an *inverse* element R^{-1} such that $R^{-1}R = \mathbf{1}$.

The set of rotation matrices possesses all these properties. This set forms a group with infinitely many elements.

Every group is fully characterized by its multiplication structure, *i.e.* the relation between the elements *via* the multiplication rules. Later, we will attempt to define this notion of "structure" more precisely in terms of formulae. Note that a group does not possess notions such as "add" or "subtract", only "multiply". There is no "zero-element" in a group.

Much use is made of the fact that the set of all transformations that leave a system invariant, together form a group. If we have two invariance transformations, we can immediately find a third, by subjecting the quantities in terms of which the theory is defined, to the two transformations in succession. Obviously, the resulting transformation must leave the system invariant as well, and so this "product transformation" belongs to our set. Thus, the first condition defining a group is fulfilled; the others usually are quite obvious as well.

For what follows, the time dependence of the wave function is immaterial, and therefore we henceforth write a rotation R of a wave function as:

$$\psi'(\vec{x}) = \psi(\vec{x}\,') = \psi(R\vec{x}) \ . \tag{2.10}$$

Applying a second rotation S , gives us

$$\psi'' = \psi'(S\vec{x}) = \psi(RS\vec{x}) \ . \tag{2.11}$$

In what follows now, we will make use of the fact that the equation $\mathcal{D}\psi = 0$ is a *linear* equation. This is in contrast to the invariance transformation R, which may or may not be linear: the sum of two matrices R and S usually is not a legitimate rotation. It *is* true that if we have two solutions ψ_1 and ψ_2 of the equation (2.1), then every linear combination of these is a solution as well:

$$\mathcal{D}\,(\lambda\,\psi_1 + \mu\,\psi_2) = \lambda\,\mathcal{D}\psi_1 + \mu\,\mathcal{D}\psi_2 = 0 \ . \tag{2.12}$$

In general: if $\psi_1, \dots, \psi_n$ are solutions of the equation in (2.1) then also every linear combination

$$\lambda_1\,\psi_1 + \lambda_2\,\psi_2 + \cdots + \lambda_n\,\psi_n \tag{2.13}$$

is a solution of (2.1).

Regarding the behavior under rotations, we now distinguish two possible situations. Either the wave function ψ is rotationally invariant, that is, upon a rotation, ψ turns into itself,

$$\psi'(\vec{x}) = \psi(\vec{x}) \quad\Longleftrightarrow\quad \psi(\vec{x}\,') = \psi(\vec{x}) \ , \tag{2.14}$$

or we have sets of linearly independent solutions $\psi_1, \dots, \psi_n$, that, upon a rotation, each transform into some linear combination of the others. To illustrate the second possibility, we can take for example the set of solutions of particles moving in all possible directions. In this case, the set $\psi_1, \dots, \psi_n$ contains an infinite number of solutions. In order to avoid complications due to the infinite number of elements in this set, we can limit ourselves either to particles at rest, or omit the momentum dependence of the wave functions. Upon a rotation, a particle at rest turns into itself, but the internal structure might change. In this case, the set of wave functions that rotate into one another usually only contains a finite number of linearly independent solutions. If the particle is in its ground state, the associated wave function is often rotationally invariant; in that case, the set only contains one wave function. If the particle is in an excited state, different excited states can emerge after a rotation.

Now let there be given such a set $\Psi = (\psi_1, \dots, \psi_n)$ of wave functions transforming into one another upon a rotation. This means that after a rotation, ψ_1 turns into some linear combination of $\psi_1, \dots, \psi_n$,

$$\psi_1'(\vec{x}) \equiv \psi_1(R\,\vec{x}) = d_{11}\,\psi_1(\vec{x}) + d_{12}\,\psi_2(\vec{x}) + \cdots + d_{1n}\,\psi_n(\vec{x}) \ , \tag{2.15}$$

and a similar expression holds for $\psi_2, \dots \psi_n$. In general, we can write

$$\psi_A' = \sum_B d_{AB}\,\psi_B. \qquad (A, B = 1, \dots, n) \ . \tag{2.16}$$

The coefficients d_{AB} depend on R and form a matrix $D(R)$, such that

$$\Psi'(\vec{x}) = \Psi(R\,\vec{x}) = D(R)\,\Psi(\vec{x}) \ , \tag{2.17}$$

where we indicated the wave functions $\psi_1, \dots, \psi_n$ as a column vector Ψ. In the cases to be discussed next, there is only a limited number of linearly independent solutions of the equation $\mathcal{D}\psi = 0$, and therefore the space of all solutions (2.15) that we obtain by rotating one of them, must be finite-dimensional.

The matrices $D(R)$ in (2.15)-(2.16) are related to the rotation matrices R in the sense that for every rotation R in 3-dimensional space a matrix $D(R)$ exists that turns the solutions ψ_A into linear combinations of the same solutions. One can, however, say more. A given rotation can either be applied at once, or be the result of several rotations performed in succession. Whatever is the case, the final result should be the same. This implies that the matrices $D(R)$ must possess certain multiplication properties. To derive these, consider two successive rotations, R and S (see Eq. (2.11)). Let R be associated with a matrix $D(R)$, and S with a matrix $D(S)$. In formulae:

$$\begin{aligned} \Psi(R\,\vec{x}) &= D(R)\,\Psi(\vec{x}) \ , \\ \Psi(S\,\vec{x}) &= D(S)\,\Psi(\vec{x}) \ . \end{aligned} \tag{2.18}$$

Obviously, the combined rotation $R\,S$ must be associated with a matrix $D(R\,S)$, so that we have

$$\Psi(R\,S\,\vec{x}) = D(R\,S)\,\Psi(\vec{x}) \ . \tag{2.19}$$

But we can also determine $\Psi(R\,S)$ using Eq. (2.18),

$$\Psi(R\,S\,\vec{x}) = D(R)\,\Psi(S\,\vec{x}) = D(R)\,D(S)\,\Psi(\vec{x}) \ . \tag{2.20}$$

therefore, one must have[4]

$$D(R)\,D(S) = D(R\,S) \ . \tag{2.21}$$

Thus, the matrices $D(R)$ must have the same multiplication rules, the same multiplication structure, as the matrices R . A mapping of the group elements R on matrices $D(R)$ with this property is said to be a 'representation' of the group. We shall study various kinds of representations of the group of rotations in three dimensions.

Summarizing: a set of matrices forms a *representation* of a group, if one has

1. Every element a of the group is mapped onto a matrix A ,
2. The product of two elements is mapped onto the product of the corresponding matrices, *i.e.* if a , b and c are associated to the matrices A , B , and C , and $c = a\,b$, then one must have $C = A\,B$.

We found the following result: Upon rotations in three-dimensional space, the wave functions of a physical system must transform as linear mappings that form a representation of the group of rotations in three dimensions.

[4]In this derivation, note the order of R en S . The correct mathematical notation is: $D(R)\Psi = \Psi{\cdot}R$, so $D(R)\cdot(D(S)\cdot\Psi) = D(R)\cdot\Psi\cdot S = (\Psi\cdot R)\cdot S = D(RS)\cdot\Psi$. It is *not* correct to say that this should equal $D(R)\cdot(\Psi\cdot S) \stackrel{?}{=} (\Psi\cdot S)\cdot R$ because the definitions (2.18) *only* hold for the given wave function Ψ , *not* for $\Psi\cdot S$.

As a simple example take the three functions

$$\psi_1(\vec{x}) = x_1\, f(r)\ , \qquad \psi_2(\vec{x}) = x_2\, f(r)\ , \qquad \psi_3(\vec{x}) = x_3\, f(r)\ , \tag{2.22}$$

where $f(r)$ only depends on the radius $r = \sqrt{\vec{x}^{\,2}}$, which is rotationally invariant. These may be, for instance, three different solutions of the Schrödinger equation (2.2). Upon a rotation, these three functions transform with a matrix $D(R)$ that happens to coincide with R itself. The condition (2.21) is trivially obeyed.

However, the above conclusion may not always hold. According to quantum mechanics, two wave functions that only differ by a factor with absolute value equal to 1, must describe the same physical situation. The wave functions ψ and $\mathrm{e}^{i\alpha}\psi$ describe the same physical situation, assuming α to be real. This leaves us the possibility of a certain multivaluedness in the definition of the matrices $D(R)$. In principle, therefore, the condition (2.21) can be replaced by a weaker condition

$$D(R_1)\, D(R_2) = \exp\left[i\alpha(R_1, R_2)\right] D\left(R_1\, R_2\right)\ , \tag{2.23}$$

where α is a real phase angle depending on R_1 and R_2 . Matrices $D(R)$ obeying (2.23) with a non-trivial phase factor form what we call a *projective representation*. Projective representations indeed occur in physics. We shall discover circumstances where every matrix R of the rotation group is associated to *two* matrices $D(R)$ en $D'(R)$, differing from one another by a phase factor, to wit, a factor -1 . One has $D'(R) = -D(R)$. This is admitted because the wave functions ψ and $-\psi$ describe the same physical situation. This multivaluedness implies that the relation (2.21) is obeyed only up to a sign, so that the phase angle α in (2.23) can be equal to 0 or π . Particles described by wave functions transforming according to a projective representation, have no analogue in classical mechanics. Examples of such particles are the electron, the proton and the neutron. Their wave functions will transform in a more complicated way than what is described in Eq. (2.10). We shall return to this topic (Chapter 6).

The physical interpretation of the quantum wave function has another implication, in the form of an important constraint that the matrices $D(R)$ must obey. A significant role is attributed to the *inner product*, a mapping that associates a complex number to a pair of wave functions, ψ_1 and ψ_2 , to be written as $\langle\psi_1 \mid \psi_2\rangle$, and obeying the following relations (see Appendix E):

$$\begin{aligned}
\langle\psi \mid \psi\rangle &\geq 0\ , \\
\langle\psi \mid \psi\rangle &= 0\ , \qquad \text{then and } \textit{only} \text{ then if } \mid \psi\rangle = 0\ , && (2.24)\\
\langle\psi_1 \mid \lambda\,\psi_2 + \mu\,\psi_3\rangle &= \lambda\,\langle\psi_1 \mid \psi_2\rangle + \mu\,\langle\psi_1 \mid \psi_3\rangle\ , && (2.25)\\
&\qquad \text{for every pair of complex numbers } \lambda \text{ and } \mu\ , \\
\langle\psi_1 \mid \psi_2\rangle^* &= \langle\psi_2 \mid \psi_1\rangle\ . && (2.26)
\end{aligned}$$

For wave functions depending on just one coordinate, such an inner product is defined by

$$\langle\psi_1 \mid \psi_2\rangle = \int_{-\infty}^{\infty} \mathrm{d}x\ \psi_1^*(x)\,\psi_2(x)\ , \tag{2.27}$$

but for our purposes the exact definition of the inner product is immaterial.

According to quantum mechanics, the absolute value of the inner product is to be interpreted as a probability. More explicitly, consider the state described by $|\,\psi\rangle$. The probability that a measurement will establish the system to be in the state $|\,\varphi\rangle$ is given by $|\langle\varphi\,|\,\psi\rangle|^2$. Now subject the system, including the measurement device, to a rotation. According to (2.17), the states will change into

$$|\,\psi\rangle \rightarrow D\,|\,\psi\rangle\ , \qquad |\,\varphi\rangle \rightarrow D\,|\,\varphi\rangle\ . \tag{2.28}$$

The corresponding change of the inner product is then

$$\langle\varphi\,|\,\psi\rangle \longrightarrow \langle\varphi\,|\,D^\dagger D\,|\,\psi\rangle\ . \tag{2.29}$$

However, if nature is invariant under rotations, the probability described by the inner product, should *not* change under rotations. The two inner products in (2.29) must be equal. Since this equality must hold for all possible pairs of states $|\,\psi\rangle$ and $|\,\varphi\rangle$, we can conclude that the matrices themselves must obey the following condition:

$$D^\dagger D = \mathbf{1}\ , \tag{2.30}$$

in other words, D must be a *unitary* matrix.[5] Since this has to hold for every matrix $D(R)$ associated to a rotation, this demand should hold for the entire representation. Thus, in this context, we shall be exclusively interested in *unitary* representations.

[5]The condition is that the absolute value of the inner product should not change, so one might suspect that it suffices to constrain $D^\dagger D$ to be equal to unity apart from a phase factor. However, $D^\dagger D$ is a hermitian, positive definite matrix, so we must conclude that this phase factor can only be equal to 1.

3. The group of rotations in three dimensions

A rotation in three-dimensional space can be represented by a 3×3 matrix of real numbers. Since upon a rotation of a set of vectors, the angles between them remain the same, the matrix in question will be *orthogonal*. These orthogonal matrices form a group, called $O(3)$. From the demand $R\tilde{R} = \mathbf{1}$, one derives that $\det(R) = \pm 1$. If we restrict ourselves to the orthogonal matrices with $\det(R) = +1$, then we call the group $SO(3)$, the *special orthogonal group in 3 dimensions*.

A rotation in three-dimensional space is completely determined by the rotation axis and the angle over which we rotate. The rotation axis can for instance be specified by a three-dimensional vector $\vec{\alpha}$; the length of this vector can then be chosen to be equal to the angle over which we rotate (in radians). Since rotations over angles that differ by a multiple of 2π, are identical, we can limit ourselves to rotation axis vectors $\vec{\alpha}$ inside (or on the surface of) a three-dimensional sphere with radius π. This gives us a natural parametrization for all rotations. Every point in this sphere of parameters corresponds to a possible rotation: the rotation axis is given by the line through this point and the center of the sphere, and the angle over which we rotate (according to a left-handed screw for instance) varies from 0 to π (rotations over angles between $-\pi$ and 0 are then associated with the vector in the opposite direction). Two opposite points on the surface of the sphere, that is, $\vec{\alpha}$ and $-\vec{\alpha}$ with $|\vec{\alpha}| = \pi$, describe the same rotation, one over an angle π and one over an angle $-\pi$, around the same axis of rotation. However, apart from this identification of diametrically opposed points on the surface of the sphere, two different points inside this parameter sphere always describe two different rotations.

From the above, it is clear that rotations can be parameterized in terms of three independent parameters, being the three components of the vectors $\vec{\alpha}$, and furthermore that the rotations depend on these parameters in a continuous fashion. To study this dependence further, consider infinitesimal rotations, or, rotations corresponding to vectors $|\vec{\alpha}| \approx 0$. First, let us limit ourselves to rotations around the z axis, so that $\vec{\alpha} = (0, 0, \alpha)$. The associated rotation follows from

$$\begin{aligned} x &\rightarrow \cos\alpha\, x + \sin\alpha\, y \ , \\ y &\rightarrow \cos\alpha\, y - \sin\alpha\, x \ , \\ z &\rightarrow z \ . \end{aligned} \tag{3.1}$$

This leads to a matrix $R(\alpha)$, equal to

$$R(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \ . \tag{3.2}$$

The rotation by an angle α can also be regarded as being the result of n successive rotations over an angle α/n. For very large values of n, the rotation by a small angle α/n will differ from the identity only infinitesimally; ignoring terms of order $(\alpha/n)^2$, we find for the associated 3×3 matrix,

$$\begin{aligned} R(\alpha/n) &= \begin{pmatrix} 1 & \alpha/n & 0 \\ -\alpha/n & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \mathrm{O}\left(\frac{\alpha^2}{n^2}\right) \\ &= \mathbf{1} + \frac{\alpha}{n} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \mathrm{O}\left(\frac{\alpha^2}{n^2}\right) \ . \end{aligned} \tag{3.3}$$

It is now possible to reconstruct the finite rotation over an angle α by taking the n^{th} power of (3.3),

$$R(\alpha) = [R(\alpha/n)]^n = \left[\mathbf{1} + \frac{\alpha}{n}T + \mathrm{O}\left(\frac{\alpha^2}{n^2}\right)\right]^n , \tag{3.4}$$

where the matrix T is given by

$$T = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} . \tag{3.5}$$

In the limit $n \to \infty$, we expect to be able to ignore terms of order $1/n^2$; furthermore, we make use of the formula

$$\mathrm{e}^A = \lim_{n\to\infty} \left(\mathbf{1} + \frac{1}{n}A\right)^n . \tag{3.6}$$

This results in

$$R(\alpha) = \exp(\alpha T) \ . \tag{3.7}$$

The exponent of this matrix can be elaborated by using the series expansion

$$\mathrm{e}^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n \ . \tag{3.8}$$

Next, we remark that

$$T^{2n} = (-)^n \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \qquad (n \geq 1) \tag{3.9}$$

from which it follows immediately that $T^{2n+1} = (-)^n T$ for $n \geq 0$. Using this, we can perform the exponentiation by separately selecting the even and odd powers. This leads to

$$\begin{aligned} \exp(\alpha T) &= \mathbf{1} + \sum_{n=1}^{\infty} \frac{(-)^n \alpha^{2n}}{(2n)!} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \sum_{n=0}^{\infty} \frac{(-)^n \alpha^{2n+1}}{(2n+1)!} T \\ &= \mathbf{1} + (\cos\alpha - 1) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \sin\alpha \ T \ , \end{aligned} \tag{3.10}$$

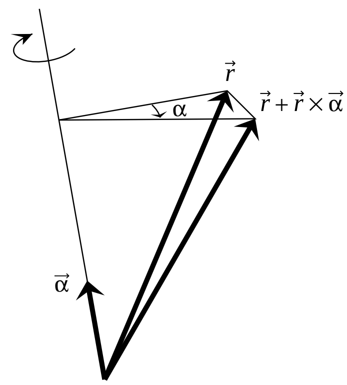


Figure 2: Infinitesimal rotation of a vector $\vec{r}$, around a rotation axis $\vec{\alpha}$

which indeed coincides with the original matrix (3.2).

Let us now consider the relation between finite and infinitesimal transformations as given by Eq. (3.7), for more general rotations. For rotations over a small angle, every $\vec{r}$ gets a small vector added to it that is orthogonal both to the vector $\vec{r}$ and the rotation axis (see Figure 2). This tiny vector is exactly equal to the outer product of $\vec{r}$ and the rotation axis vector $\vec{\alpha}$ (where it was assumed that $|\vec{\alpha}| \approx 0$), so that

$$\vec{r} \to \vec{r} + \vec{r} \times \vec{\alpha} + \mathrm{O}\left(|\vec{\alpha}|^2\right) \ . \tag{3.11}$$

therefore, in case of a general rotation axis vector $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ one can write

$$\begin{aligned} x &\to x + \alpha_3\, y - \alpha_2\, z + \mathrm{O}\left(|\vec{\alpha}|^2\right) \ , \\ y &\to y + \alpha_1\, z - \alpha_3\, x + \mathrm{O}\left(|\vec{\alpha}|^2\right) \ , \\ z &\to z + \alpha_2\, x - \alpha_1\, y + \mathrm{O}\left(|\vec{\alpha}|^2\right) \ . \end{aligned} \tag{3.12}$$

Infinitesimal rotations can therefore be written as follows:

$$R(\vec{\alpha}) = \mathbf{1} + i\Big(\alpha_1 L_1 + \alpha_2 L_2 + \alpha_3 L_3\Big) + \mathrm{O}\left(|\vec{\alpha}|^2\right) \ , \tag{3.13}$$

where we added a factor i in order to conform to the usual notations, and the hermitian matrices L_1 , L_2 en L_3 are defined by

$$\begin{aligned} L_1 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \ , \\ L_2 &= \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \ , \\ L_3 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ . \end{aligned} \tag{3.14}$$

Above result can be compressed in one expression by using the completely skew-symmetric epsilon tensor,

$$(L_i)_{jk} = -i\epsilon_{ijk} \ . \tag{3.15}$$

Indeed, we can easily check that

$$\begin{aligned}(L_1)_{23} = -(L_1)_{32} = -i\epsilon_{123} = -i \ , \\ (L_2)_{31} = -(L_2)_{13} = -i\epsilon_{231} = -i \ , \\ (L_3)_{12} = -(L_3)_{21} = -i\epsilon_{312} = -i \ .\end{aligned} \tag{3.16}$$

Again, we can consider $R(\vec{\alpha})$ as being formed out of n successive rotations with rotation axis $\vec{\alpha}/n$,

$$\begin{aligned} R(\vec{\alpha}) &= [R(\vec{\alpha}/n)]^n \\ &= \left[\mathbf{1} + \frac{1}{n}\Big(i\alpha_1 L_1 + i\alpha_2 L_2 + i\alpha_3 L_3\Big) + \mathrm{O}\left(\frac{|\vec{\alpha}|^2}{n^2}\right)\right]^n \ . \end{aligned} \tag{3.17}$$

Employing (3.4), we find then the following expression in the limit $n \to \infty$,

$$R(\vec{\alpha}) = \exp\left(i\sum_k \alpha_k L_k\right) \ . \tag{3.18}$$

The correctness of Eq. (3.18) can be checked in a different way. First, we note that the following multiplication rule holds for rotations around one common axis of rotation, but with different rotation angles:

$$R(s\vec{\alpha})\, R(t\vec{\alpha}) = R((s+t)\vec{\alpha}) \ , \tag{3.19}$$

where s and t are real numbers. The rotations $R(s\vec{\alpha})$ with one common axis of rotation define a commuting subgroup of the complete rotation group. This is not difficult to see: The matrices $R(s\vec{\alpha})$ (with a fixed vector $\vec{\alpha}$ and a variable s) define a group, where the result of a multiplication does not depend on the order in the product,

$$R(s\vec{\alpha})\, R(t\vec{\alpha}) = R(t\vec{\alpha})\, R(s\vec{\alpha}) \ . \tag{3.20}$$

This subgroup is the group $SO(2)$, the group of the two-dimensional rotations (the axis of rotation stays the same under these rotations, only the components of a vector that are orthogonal to the axis of rotation are rotated). Using Eq. (3.19), we can simply deduce the following differential equation for $R(s\vec{\alpha})$,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}s} R(s\vec{\alpha}) &= \lim_{\Delta\to 0} \frac{R((s+\Delta)\vec{\alpha}) - R(s\vec{\alpha})}{\Delta} \\ &= \lim_{\Delta\to 0} \frac{R(\Delta\vec{\alpha}) - \mathbf{1}}{\Delta} R(s\vec{\alpha}) \\ &= \left(i\sum_k \alpha_k L_k\right) R(s\vec{\alpha}) \ , \end{aligned} \tag{3.21}$$

where first Eq. (3.19) was used, and subsequently (3.13). Now it is easy to verify that the solution of this differential equation is exactly given by Eq. (3.18).

Yet an other way to ascertain that the matrices (3.18) represent rotations, is to prove that these matrices are orthogonal and have determinant equal to 1, which means that the following relations are fulfilled

$$\widetilde{R(\vec{\alpha})} = [R(\vec{\alpha})]^{-1} = R(-\vec{\alpha}) \ , \qquad \det R(\vec{\alpha}) = 1 \ , \tag{3.22}$$

The proof follows from the following properties for a general matrix A (see also Appendix C),

$$\widetilde{(\mathrm{e}^{A})} = \mathrm{e}^{\widetilde{A}} \ , \qquad \det\left(\mathrm{e}^{A}\right) = \mathrm{e}^{\mathrm{Tr}\, A} \ . \tag{3.23}$$

From this, it follows that the matrices (3.18) obey Eqs. (3.22) provided that the matrix $i\sum_k \alpha_k L_k$ be real and skew-symmetric. This indeed turns out to be the case; from the definitions (3.15) it follows that $i\sum_k \alpha_k L_k$ in fact represents the most general real, and skew-symmetric 3×3 matrix.

The above question may actually be turned around: can *all* rotations be written in the form of Eq. (3.18)? The answer to this question is not quite so easy to give. In principle, the exponentiation in (3.18) can be performed explicitly *via* the power series expansion (3.8), and the result can be compared with the most general rotation matrix. It will turn out that the answer is affirmative: all rotations can indeed be written in the form of Eq. (3.18). This, however, is not the case for all groups. The so-called *non-compact* groups contain elements that cannot be written as a product of a finite number of such exponentials. These groups are called non-compact, because the volume of parameter space is non-compact. The rotation group, where all possible group elements are defined in terms of the parameters α_k that are restricted to the insides of a sphere with radius π, is a compact group. Within the frame of these lectures, non-compact groups will play no role, but such groups are not unimportant in physics. The Lorentz group, for example, which is the group consisting of all lorentz transformations, is an example of a non-compact group.

From the preceding discussion it will be clear that the matrices L_k, associated with the infinitesimal transformations, will be important, and at least for the compact groups, they will completely determine the group elements, by means of the exponentiation (3.18). This is why these matrices are called the *generators* of the group. Although our discussion was confined to the rotation group, the above can be applied to all *Lie groups*[6]: a group whose elements depend analytically on a finite number of parameters, in our case α_1, α_2, and α_3. In the case that the group elements take the form of matrices, this means that the matrix elements must be differentiable functions of the parameters.[7] The number of linearly independent parameters defines the *dimension* of the Lie group, not to be confused

[6]Named after the Norwegian mathematician Sophus Lie, 1842-1899

[7]This is clearly the case for the rotation group. In the general case, the above requirement can be somewhat weakened; for a general Lie group it suffices to require the elements as functions of the parameters to be twice differentiable.

with the dimension of the matrices considered.[8] The number of linearly independent generators must obviously be equal to the dimension of the group.

One of the most essential ingredients of a group, is its multiplication structure, according to which the product of two rotations $R(\vec{\alpha})$ and $R(\vec{\beta})$, again should be a rotation,

$$R(\vec{\alpha})\, R(\vec{\beta}) = R(\vec{\gamma}) \ , \tag{3.24}$$

where $\vec{\gamma}$ depends on $\vec{\alpha}$ and $\vec{\beta}$. The exact dependence fixes the multiplication structure of the group. The fact that such a vector function $\vec{\gamma}(\vec{\alpha},\vec{\beta})$ must exist, has implications for the product of generators. To derive these, we expand (3.24) in powers[9] of α en β,

$$\begin{aligned} \mathrm{e}^{i\vec{\alpha}\cdot\vec{L}}\, \mathrm{e}^{i\vec{\beta}\cdot\vec{L}} &= \Big(\mathbf{1} + i\alpha_k L_k + \mathrm{O}(\alpha^2)\Big)\Big(\mathbf{1} + i\beta_l L_l + \mathrm{O}(\beta^2)\Big) \\ &= \mathbf{1} + i(\alpha+\beta)_k\, L_k - \alpha_k\beta_l\, L_k L_l + \mathrm{O}(\alpha^2) + \mathrm{O}(\beta^2) \\ &= \mathbf{1} + i(\alpha+\beta)_k\, L_k - \tfrac{1}{2}(\alpha+\beta)_k(\alpha+\beta)_l\, L_k L_l \\ &\quad -\tfrac{1}{2}\alpha_k\beta_l\, [L_k, L_l] + \mathrm{O}(\alpha^2) + \mathrm{O}(\beta^2) \ . \end{aligned} \tag{3.25}$$

The first three terms are recognized as the beginning of the power series of $\exp(i(\vec{\alpha}+\vec{\beta})\cdot\vec{L})$. If the fourth term would vanish, that is, if the matrices L_k and L_l commute, then indeed $\gamma_k = \alpha_k + \beta_k$. However, it will turn out that the generators of the rotation group do *not* commute. Since it must be possible in any case to write the r.h.s. of the equation again in the form of the power series for $\exp(i\vec{\gamma}\cdot\vec{L})$, it must be possible to rewrite the commutators of the generators in terms of some linear combination of the generators. in other words, we must have

$$[L_i, L_j] = c_{ij}^k\, L_k \ , \tag{3.26}$$

where the constants c_{ij}^k are called the *structure constants* of the group, because they (nearly) completely determine the multiplication structure of the group. Note that, since the generators L_k are hermitian, the structure constants must be purely imaginary.

Before continuing, we first verify whether the generators (3.15) obey to the demand (3.26). After explicit matrix multiplications, we find this indeed to be the case:

$$[L_1, L_2] = iL_3 \ , \qquad [L_2, L_3] = iL_1 \ , \qquad [L_3, L_1] = iL_2 \ , \tag{3.27}$$

or,

$$[L_i, L_j] = i\epsilon_{ijk}\, L_k \ . \tag{3.28}$$

Making use of Eq. (3.26), we can now deduce the following result for $\vec{\gamma}(\vec{\alpha},\vec{\beta})$:

$$\gamma_k = \alpha_k + \beta_k + \tfrac{i}{2} c_{mn}^k\, \alpha_m\beta_n + \mathrm{O}(\alpha^2) + \mathrm{O}(\beta^2) \ . \tag{3.29}$$

[8] For the rotation group in three dimensions the dimension of the group and that of the matrices are both 3, but this is a coincidence: the dimension of the rotation group in d dimensions is $\frac{1}{2}d(d-1)$.

[9] The notation $\vec{\alpha}\cdot\vec{L}$ is here intended to mean $\alpha_1 L_1 + \alpha_2 L_2 + \alpha_3 L_3$. In Eq. (3.25) we also used summation convention: if in one term of an expression an index occurs twice, this means that it is summed over, even if the summation sign is not explicitly shown. So, $\alpha_k L_k \equiv \sum_k \alpha_k L_k$. From now on, this convention will be frequently used.

In principle, the higher order contributions can be determined by means of iteration; for example, we find

$$\gamma_k = \alpha_k + \beta_k + \tfrac{i}{2} c^k_{mn}\, \alpha_m \beta_n - \tfrac{1}{12}\left(\alpha_m \alpha_n \beta_p + \beta_m \beta_n \alpha_p\right) c^k_{mq} c^q_{np} + \cdots \ . \tag{3.30}$$

The fact that all terms in this iteration can be expressed in terms of the structure constants follows from the Campbell-Baker-Hausdorff formula, which expresses the logarithm of $(\exp A\, \exp B)$ in terms of a power series consisting exclusively of repeated commutators of the matrices A and B. Thus, the multiplication structure of the group is determined by the structure constants (at least for all those group elements that reside in some finite domain in the neighborhood of the identity). The CBH formula is explained in Appendix D.

Imagine that we can find matrices A_k, different from the matrices L_k, obeying the same commutation relations (3.26) as the L_k. In that case, by means of exponentiation, we can determine the corresponding group elements, which will have the same multiplication rules as the elements of the original group. In other words, we find a representation of the group this way. On the other hand, for every representation of the group, we can construct the corresponding generators, using the infinitesimal transformations, and they will obey the same commutation rules (3.26), with the same structure constants. Thus, we have found a direct relation between group representations and the matrix relations (3.26) (In more mathematical terms: the generators L_k, together with the commutation relations (3.26), define an algebra, called the *Lie algebra*. Matrices A_k with the same commutation relations then define a *representation* of the Lie algebra.)

One can easily check that the structure constants also must obey certain relations. This follows from the so-called *Jacobi identity*, which holds for any triple of matrices A, B and C,

$$[[A,B]\,,C] + [[B,C]\,,A] + [[C,A]\,,B] = 0 \ . \tag{3.31}$$

This identity can be proven by explicitly writing the commutators and using the associativity of the multiplication (See chapter 2); one then obtains 12 terms that cancel out pairwise. Using the Jacobi identity with $A = L_i$, $B = L_j$ en $C = L_k$, we deduce the following equation for the structure constants,

$$c^m_{ij}\, c^n_{mk} + c^m_{jk}\, c^n_{mi} + c^m_{ki}\, c^n_{mj} = 0 \ , \tag{3.32}$$

where use was made of (3.26). The equation (3.32) is also called the Jacobi identity. For the rotation group, this implies the following equation for the ϵ-tensors:

$$\epsilon_{ijm}\, \epsilon_{mkn} + \epsilon_{jkm}\, \epsilon_{min} + \epsilon_{kim}\, \epsilon_{mjn} = 0 \ , \tag{3.33}$$

which will be frequently used later. The validity of Eq. (3.33) can be derived directly from the identity

$$\epsilon_{ijm}\, \epsilon_{mkl} = \delta_{ik}\, \delta_{jl} - \delta_{il}\, \delta_{jk} \ , \tag{3.34}$$

which is easy to prove (for instance by choosing a couple of values for two of the indices).

Equation (3.32) has another consequence. Let us define n $n \times n$ matrices C_i according to

$$(C_i)_j{}^k \equiv -c_{ij}^k \; , \tag{3.35}$$

where n is the dimension of the Lie group. We can then write (3.32) as

$$(c_{ij}^m\, C_m)_k{}^n + (C_j C_i)_k{}^n - (C_i C_j)_k{}^n = 0 \; , \qquad \text{or} \qquad C_i\, C_j - C_j\, C_i = c_{ij}^k\, C_k \; . \tag{3.36}$$

These are exactly the same commutation relations as the ones we used to define the structure constants, in Eq. (3.26). The matrices C_i thus define a representation of the Lie algebra based on (3.26). Through exponentiation of the matrices C_i, we can then define a group with the same multiplication properties (at least in some finite region surrounding the identity) as the original Lie group, consisting of $n \times n$ matrices, where n is the dimension of the Lie group. This representation is called the *adjoint representation*.

Applying the above to the case of the rotation group leads to something of a disappointment. Since in this case $c_{ij}^k = i\epsilon_{ijk}$, the matrices C_i are simply equal to the matrices L_i (see Eq. (3.15), and so we recovered the original three-dimensional rotations. The adjoint representation thus coincides with the original group. This, however, is rather the exception than the rule, as will be seen later.

4. More about representations

In the previous chapter the properties of the group of three-dimensional rotations were discussed. Now, we return to the representations of this group. First, we note that, starting from a given representation, for instance by the matrices D acting on the wave functions that we combined in a column vector ψ, we can obtain an other representation, by constructing an other vector ψ. For instance, rearrange ψ in wave functions $\hat{\psi}$ according to

$$\hat{\psi} = U\psi \ . \tag{4.1}$$

Under rotations, $\hat{\psi}$ then transforms according to

$$\hat{\psi} \to \hat{\psi}' = \hat{D}\hat{\psi} \ , \tag{4.2}$$

where $\hat{D}$ is given by

$$\hat{D} = U D\, U^{-1} \ . \tag{4.3}$$

Both the original matrices D and the matrices $\hat{D}$ define a representation of the rotation group, but such representations will not be considered as fundamentally different. This is why representations that are related according to (4.3), are called *equivalent representations*. This allows us to formulate an important result in representation theory:

All finite dimensional representations of finite or compact groups are unitary.

With this we mean that all representations can be chosen to be unitary *via* a redefinition (4.3), so that all matrices D belonging to the representation obey $D^\dagger = D^{-1}$. We will not prove this here.

Up to here, we have primarily discussed one special representation of the group of rotations, being the representation defined by rotating the three-dimensional vector $\vec{x} = (x_1, x_2, x_3)$. There is an easy way to construct larger representations: just consider two vectors, $\vec{x}$ and $\vec{y}$, both transforming the usual way under rotations. Together, they form a six-dimensional vector $\vec{z} = (x_1, x_2, x_3, y_1, y_2, y_3)$, transforming under rotations as

$$\vec{z} \to \vec{z}\,' = D\vec{z} \ , \tag{4.4}$$

where the matrix D can be decomposed in 3×3 matrices in the following way:

$$D = \begin{pmatrix} R & 0 \\ 0 & R \end{pmatrix} \ . \tag{4.5}$$

Such a representation is called *reducible*, because the six-dimensional space can be split up in two invariant three-dimensional subspaces. This reducible six-dimensional representation can therefore be regarded as the *direct sum* of two three-dimensional representations, and we write

$$\mathbf{6} = \mathbf{3} \oplus \mathbf{3} \ . \tag{4.6}$$

The sum representation can occur if we consider a particle that can be in a superposition of two different kinds of quantum states.

It will be clear that representations that do *not* leave any subspace invariant, and therefore cannot be described in a block diagonal form such as in Eq. (4.5), are considered to be *irreducible representations*.

Other representations can be obtained by constructing so-called *product representations*. Consider for instance a system of two (free) particles with wave functions $\psi_1(\vec{x})$ and $\psi_2(\vec{y})$, where $\vec{x}$ and $\vec{y}$ are the coordinates of the particles. The wave functions $\Psi(\vec{x},\vec{y})$ of the combined system then consist of all possible products of wave functions ψ_1 and ψ_2. We call this a *tensor product*, which is denoted by

$$\Psi = \psi_1 \otimes \psi_2 \ . \tag{4.7}$$

Under rotations of both $\vec{x}$ and $\vec{y}$, this Ψ transforms accordingly, but the corresponding representation is more complicated than the ones associated to the separate wave functions ψ_1 and ψ_2. Often, such a product representation is not irreducible, and can be decomposed into a number of distinct representations that are irreducible. Let us demonstrate this phenomenon first in the following example. Let three possible functions ψ_i^1 be given by the coordinates x_i and three possible functions ψ_j^2 by the coordinates y_j. Thus, both the ψ_i^1's and the ψ_j^2's transform according to the three-dimensional representation of the rotation group. The product representation works on all possible products of ψ_i^1 and ψ_j^2, and therefore we can distinguish nine independent functions,

$$T_{ij}(\vec{x},\vec{y}) = x_i\, y_j \ , \tag{4.8}$$

transforming under rotations as

$$T_{ij} \to T'_{ij} = R_{ii'}R_{jj'}\, T_{i'j'} \ . \tag{4.9}$$

This nine-dimensional representation however is not irreducible. For instance, the symmetric part and the skew-symmetric part of T_{ij}, defined by $T_{(ij)} \equiv \frac{1}{2}(T_{ij}+T_{ji})$, and $T_{[ij]} \equiv \frac{1}{2}(T_{ij}-T_{ji})$, transform separately and independently under rotations. This follows directly by restricting ourselves only to the (skew-)symmetric part of T'_{ij}, and observing that the (anti)symmetry in i and j of $\frac{1}{2}(R_{ii'}R_{jj'} \pm R_{ji'}R_{ij'})$ implies the (anti)symmetry in i' en j'. This is why we write

$$T_{(ij)} \to T'_{(ij)} = R_{ii'}R_{jj'}\, T_{(i'j')} \ , \qquad T_{[ij]} \to T'_{[ij]} = R_{ii'}R_{jj'}\, T_{[i'j']} \ . \tag{4.10}$$

The skew-symmetric part of T_{ij} contains three independent components, transforming as a three-dimensional representation of the rotation group. The symmetric part of T_{ij} contains the remaining six components, which however do not transform as an irreducible transformation. This follows immediately from the fact that the trace of T_{ij} is equal to

$$T_{ii} = \vec{x}\cdot\vec{y} \ , \tag{4.11}$$

and therefore invariant under rotations. We must conclude that T_{ij} can be decomposed in three independent tensors[10],

$$T_{ij} \to \left\{ \begin{array}{l} T = \vec{x} \cdot \vec{y} \\ \\ T_i = \epsilon_{ijk} x_j y_k \\ \\ S_{ij} = x_i y_j + x_j y_i - \frac{2}{3} \delta_{ij} (\vec{x} \cdot \vec{y}) \end{array} \right\} \; . \tag{4.12}$$

Note that we used the epsilon symbol to describe the skew-symmetric part of T_{ij} again as a three-dimensional vector $\vec{T}$ (it is nothing but the outer product $\vec{x} \times \vec{y}$). Furthermore, we made the symmetric part S_{ij} traceless by adding an extra term proportional to δ_{ij}. The consequence of this is that S_{ij} consists of only *five* independent components. Under rotations, the terms listed above transform into expressions of the same type; the five independent components of S_{ij} transform into one another.[11] In short, the product of two three-dimensional representations can be written as

$$\mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus \mathbf{3} \oplus \mathbf{5} \; , \tag{4.13}$$

where the representations are characterized by their dimensions (temporarily ignoring the fact that inequivalent irreducible representations might exist with equal numbers of dimensions; they don't here, as we will see later).

The procedure followed in this example, rests on two features; first, we use that the symmetry properties of tensors do not change under the transformations, and secondly we make use of the existence of two invariant tensors, to wit:

$$T_{ij} = \delta_{ij} \; , \qquad T_{ijk} = \epsilon_{ijk} \; . \tag{4.14}$$

An invariant tensor is a tensor that does not change at all under the group transformations, as they act according to the index structure of the tensor, so that

$$T_{ijk\cdots} \to T'_{ijk\cdots} = R_{ii'} R_{jj'} R_{kk'} \cdots \; T_{i'j'k'\cdots} = T_{ijk\cdots} \; . \tag{4.15}$$

Indeed, both tensors δ_{ij} and ϵ_{ijk} obey (4.15), since the equation

$$R_{ii'} R_{jj'} \, \delta_{i'j'} = \delta_{ij} \tag{4.16}$$

is fulfilled because the R_{ij} are orthogonal matrices, and

$$R_{ii'} R_{jj'} R_{kk'} \; \epsilon_{i'j'k'} = \det R \, \epsilon_{ijk} = \epsilon_{ijk} \tag{4.17}$$

[10] In the second equation, again summation convention is used, see an earlier footnote.

[11] For each of these representations, we can indicate the matrices $D(R)$ that are defined in chapter 2. For the first representation, we have that $D(R) = 1$. In the second representation, we have 3×3 matrices $D(R)$ equal to the matrix R. For the third representation, we have 5×5 matrices $D(R)$. The indices of this correspond to the symmetric, traceless index pairs ij. The matrices $D(R)$ can be written as

$$D(R)_{(ij)\,(kl)} = \frac{1}{2} \left(R_{ik} \, R_{jl} + R_{il} \, R_{jk} \right) - \frac{1}{3} \delta_{ij} \, \delta_{kl} \; .$$

holds because the rotation matrices R_{ij} have $\det R = 1$. For every given tensor $T_{ijk\cdots}$ we can contract the indices using invariant tensors. It is then evident that tensors contracted that way span invariant subspaces, in other words, under rotations they will transform into tensors that are formed the same way. For example, let $T_{ijk\cdots}$ be a tensor transforming like

$$T_{ijk\cdots} \to T'_{ijk\cdots} = R_{ii'} R_{jj'} R_{kk'} \cdots \; T_{i'j'k'\cdots} \; . \tag{4.18}$$

Now, form the tensor

$$\hat{T}_{klm\cdots} \equiv \delta_{ij} \, T_{ijklm\cdots} \tag{4.19}$$

which has two indices less. By using Eq, (4.16), it is now easy to check that $\hat{T}$ transforms as

$$\hat{T}_{klm\cdots} \to \hat{T}'_{klm\cdots} = R_{kk'} R_{ll'} R_{mm'} \cdots \; \hat{T}_{k'l'm'\cdots} \; , \tag{4.20}$$

and, in a similar way, we can verify that contractions with one or more δ and ϵ tensors, produce tensors that span invariant subspaces. Using the example discussed earlier, we can write the expansion as

$$T_{ij} = \frac{1}{2} \epsilon_{ijk} \left(\epsilon_{klm} \, T_{lm} \right) + \frac{1}{2} \left(T_{ij} + T_{ji} - \frac{2}{3} \delta_{ij} T_{kk} \right) + \frac{1}{3} \delta_{ij} \, T_{kk} \; , \tag{4.21}$$

where the first term can also be written as $\frac{1}{2}(T_{ij} - T_{ji})$, by using the identity (3.34),

$$\epsilon_{ijk} \, \epsilon_{klm} = \delta_{il} \delta_{jm} - \delta_{im} \delta_{jl} \; , \tag{4.22}$$

and the second term in (4.21) is constructed in such a way that it is traceless:

$$\delta_{ij} \left(T_{ij} + T_{ji} - \frac{2}{3} \delta_{ij} T_{kk} \right) = 0 \; . \tag{4.23}$$

5. Ladder operators

Let us consider a representation of the rotation group, generated by hermitian matrices I_1 , I_2 and I_3 , which obey the same commutation rules as L_1 , L_2 and L_3 , given in Eq. (3.15),

$$[I_1, I_2] = iI_3 \ , \qquad [I_2, I_3] = iI_1 \ , \qquad [I_3, I_1] = iI_2 \ , \tag{5.1}$$

or in shorthand:

$$[I_i, I_j] = i\epsilon_{ijk}\, I_k \ . \tag{5.2}$$

We demand the matrices $\exp(i\alpha_k I_k)$ to be unitary; therefore, the I_i are hermitian: $I_i^\dagger = I_i$. Starting from this information, we now wish to determine all sets of irreducible matrices I_i with these properties. This is the way to determine all (finite-dimensional, unitary) representations of the group of rotations in three dimensions.

To this end, first define the linear combinations

$$I_\pm = I_1 \pm iI_2 \ , \tag{5.3}$$

so that $(I_\pm)^\dagger = I_\mp$, and

$$[I_3, I_\pm] = [I_3, I_1] \pm i[I_3, I_2] = iI_2 \pm I_1 = \pm I_\pm \ . \tag{5.4}$$

So we have for any state $\mid \psi\rangle$,

$$I_3 \Big(I_+ \mid \psi\rangle \Big) = I_+(I_3 + 1) \mid \psi\rangle \ . \tag{5.5}$$

A *Casimir operator* is a combination of operators for a representation constructed in such a way that it commutes with all generators. *Schur's lemma* states the following: if and only if the representation is irreducible, every Casimir operator will be a multiple of the unit matrix.

In the case of the three-dimensional rotations, we have such a Casimir operator:

$$\vec{I}^{\,2} \equiv I_1^2 + I_2^2 + I_3^2 \ . \tag{5.6}$$

We derive from Eq. (5.1):

$$[\vec{I}^{\,2}, I_1] \ = \ [\vec{I}^{\,2}, I_2] \ = \ [\vec{I}^{\,2}, I_3] \ = \ 0 \ . \tag{5.7}$$

Since $\vec{I}^{\,2}$ en I_3 are two commuting matrices, we can find a basis of states such that $\vec{I}^{\,2}$ and I_3 both at the same time take a diagonal form, with real eigenvalues. Furthermore, the eigenvalues of $\vec{I}^{\,2}$ must be positive (or zero), because we have

$$\langle\psi \mid \vec{I}^{\,2} \mid \psi\rangle = \Big| I_1 \mid \psi\rangle \Big|^2 + \Big| I_2 \mid \psi\rangle \Big|^2 + \Big| I_3 \mid \psi\rangle \Big|^2 \geq 0 \ . \tag{5.8}$$

It will turn out to be convenient to write the eigenvalues of $\vec{I}^2$ as $\ell(\ell+1)$, where $\ell \geq 0$ (The reason for this strange expression will become clear shortly; for the time being, consider this merely as a notation).

Now, consider a state $\mid \ell, m\rangle$ that is an eigenstate of $\vec{I}^{\,2}$ and I_3 , with eigenvalues $\ell(\ell+1)$ and m ,

$$\vec{I}^{\,2} \mid \ell, m\rangle = \ell(\ell+1) \mid \ell, m\rangle \ , \qquad I_3 \mid \ell, m\rangle = m \mid \ell, m\rangle \ . \tag{5.9}$$

From Eqs. (5.5) and (5.7), one derives that

$$\begin{aligned} I_3\Big(I_+ \mid \ell, m\rangle\Big) &= (m+1)\Big(I_+ \mid \ell, m\rangle\Big) \ , \\ \vec{I}^2\Big(I_+ \mid \ell, m\rangle\Big) &= \ell(\ell+1)\Big(I_+ \mid \ell, m\rangle\Big) \ . \end{aligned} \tag{5.10}$$

Substituting $I_+ \mid \ell, m\rangle = \mid \psi\rangle$, we have

$$I_3 \mid \psi\rangle = (m+1) \mid \psi\rangle \ , \qquad \vec{I}^2 \mid \psi\rangle = \ell(\ell+1) \mid \psi\rangle \ , \tag{5.11}$$

in other words, $\mid \psi\rangle$ is a new eigenvector of I_3 and $\vec{I}^2$ with eigenvalues $m' = m+1$, and $\ell' = \ell$, unless

$$\mid \psi\rangle \equiv I_+ \mid \ell, m\rangle \stackrel{?}{=} 0 \ . \tag{5.12}$$

Furthermore, we find

$$\begin{aligned} \langle\psi \mid \psi\rangle &= \langle \ell, m \mid I^- I^+ \mid \ell, m\rangle \\ &= \langle \ell, m \mid I_1^2 + I_2^2 + i[I_1, I_2] \mid \ell, m\rangle \\ &= \langle \ell, m \mid I_1^2 + I_2^2 - I_3 \mid \ell, m\rangle \\ &= \langle \ell, m \mid \vec{I}^{\,2} - I_3(I_3+1) \mid \ell, m\rangle \ , \end{aligned} \tag{5.13}$$

where we made use of: $I_+^\dagger = I_-$. And so, using Eq. (5.9), we find

$$\langle\psi \mid \psi\rangle = \Big(\ell(\ell+1) - m(m+1)\Big) \langle \ell, m \mid \ell, m\rangle \ . \tag{5.14}$$

If we now assume that $\mid \ell, m\rangle$ is a normalized state (so, $\langle \ell, m \mid \ell, m\rangle = 1$), then $\mid \psi\rangle$ can be written as a normalized state $\mid \ell, m+1\rangle$ multiplied by a proportionality factor that is given by (5.14). This factor is fixed up to a phase factor, which we absorb in the definition of $\mid \ell, m+1\rangle$. This way, we conclude that

$$I_+ \mid \ell, m\rangle = \sqrt{\ell(\ell+1) - m(m+1)} \ \mid \ell, m+1\rangle \ . \tag{5.15}$$

Repeating this procedure, the operator I_+ produces states with ever increasing eigenvalues of I_3 :

$$\mid \ell, m\rangle \xrightarrow{I_+} \mid \ell, m+1\rangle \xrightarrow{I_+} \mid \ell, m+2\rangle \xrightarrow{I_+} \mid \ell, m+3\rangle \xrightarrow{I_+} \text{ etc}\,. \tag{5.16}$$

This is why I_+ will be called "ladder operator" or "step operator". However, we are interested in finite matrices I_i , and this implies that the series (5.16) has to come to an end somewhere. According to Eq. (5.15), this only happens if, in the series (5.16), a state emerges for which the eigenalue m of I_3 equals ℓ. This, in turn, requires that the original eigenvalue m of the state we started off with, differs from ℓ by an integer. The necessity of this in fact already follows from Eq. (5.14): since $\langle\psi \mid \psi\rangle$ and $\langle \ell, m \mid \ell, m\rangle$ must have non negative norms, one must have $\ell(\ell+1) - m(m+1) \geq 0$, and also $-\ell - 1 \leq m \leq \ell$. In order to ensure that the series (5.15) terminates, as soon as m approaches values greater than its allowed limit, we must demand that $\ell - m$ be a positive integer. therefore, we find

$$\mid \ell, m\rangle \xrightarrow{I_+} \mid \ell, m+1\rangle \xrightarrow{I_+} \cdots\cdots \xrightarrow{I_+} \mid \ell, \ell\rangle \ , \tag{5.17}$$

where the vector $\mid \ell, \ell\rangle$ with the highest eigenvalue of I_3 obeys

$$I_+ \mid \ell, \ell\rangle = 0 \ . \tag{5.18}$$

It is now easy to continue by observing that the matrix I_- is also a ladder operator, but one generating lower eigenvalues of I_3. Starting from a state $\mid \ell, m\rangle$, we can construct states with decreasing eigenvalues of I_3:

$$\text{etc.} \xleftarrow{I_-} \mid \ell, m-3\rangle \xleftarrow{I_-} \mid \ell, m-2\rangle \xleftarrow{I_-} \mid \ell, m-1\rangle \xleftarrow{I_-} \mid \ell, m\rangle \tag{5.19}$$

Repeating the same manipulations as the ones for I_+, shows that for $\mid \psi\rangle = I_- \mid \ell, m\rangle$,

$$\langle\psi \mid \psi\rangle = \Big[\ell(\ell+1) - m(m-1)\Big] \langle \ell, m \mid \ell, m\rangle \ , \tag{5.20}$$

so it follows that we must have $\ell(\ell+1) - m(m-1) \geq 0$, and subsequently $\ell(\ell+1) - m(m-1) \geq 0$, that is, $-\ell \leq m \leq \ell + 1$. Since we must require the series (5.19) to terminate as well, there must be a state in the series with minimal eigenvalue $m = -\ell$, which guarantees that

$$I_- \mid \ell, -\ell\rangle = 0 \ . \tag{5.21}$$

Again, we encounter an undetermined phase factor. It seems that we have the freedom to choose it any way we like, so again we fix the phase factor to be $+1$, but we return to this phase factor shortly:

$$I_- \mid \ell, m\rangle = \sqrt{\ell(\ell+1) - m(m-1)} \mid \ell, m-1\rangle \ . \tag{5.22}$$

Starting from a given state $\mid \ell, m\rangle$, we now have constructed $\ell - m$ states with eigenvalues $m+1,\ m+2,\ \ldots,\ \ell$ and $\ell + m$ states with I_3 eigenvalues $m-1,\ m-2,\ \ldots\ , -\ell$. Thus, in total we found $1 + (\ell - m) + (\ell + m) = 2\ell + 1$ states. This is why $2\ell + 1$ must be an integral number, so that ℓ, and therefore also m, are either both integers or both an integer plus $\frac{1}{2}$.

Above arguments do not quite suffice to prove that we indeed found all states. In principle, it might be possible to apply arbitrary sequences of I_+ and I_- operators, to find many more states. Suppose we apply I_+ and subsequently I_-. We get a state with the same values of both ℓ and m as before. But is this the same state? Indeed, the answer is yes — and also the phase is +1 ! Note that

$$I_- \, I_+ = I_1^2 + I_2^2 + i(I_1 \, I_2 - I_2 \, I_1) = (\vec{I})^2 - I_3^2 - I_3 = \Big(\ell(\ell+1) - m(m+1)\Big)\mathbf{1} \ . \quad (5.23)$$

This ensures that, if we apply (5.15) and (5.22) in succession, we get back exactly the same state as the one we started off with (correctly normalized, and with a phase factor +1).

By way of exercise, we verify that the operators I_+, I_- and I_3 exclusively act on this single series of states $\mid \ell, m\rangle$ as prescribed by Eqs. (5.9), (5.15), and (5.22). Checking the commutation rules,

$$[I_3, I_\pm,] = \pm I_\pm \ , \qquad [I_+, I_-] = 2I_3 \ , \quad (5.24)$$

we indeed find

$$\begin{aligned} (I_3 I_\pm - I_\pm I_3) \mid \ell, m\rangle &= (m\pm 1)\sqrt{\ell(\ell+1) - m(m\pm 1)} \mid \ell, m\pm 1\rangle \\ &\quad - m\sqrt{\ell(\ell+1) - m(m\pm 1)} \mid \ell, m\pm 1\rangle \\ &= \pm\sqrt{\ell(\ell+1) - m(m\pm 1)} \mid \ell, m\pm 1\rangle \\ &= \pm I_\pm \mid \ell, m\rangle \ , \end{aligned} \quad (5.25)$$

$$\begin{aligned} (I_+ I_- - I_- I_+) \mid \ell, m\rangle &= \sqrt{\ell(\ell+1) - (m-1)m}\ \sqrt{\ell(\ell+1) - (m-1)m} \mid \ell, m\rangle \\ &\quad - \sqrt{\ell(\ell+1) - (m+1)m}\ \sqrt{\ell(\ell+1) - (m+1)m} \mid \ell, m\rangle \\ &= 2m \mid \ell, m\rangle \\ &= 2I_3 \mid \ell, m\rangle \ . \end{aligned} \quad (5.26)$$

Summarizing, we found that an irreducible representation of I_1 , I_2 , I_3 can be characterized by a number ℓ , and it acts on a space spanned by $2\ell+1$ states $\mid \ell, m\rangle$ for which

$$\begin{aligned} \vec{I}^2 \mid \ell, m\rangle &= \ell(\ell+1) \mid \ell, m\rangle \ , \\ I_3 \mid \ell, m\rangle &= m \mid \ell, m\rangle \ , \\ I_\pm \mid \ell, m\rangle &= \sqrt{\ell(\ell+1) - m(m\pm 1)} \mid \ell, m\pm 1\rangle \ , \end{aligned} \quad (5.27)$$

with $m = -\ell,\ -\ell+1,\ ,-\ell+2,\ \cdots,\ \ell-2,\ \ell-1,\ \ell$. Either both ℓ and m are integers, or they are both integers plus $\frac{1}{2}$. Of course, we always have $I_1 = \frac{1}{2}(I_+ + I_-)$ and $I_2 = \frac{1}{2i}(I_+ - I_-)$.

We now provide some examples, being the representations for $\ell = 0,\ \frac{1}{2},\ 1$, and $\frac{3}{2}$:

- For $\ell = 0$, we find the trivial representation. There is only one state, $|0, 0\rangle$, and $I_i|0, 0\rangle = 0$ for $i = 1, 2, 3$.

- For $\ell = \frac{1}{2}$, we find a two-dimensional representation. There are two basis elements, $|\,\frac{1}{2},\frac{1}{2}\rangle$ and $|\,\frac{1}{2},-\frac{1}{2}\rangle$, for which, according to Eq. (5.27), we have

$$\begin{aligned} I_+\,|\,\tfrac{1}{2},-\tfrac{1}{2}\rangle &= |\,\tfrac{1}{2},\tfrac{1}{2}\rangle\ , \\ I_+\,|\,\tfrac{1}{2},\tfrac{1}{2}\rangle &= 0\ , \\ I_-\,|\,\tfrac{1}{2},\tfrac{1}{2}\rangle &= |\,\tfrac{1}{2},-\tfrac{1}{2}\rangle\ , \\ I_-\,|\,\tfrac{1}{2},-\tfrac{1}{2}\rangle &= 0\ . \end{aligned} \tag{5.28}$$

This way, we find the matrices

$$I_3 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}\ , \qquad I_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\ , \qquad I_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\ . \tag{5.29}$$

The matrices I_1 , I_2 en I_3 following from this calculation, are the matrices $\frac{1}{2}\tau_i$ that will be introduced in Chapter 6.

- For $I = 1$ we find a three-dimensional representation. There are three basis elements, $|\,1,1\rangle$, $|\,1,0\rangle$ and $|\,1,-1\rangle$, for which, according to Eq. (5.27), we have

$$\begin{aligned} I_+\,|\,1,-1\rangle &= \sqrt{2}\,|\,1,0\rangle\ , \\ I_+\,|\,0\rangle &= \sqrt{2}\,|\,1,1\rangle\ , \\ I_+\,|\,1,1\rangle &= 0\ , \\ I_-\,|\,1,1\rangle &= \sqrt{2}\,|\,1,0\rangle\ , \\ I_-\,|\,1,0\rangle &= \sqrt{2}\,|\,1,-1\rangle\ , \\ I_-\,|\,-1,-1\rangle &= 0\ . \end{aligned} \tag{5.30}$$

This way, we find the matrices

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad I_+ = \begin{pmatrix} 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 \end{pmatrix}, \qquad I_- = \begin{pmatrix} 0 & 0 & 0 \\ \sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{pmatrix}\ . \tag{5.31}$$

The matrices I_1 , I_2 en I_3 are here equal to the matrices L_i , but in a different (complex) basis, where L_3 is diagonal.

- For $l = \frac{3}{2}$, we find a four dimensional representation. We have the basis elements $|\,\frac{3}{2},\frac{3}{2}\rangle$, $|\,\frac{3}{2},\frac{1}{2}\rangle$, $|\,\frac{3}{2},-\frac{1}{2}\rangle$ en $|\,\frac{3}{2},-\frac{3}{2}\rangle$, for which, according to Eq. (5.27),

$$\begin{aligned} I_+\,|\,\tfrac{3}{2},-\tfrac{3}{2}\rangle &= \sqrt{3}\,|\,\tfrac{3}{2},-\tfrac{1}{2}\rangle\ , \\ I_+\,|\,\tfrac{3}{2},-\tfrac{1}{2}\rangle &= 2\,|\,\tfrac{3}{2},\tfrac{1}{2}\rangle\ , \\ I_+\,|\,\tfrac{3}{2},\tfrac{1}{2}\rangle &= \sqrt{3}\,|\,\tfrac{3}{2},\tfrac{3}{2}\rangle\ , \\ I_+\,|\,\tfrac{3}{2},\tfrac{3}{2}\rangle &= 0\ . \end{aligned} \tag{5.32}$$

This way, we find the marices

$$I_3 = \begin{pmatrix} \frac{3}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & -\frac{3}{2} \end{pmatrix}, \qquad I_+ = \begin{pmatrix} 0 & \sqrt{3} & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & \sqrt{3} \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{5.33}$$

The matrix I_- can be derived in a similar way from Eq. (5.27), or can be obtained directly by hermitian conjugation: $I_- = I_+^\dagger$.

6. The group $SU(2)$

In Chapter 4, we only saw irreducible representations of the three-dimensional rotation group that were all odd dimensional. Chapter 5, however, showed the complete set of all irreducible representations of this group, and as many of them are even as there are odd ones. More understanding of the even-dimensional representations is needed. To this end, we subject the simplest example of these, the one with $\ell = \frac{1}{2}$, to a closer inspection. Clearly, we have vectors forming a two-dimensional space, which will be called *spinors*. Every rotation in a three-dimensional space must be associated to a unitary transformation in this spinor space. If $R = \exp(i\sum_k \alpha_k L_k)$, then the associated transformation X is written as $X = \exp(i\sum_k \alpha_k I_k)$, where the generators I_k follow from Eq. (5.29):

$$I_1 = \frac{I_+ + I_-}{2} = \tfrac{1}{2}\tau_1 \,, \qquad I_2 = \frac{I_+ - I_-}{2i} = \tfrac{1}{2}\tau_2 \,, \qquad I_3 = \tfrac{1}{2}\tau_3 \,. \tag{6.1}$$

Here, we have introduced the following three fundamental 2×2 matrices: [12]

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \qquad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \qquad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{6.2}$$

These τ-matrices obey the following product rules:

$$\tau_i\,\tau_j = \delta_{ij}\mathbf{1} + i\epsilon_{ijk}\,\tau_k \,, \tag{6.3}$$

as can easily be established. Since $[\tau_i,\,\tau_j] = \tau_i\,\tau_j - \tau_j\,\tau_i$, we find that the generators I_k indeed obey the correct commutation rules:

$$\left[\frac{\tau_i}{2},\frac{\tau_j}{2}\right] = i\epsilon_{ijk}\,\frac{\tau_k}{2} \,. \tag{6.4}$$

The three τ matrices are *hermitian* and *traceless*:

$$\tau_i = \tau_i^\dagger \,; \qquad \mathrm{Tr}\,(\tau_i) = 0 \,. \tag{6.5}$$

For rotations over tiny angles, $|\vec\alpha| \ll 1$, the associated matrix $X(\vec\alpha)$ takes the following form:

$$X(\vec\alpha) = \mathbf{1} + iB + \mathrm{O}(B^2) \,; \qquad B = \alpha_i\frac{\tau_i}{2} \,. \tag{6.6}$$

One readily verifies that $X(\vec\alpha)$ is unitary and that its determinant equals 1:

$$\begin{aligned} \big(\mathbf{1} + iB + \mathrm{O}(B^2)\big)^\dagger &= \big(\mathbf{1} + iB + \mathrm{O}(B^2)\big)^{-1} = \mathbf{1} - iB + \mathrm{O}(B^2) \,; \\ \det\big(\mathbf{1} + iB + \mathrm{O}(B^2)\big) &= 1 + i\,\mathrm{Tr}\,B + \mathrm{O}(B^2) = 1 \,, \end{aligned} \tag{6.7}$$

since

$$B^\dagger = B \,, \qquad \mathrm{Tr}\,B = 0 \,. \tag{6.8}$$

[12] Also called Pauli matrices, and often indicated as σ_i.

The finite transformation $X(\vec{\alpha})$ is found by exponentiation of (6.6), exactly in accordance with the limiting procedure displayed in Chapter 3:

$$X(\vec{\alpha}) = \lim_{n\to\infty} \left\{ \mathbf{1} + i\frac{\alpha_i}{n}\frac{\tau_i}{2} \right\}^n = \exp\left(i\alpha_i \frac{\tau_i}{2} \right) \ . \tag{6.9}$$

The matrices $\frac{1}{2}\tau_i$ are therefore the generators of the rotations for the $\ell = \frac{1}{2}$ representation. They do require the coefficients $\frac{1}{2}$ in order to obey exactly the same commutation rules as the generators L_i of the rotation group in three dimensions, see Eq. (6.4).

By making use of the product property of the τ-matrices, we can calculate the exponential expression for $X(\vec{\alpha})$. This is done as follows:

$$\begin{aligned} X(\vec{\alpha}) &= \mathrm{e}^{i\alpha_i\tau_i/2} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{i\alpha_j\tau_j}{2}\right)^n \\ &= \sum_{n=0}^{\infty} \frac{1}{(2n)!}\left(\frac{i\alpha_j\tau_j}{2}\right)^{2n} + \sum_{n=0}^{\infty} \frac{1}{(2n+1)!}\left(\frac{i\alpha_j\tau_j}{2}\right)^{2n+1} \ , \end{aligned} \tag{6.10}$$

where, in the last line, we do the summation over the even and the odd powers of $(i\alpha_j\tau_j)$ separately. Now we note that

$$(i\alpha_j\tau_j)^2 = -\alpha_j\alpha_k\,\tau_j\,\tau_k = -\alpha^2\,\mathbf{1}, \tag{6.11}$$

where use was made of Eq. (6.3), and α is defined as

$$\alpha = \sqrt{\alpha_1^2 + \alpha_2^2 + \alpha_3^2} \ . \tag{6.12}$$

From Eq. (6.11) it immediately follows that

$$(i\alpha_j\tau_j)^{2n} = (-)^n\,\alpha^{2n}\,\mathbf{1}, \qquad (i\alpha_j\tau_j)^{2n+1} = (-)^n\,\alpha^{2n}\,(i\alpha_j\tau_j) \ , \tag{6.13}$$

so that we can write Eq. (6.10) as

$$\begin{aligned} X(\vec{\alpha}) &= \left\{ \sum_{n=0}^{\infty} \frac{(-)^n}{(2n)!}\left(\frac{\alpha}{2}\right)^{2n} \right\} \mathbf{1} + \left\{ \sum_{n=0}^{\infty} \frac{(-)^n}{(2n+1)!}\left(\frac{\alpha}{2}\right)^{2n+1} \right\} \left(\frac{i\alpha_j\tau_j}{\alpha}\right) \\ &= \cos\frac{\alpha}{2}\,\mathbf{1} + i\sin\frac{\alpha}{2}\,\frac{\alpha_j\tau_j}{\alpha} \ . \end{aligned} \tag{6.14}$$

It so happens that every 2×2 matrix can be decomposed in the unit matrix $\mathbf{1}$ and τ_i:

$$X = c_0\,\mathbf{1} + ic_i\,\tau_i. \tag{6.15}$$

If we furthermore use the product rule (6.3) and Eq. (6.5), and also

$$\mathrm{Tr}\,(\mathbf{1}) = 2 \ , \tag{6.16}$$

the coefficients c_0 and c_i can be determined for every 2×2 matrix X :

$$c_0 = \tfrac{1}{2}\,\mathrm{Tr}\,(X)\ ; \qquad c_i = \tfrac{1}{2}\,\mathrm{Tr}\,(X\,\tau_i)\ . \tag{6.17}$$

In our case, we read off the coefficients c_0 and c_i directly from Eq. (6.14):

$$c_0 = \cos\frac{\alpha}{2}, \qquad c_i = \frac{\alpha_i}{\alpha}\,\sin\frac{\alpha}{2}. \tag{6.18}$$

It is clear that all these coefficients are real. Furthermore, we simply establish:

$$c_0^2 + c_i^2 = 1. \tag{6.19}$$

The expression (6.15) for $X(\vec{\alpha})$ can now also be written in terms of two complex parameters a en b ,

$$X = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}\ , \tag{6.20}$$

with $|a|^2 + |b|^2 = 1$. Matrices of the form (6.20) with generic a and b obeying $|a|^2 + |b|^2 = 1$ form the elements of the group $SU(2)$, the group of unitary 2×2 matrices with determinant 1, because [13] they obey:

$$X^\dagger = X^{-1}, \qquad \det X = 1\ . \tag{6.21}$$

It should be clear that these matrices form a group: if X_1 en X_2 both obey (6.21) and (6.20), then also $X_3 = X_1 X_2$ and so this matrix also is an element of the group. Furthermore, the unit matrix and the inverse matrix obey (6.20) en (6.21), so they also are in the group, while associativity for the multiplication is evident as well.

In chapter 3 we established that the rotations can be parameterized by vectors $\vec{\alpha}$ that lie in a sphere with radius $\alpha = \pi$. The direction of $\vec{\alpha}$ coincides with the axis of rotation, and its length α equals the angle of rotation. Since rotations over $+\pi$ and $-\pi$ radians are equal, we established that

$$R(\vec{\alpha}) = R(-\vec{\alpha})\ , \qquad \text{if} \qquad \alpha = \pi\ . \tag{6.22}$$

As we see in Eq. (6.14), the elements of $SU(2)$ can be parameterized by the same vectors $\vec{\alpha}$. However, to parameterize all elements $X(\vec{\alpha})$, the radius of the sphere must be taken to be *twice as large*, that is, equal to 2π . Again consider two vectors in opposite directions, $\vec{\alpha}$ and $\vec{\alpha}\,'$, in this sphere, such that the lengths $\alpha + \alpha' = 2\pi$, so that they yield the same rotation,

$$R(\vec{\alpha}\,') = R(\vec{\alpha})\ , \tag{6.23}$$

just because they rotate over the same axis with a difference of 2π in the angles. *The two associated $SU(2)$ elements, $X(\vec{\alpha}\,')$ and $X(\vec{\alpha})$, however, are opposite to each other:*

$$X(\vec{\alpha}\,') = -X(\vec{\alpha})\ . \tag{6.24}$$

[13] Similarly, the complex numbers with norm 1 form the group $U(1)$, which simply consists of all phase factors $\exp i\alpha$.

This follows from Eqs. (6.14), (6.18) and the fact that $\cos\frac{\alpha'}{2} = -\cos\frac{\alpha}{2}$ en $\sin\frac{\alpha'}{2} = \sin\frac{\alpha}{2}$.

The above implies that, strictly speaking, the elements of $SU(2)$ are not a representation of the three-dimensional rotation group, but a *projective* representation. After all, in the product of the rotations

$$R(\vec{\alpha})\, R(\vec{\beta}) = R(\vec{\gamma}), \tag{6.25}$$

with α , β , we would also have $\gamma \le \pi$, but the product of the associated $SU(2)$ matrices,

$$X(\vec{\alpha})\, X(\vec{\beta}) = \pm X(\vec{\gamma}) \ , \tag{6.26}$$

the value of γ depends on α and β but its length can be either larger or smaller than π , so we may or may not have to include a minus sign in the equation[14] if we wish to restrict ourselves to vectors shorter than π . The group $SU(2)$ does have the same structure constants, and thus the same group product structure, as the rotation group, but the latter only holds true in a small domain surrounding the unit element, and not exactly for the entire group.

A spinor φ^α transforms as follows:

$$\varphi^\alpha \to \varphi^{\alpha\prime} = X^\alpha_{\ \beta}\, \varphi^\beta \ . \tag{6.27}$$

The complex conjugated vectors then transform as

$$\varphi^*_\alpha \to \varphi^{*\prime}_\alpha = (X^\alpha_{\ \beta})^*\, \varphi^*_\beta = (X^\dagger)^\beta_{\ \alpha}\, \varphi^*_\beta \ . \tag{6.28}$$

Here, we introduced an important new notation: the indices are sometimes in a raised position (superscripts), and sometimes lowered (subscripts). This is done to indicate that spinors with superscripts, such as in (6.27), transform differently under a rotation than spinors with subscripts, such as (6.28). Upon complex conjugation, a superscript index becomes a subscript, and *vice versa*. Subsequently, we limit our summation convention to be applied only in those cases where one superscript index is identified with one subscript index:

$$\phi_\alpha \psi^\alpha \equiv \sum_{\alpha=1}^{2} \phi_\alpha \psi^\alpha \ . \tag{6.29}$$

In contrast to the case of the rotation group, one cannot apply group-invariant summations with two superscript or two subscript indices, since

$$X^\alpha_{\ \alpha'}\, X^\beta_{\ \beta'}\, \delta^{\alpha'\beta'} = \sum_\gamma X^\alpha_\gamma X^\beta_\gamma \ne \delta^{\alpha\beta} \ , \tag{6.30}$$

because X in general is not orthogonal, but unitary. The only allowed Kronecker delta function is one with one superscript and one subscript index: δ^α_β . A summation such as in Eq.(6.29) *is* covariant:

$$\sum_{\alpha=1}^{2} \phi'_\alpha \psi'^\alpha = (X^\alpha_{\ \beta})^* X^\alpha_{\ \gamma} \phi_\beta \psi^\gamma = (X^\dagger\, X)^\beta_\gamma \phi_\beta \psi^\gamma = \delta^\beta_\gamma \phi_\beta \psi^\gamma = \sum_{\beta=1}^{2} \phi_\beta \psi^\beta \ , \tag{6.31}$$

[14]On the other hand, we may state that the three-dimensional rotations are a representation of the group $SU(2)$.

where unitarity, according to the first of Eqs. (6.21), is used.

We do have two other *invariant tensors* however, to wit: $\varepsilon_{\alpha\beta}$ and $\varepsilon^{\alpha\beta}$, which, as usual, are defined by

$$\varepsilon^{\alpha\beta} = e_{\alpha\beta} = -\varepsilon_{\beta\alpha} \ , \qquad \varepsilon^{12} = \varepsilon_{12} = 1 \ . \tag{6.32}$$

By observing that

$$X^{\alpha}_{\ \alpha'} \, X^{\beta}_{\ \beta'} \, \epsilon^{\alpha'\beta'} = \det X \, \epsilon^{\alpha\beta} = \epsilon^{\alpha\beta} \ , \tag{6.33}$$

where the second of Eqs. (6.21) was used, we note that $\varepsilon_{\alpha\beta}$ and $\varepsilon^{\alpha\beta}$ after the transformation take the same form as before.

From this, one derives that the representation generated by the matrices X^* is *equivalent* to the original representation. With every *co*-spinor φ^α we have a *contra*-spinor,

$$\psi_\alpha \stackrel{\text{def}}{=} \varepsilon_{\alpha\beta}\varphi^\beta \ , \tag{6.34}$$

transforming as in Eq. (6.28).

The fact that X and X^* are equivalent can also be demonstrated by writing $\varepsilon_{\alpha\beta}$ as a matrix:

$$\begin{aligned} \varepsilon \, X \varepsilon^{-1} &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} a^* & b^* \\ -b & a \end{pmatrix} \\ &= X^* \ , \end{aligned} \tag{6.35}$$

since $\varepsilon^2 = -\mathbf{1}$. From this, it follows that the two representations given by (6.27) and (6.28) are equivalent according to the definition given in Eq. (4.3).

Now, let us attempt to find *all* representations of the group $SU(2)$, rather than $SO(3)$. To this end, we let the $SU(2)$ matrices act in an abstract vector space with *complex* coordinates [15] φ^α, where $\alpha = 1, \ 2$. We consider all *analytic* functions f of these two coordinates. Perform the Taylor series expansion of such a function at the origin. At the N^{th} order, the Taylor expansion terms form homogeneous, symmetric polynomials in φ^α of degree N. Obviously, N is a non negative integer. Since f is analytic, the complex conjugated spinors, φ^*_α are not allowed to enter in these polynomials. Write

$$Y^{\alpha_1\alpha_2\cdots\alpha_N} = \varphi^{\alpha_1}\,\varphi^{\alpha_2}\cdots\varphi^{\alpha_N}. \tag{6.36}$$

Under $SU(2)$, these polynomials transform as follows:

$$Y^{\alpha_1\alpha_2\cdots\alpha_N} \rightarrow Y^{\alpha_1\alpha_2\cdots\alpha_N\,\prime} = X^{\alpha_1}_{\ \alpha_1'}\,X^{\alpha_2}_{\ \alpha_2'}\cdots X^{\alpha_N}_{\ \alpha_N'}\;Y^{\alpha_1'\alpha_2'\cdots\alpha_N'} \ . \tag{6.37}$$

In view of the above, we expect that the tensors $Y^{\alpha_1\alpha_2\cdots\alpha_N}$ (which, because of the symmetry under interchange of the indices, do not depend on the way these indices are

[15]The coordinates φ^α are therefore slightly more difficult to interpret.

ordered), should transform as representations of $SU(2)$. Indeed, they are *irreducble* representations. The independent coefficients of these polynomials are completely characterized by specifying the number p_1 of indices that are equal to 1 (the remaining indices, their number being $p_2 = N - p_1$, must be equal to 2), and so we find the number of independent coefficients in a polynomial of degree N to be

$$\sum_{p_1=0}^{N} = N + 1 \ . \tag{6.38}$$

Thus, here we have representations of dimension $N+1$, for any non negative integer N.

Subsequently, we can write the $SU(2)$ generators, acting on functions of the coordinates φ, as differential operators. This leads to

$$L_i^{SU(2)} = -\frac{1}{2}(\tau_i)^\alpha_{\ \beta}\,\varphi^\beta\,\frac{\partial}{\partial\varphi^\alpha} \ , \tag{6.39}$$

so that infinitesimal $SU(2)$ transformations on functions $f(\varphi)$ can be written as

$$\begin{aligned} f(\varphi) \to f'(\varphi) &= \Big(\mathbf{1} - i\vec\alpha\cdot\vec L^{\,SU(2)} + \mathrm{O}(\alpha^2)\Big)\, f(\varphi) \\ &= f(\varphi) + \frac{i}{2}\alpha_j\,(\tau_j)^\alpha_{\ \beta}\,\varphi^\beta\,\frac{\partial f(\varphi)}{\partial\varphi^\alpha} + \mathrm{O}(\alpha^2) \ . \end{aligned} \tag{6.40}$$

Note in passing that the index α in $\frac{\partial}{\partial\phi^\alpha}$ is treated as a *subscript* index.

Making use of Eq. (6.39), we can now derive the Casimir operator $(\vec L^{SU(2)})^2$ as a differential operator,

$$\begin{aligned} (L_i^{SU(2)})^2 &= \frac{1}{4}\sum_i\left(\tau^{\alpha}_{i\,\beta}\,\varphi^\beta\,\frac{\partial}{\partial\varphi^\alpha}\right)\left(\tau^{\gamma}_{i\,\delta}\,\varphi^\delta\,\frac{\partial}{\partial\varphi^\gamma}\right) \\ &= \frac{1}{4}(\,-\,\delta^\alpha_\beta\,\delta^\gamma_\delta + 2\,\delta^\alpha_\delta\,\delta^\gamma_\beta)\,\varphi^\beta\,\frac{\partial}{\partial\varphi^\alpha}\,\varphi^\delta\,\frac{\partial}{\partial\varphi^\gamma} \\ &= -\frac{1}{4}\,\varphi^\alpha\,\frac{\partial}{\partial\varphi^\alpha}\,\varphi^\gamma\,\frac{\partial}{\partial\varphi^\gamma} + \frac{1}{2}\,\varphi^\beta\,\frac{\partial}{\partial\varphi^\alpha}\,\varphi^\alpha\,\frac{\partial}{\partial\varphi^\beta} \\ &= \frac{1}{4}\left(\varphi^\alpha\frac{\partial}{\partial\varphi^\alpha}\right)^2 + \frac{1}{2}\left(\varphi^\alpha\frac{\partial}{\partial\varphi^\alpha}\right) \ . \end{aligned} \tag{6.41}$$

It is easy to see that the last two lines of Eq. (6.41) are equal by writing all derivatives to the right of the coordinates. The transition from the first to the second line is less trivial. There, use was made of the identity

$$\sum_i (\tau_i)^\alpha_{\ \beta}\,(\tau_i)^\gamma_{\ \delta} = -\delta^\alpha_\beta\,\delta^\gamma_\delta + 2\,\delta^\alpha_\delta\,\delta^\gamma_\beta \ . \tag{6.42}$$

A convenient way to derive this equation is by first multiplying it with an arbitrary matrix X^γ_δ, after which one uses the decomposition rule (6.15) and Eq. (6.17) for this X. If now the derivative of this equation is taken with respect to $X^{\gamma}_{\ \delta}$, we directly end up with the identity (6.42). Evidently, the validity of (6.42) can also be verified by choosing specific values for the indices α , β , γ and δ .

Now, let the operator (6.41) act on the polynomials $Y^{\alpha_1\alpha_2\cdots\alpha_{2s}}$. Using the fact that

$$\left(\varphi^\alpha \frac{\partial}{\partial\varphi^\alpha}\right) Y^{\alpha_1\alpha_2\cdots\alpha_N} = N\, Y^{\alpha_1\alpha_2\cdots\alpha_{2s}}, \tag{6.43}$$

we find directly the result:

$$(L_i^{SU(2)})^2\, Y^{\alpha_1\alpha_2\cdots\alpha_N} = \tfrac{1}{2}N(\tfrac{1}{2}N+1)\, Y^{\alpha_1\alpha_2\cdots\alpha_N} \ . \tag{6.44}$$

Thus, we recognize the representations ℓ of chapter 5, if we write $\ell = s,\ s = \frac{1}{2}N$. We succeeded to (re)construct $(2s+1)$ dimensional representations of $SU(2)$, where s is an integer or an integer plus $\frac{1}{2}$. In these representations, the eigenvalue of the Casimir operator, according to Eq. (6.44), equals $s(s+1)$. In Chapter 5, it was shown that this completes the set of all irreducible representations of $SU(2)$.

We expect that, for integral values of s , the representations coincide with the representation of the rotation group found in chapter 4. This indeed turns out to be the case. To see this, consider the tensors Y with an *even* number of indices. Then, arrange the factors φ^α in pairs, and use in each pair the ε tensor to lower one of the superscript indices to obtain a subscript index:

$$Y^{\alpha\beta} = \varphi^\alpha\,\varphi^\beta\ ; \qquad \hat{Y}^\alpha_\beta = \varepsilon_{\beta\gamma}Y^{\alpha\gamma} = \varepsilon_{\beta\gamma}\varphi^\alpha\varphi^\gamma\ . \tag{6.45}$$

We read off easily that $\mathrm{Sp}(\hat{Y}) = 0$, so that, according to the decomposition (6.15), $\hat{Y}$ can be written as

$$\hat{Y} = \tfrac{1}{2}\sum_i x_i\,\tau_i\ ; \qquad x_i = \hat{Y}^\alpha_\beta\,(\tau_i)^\beta_{\ \alpha}\ . \tag{6.46}$$

Under $SU(2)$, the quantities x_i transform as

$$x_i\ \to\ x'_i\ =\ X^\alpha_{\ \alpha'}\,(X^{-1})^{\beta'}_{\ \beta}\,Y^{\alpha'}_{\beta'}\,(\tau_i)^\beta_{\ \alpha}\ , \tag{6.47}$$

where use was made of the transformation rules for superscript and subscript indices. And now we prove that

$$X^{-1}(\vec{\alpha})\,\tau_i\,X(\vec{\alpha}) = R(\vec{\alpha})_{ij}\,\tau_j\ , \tag{6.48}$$

so that the tensors x_i actually transform exactly like the coordinates x_i in chapter 4. We verify the validity of the transformation rule (6.48) for infinitesimal transformations. One then has

$$\begin{aligned} X^{-1}(\vec{\alpha})\,\tau_i\,X(\vec{\alpha}) &\approx \left(\mathbf{1} - \frac{i}{2}\alpha_j\,\tau_j + \mathrm{O}(\alpha^2)\right)\,\tau_i\,\left(\mathbf{1} + \frac{i}{2}\alpha_k\,\tau_k + \mathrm{O}(\alpha^2)\right) \\ &= \left(\tau_i + \frac{i}{2}\alpha_j\,[\tau_i,\tau_j] + \mathrm{O}(\alpha^2)\right) \\ &= \tau_i + \epsilon_{ijk}\,\tau_j\,\alpha_k + \mathrm{O}(\alpha^2)\ , \end{aligned} \tag{6.49}$$

which indeed takes the same form as infinitesimal rotations of the coordinates x_i .

The rotation operator (6.39) is an exact analogue of the generator of rotations in x space:

$$L_i = -i\varepsilon_{ijk}x_j\frac{\partial}{\partial x_k} \ , \tag{6.50}$$

which we obtain if we apply an infinitesimal rotation (3.11) to a function $\psi(\vec{r})$:

$$\psi(\vec{r}) \to \psi(\vec{r} + \vec{r} \times \vec{\alpha}) = (\mathbf{1} + i\alpha_k L_k)\psi(\vec{r}) \ . \tag{6.51}$$