NPTEL COURSE
SELECTED TOPICS IN MATHEMATICAL PHYSICS

V. Balakrishnan
Department of Physics, Indian Institute of Technology Madras
Chennai 600 036, India

Note: All figures have been omitted in this write-up. The reader will find it an instructive exercise to fill in all necessary figures.

Table of contents

1. **Analytic functions of a complex variable**: Complex numbers. Equations to curves in the plane in terms of z and z^*. The Riemann sphere and stereographic projection. Analytic functions of z and the Cauchy-Riemann conditions. The real and imaginary parts of an analytic function. The derivative of an analytic function. Power series as analytic functions. Convergence of power series.

2. **Calculus of residues**: Cauchy's integral theorem. Singularities— removable singularity, simple pole, multiple pole, essential singularity. Laurent series. Singularity at infinity. Accumulation point of poles. Meromorphic function. Cauchy's integral formula. Contour integration. Residue at infinity. Summation of series using contour integration. Solving difference equations using contour integration. Mittag-Leffler expansions of meromorphic functions.

3. **Linear response; dispersion relations**: Causal, linear, retarded response. Dynamic susceptibility. Symmetry properties of the dynamic susceptibility. Dispersion relations. Admittance of an LCR circuit. Subtracted dispersion relations. Hilbert transform pairs. Discrete and continuous relaxation spectra.

4. **Analytic continuation and the gamma function**: The gamma function. Connection with the gaussian integral. Analytic properties of the gamma function. Mittag-Leffler expansion of $\Gamma(z)$. The logarithmic derivative of the gamma function. Infinite product representation for the gamma function. The beta function. Reflection formula for $\Gamma(z)$. Legendre's duplication formula for $\Gamma(z)$.

5. **Möbius transformations**: Conformal mapping. Definition of a Möbius transformation. Fixed points of a Möbius transformation. The cross-ratio. Normal form of a Möbius transformation. Iterates of a Möbius transformation. Classification of Möbius transformations. The isometric circle. Group properies; the Möbius group. The Möbius group over the reals. The modular group. The invariance group of the unit circle. Connection with the pseudo-unitary group $SU(1,1)$. The group of cross-ratios.

6. **Multivalued functions; integral representations**: Branch points and branch cuts. Algebraic and logarithmic branch points, winding point. Contour integrals in the presence of branch points. An integral involving a class of rational functions. Contour integral representations for the gamma, beta, and Riemann zeta functions. Connection with Bernoulli numbers. The statement of the Riemann hypothesis. Contour integral representations of the Legendre functions $P_\nu(z)$ and $Q_\nu(z)$. Singularities of functions defined by integrals. End-point and pinch singularities, examples. Singularities of the Legendre functions. Dispersion relations for the Legendre functions.

7. **Laplace transforms**: Definition of the Laplace transform. The convolution theorem. Laplace transforms of derivatives. The inverse transform, Mellin's formula. Laplace transform of the Bessel and modified Bessel functions of the first kind. The LCR series circuit. Laplace transforms and random processes: the Poisson process, biased random walk on a linear lattice and on a d-dimensional lattice.

8. **Fourier transforms**: Fourier integrals. Parseval's formula for Fourier transforms. Fourier transform of the δ-function. Relative 'spreads' of a Fourier transform pair. The convolution theorem. Generalization of Parseval's formula. Iterates of the Fourier transform operator $\mathcal{F}$. Eigenvalues and eigenfunctions of $\mathcal{F}$. Unitarity of the Fourier transformation. The Fourier transform in d dimensions. The Poisson summation formula. Some illustrative examples. Generalization to higher dimensions.

9. **Quiz 1; solutions**.

10. **The fundamental Green function for ∇^2**: Green functions. Poisson's equation; the fundamental Green function for ∇^2. Solution for a spherically symmetric source. The Coulomb potential in d dimensions. Ultraspherical coordinates. A divergence problem. Dimensional regularization. A direct derivation using Gauss' Theorem. The Coulomb potential in $d = 2$ dimensions. A direct derivation once again.

11. **The diffusion equation**: Fick's laws of diffusion. The fundamental solution in d dimensions. Solution for an arbitrary initial distribution. Moments of the distance travelled in time t. Diffusion in one dimension: continuum limit of a random walk. The Smoluchowski equation. Sedimentation. Absorbing and reflecting boundary conditions. Free diffusion on a semi-infinite line. Finite boundaries: solution by the method of images. Solution by separation of variables. Survival probability. First-passage-time distribution and mean first-passage time. Connection with the Schrödinger equation for a free particle. Spreading of a quantum mechanical wave packet.

12. **The Green function for (∇^2+k^2); nonrelativistic scattering**: The Helmholtz operator. The scattering amplitude; differential and total cross-sections. Integral equation for scattering. Green function for the Helmholtz operator. Exact formula for the scattering amplitude. Scattering geometry and the momentum transfer. Born series and the Born approximation. The Yukawa and Coulomb potentials. The Rutherford scattering formula.

13. **The wave equation**: Formal solution for the causal Green function. The solution in $(1+1)$, $(2+1)$ and $(3+1)$ dimensions. Retarded solution of the wave equation. Remarks on propagation in spatial dimensions $d > 3$.

14. **The rotation group and all that**: Rotations of the coordinate axes. Orthogonality of rotation matrices. Proper and improper rotations. Generators of infinitesimal rotations in 3 dimensions. The general rotation matrix in 3 dimensions. The finite rotation formula for a vector. Relation between the groups $SO(3)$ and $SU(2)$. Rotation generators in 3 dimensions transform like a vector. Derivation of Hadamard's lemma. The general form of the elements of $U(2)$ and $SU(2)$. The parameter spaces of $SU(2)$ and $SO(3)$. Tensor and spinor representations. The universal covering group of a Lie group. The group $SO(2)$ and its covering group. The groups $SO(n)$ and Spin (n). Parameter spaces of $U(n)$ and $SU(n)$. A bit about the first homotopy group of a space.

15. **Quiz 2; solutions**.

1 Analytic functions of a complex variable

We begin this course with the study of analytic functions of a complex variable. The theory of functions of a complex variable (and its generalization, the theory of functions of *several* complex variables) is one of the richest and most beautiful branches of mathematics, with deep results and far-reaching implications and applications. It is also a vast subject. Only its most rudimentary aspects will be dealt with in what follows. No formal proofs of theorems will be given, as usual, and the treatment of standard material will be heuristic rather than rigorous. I shall assume that you already have some familiarity with analytic functions of a complex variable, from an earlier course on mathematical methods in physics.

Complex numbers: Let's recapitulate some elementary properties of complex numbers. Given a complex number $z = x + iy = re^{i\theta}$, its real and imaginary parts are x and y, respectively; its modulus and argument are r and θ, respectively. Its complex conjugate is $z^* = x - iy = re^{-i\theta}$. It is important to remember that the specification of a complex number implies the specification of two *independent* pieces of information, namely, x and y, or r and θ. be z and z^* themselves. Recall also the basic relations

$$x = r\,\cos\,\theta = \frac{(z+z^*)}{2}\,, \quad y = r\,\sin\,\theta = \frac{(z-z^*)}{2i}\,,$$

as well as

$$r = (x^2+y^2)^{1/2} = zz^* = |z|^2, \quad \theta = \tan^{-1}\left(\frac{y}{x}\right) = \frac{1}{2i}\,\ln\,\left(\frac{z}{z^*}\right).$$

1. Show that the real and imaginary parts of the complex numbers listed below are as indicated. Remember the standard phase convention, according to which $i = e^{i\pi/2}$ and $-i = e^{-i\pi/2} = e^{3\pi i/2}$.

(a) $(i)^{i^i} = \cos\left(\frac{1}{2}\pi e^{-\pi/2}\right) + i\,\sin\left(\frac{1}{2}\pi e^{-\pi/2}\right)$.

(b) $\displaystyle\sum_{n=0}^{\infty} \frac{(i\pi)^{2n+1}}{(2n+1)!} = 0.$

(c) $\displaystyle\sum_{n=0}^{\infty} \frac{1}{(2n)!}\,\left(\frac{1+i}{\sqrt{2}}\right)^n = \cosh\left(\cos\frac{\pi}{8}\right)\,\cos\left(\sin\frac{\pi}{8}\right) + i\,\sinh\left(\cos\frac{\pi}{8}\right)\,\sinh\left(\sin\frac{\pi}{8}\right).$

(d) The iterated square root

$$\begin{aligned}\sqrt{i+\sqrt{i+\sqrt{i+\sqrt{\ldots}\ ad\ inf.}}} &= \frac{1}{2} + \frac{\sqrt{\sqrt{17}+1}}{2\sqrt{2}} + i\,\frac{\sqrt{\sqrt{17}-1}}{2\sqrt{2}} \\ &\simeq 1.3002 + 0.6248\,i.\end{aligned}$$

(The other root is $\simeq -0.3002 - 0.6248\,i$.)

(e) The infinite continued fraction

$$\frac{i}{1+}\,\frac{i}{1+}\,\frac{i}{1+}\cdots \; ad\ inf. \quad = \quad -\frac{1}{2}+\frac{\sqrt{\sqrt{17}+1}}{2\sqrt{2}}+i\,\frac{\sqrt{\sqrt{17}-1}}{2\sqrt{2}}$$
$$\simeq \; 0.3002+0.6248\,i.$$

(The other root is $\simeq -1.3002-0.6248\,i$.)

(f) The infinite continued fraction

$$\frac{1}{i+}\,\frac{1}{i+}\,\frac{1}{i+}\cdots \; ad\ inf. \; = e^{-i\pi/6}=\frac{\sqrt{3}-i}{2}\,.$$

(The other root is $e^{-5\pi i/6}=-\frac{1}{2}(\sqrt{3}+i)$.)

Hint: In (d), (e) and (f), calling the left-hand side z, you have to solve the respective equations $z=(i+z)^{1/2}$, $z=i/(1+z)$ and $z=1/(i+z)$. Each of these reduces to a quadratic equation, and you must take care to identify the correct root in each case.

Equations to curves in the plane in terms of z and z^*: We know that the Cartesian coordinates x and y labeling points on a plane are linearly independent of each other. It is important to recognize that, equivalently, the complex variable $z=x+iy$ and its complex conjugate $z^*=x-iy$ are linearly independent. Once you bear this in mind, it becomes much easier to understand the concept of analytic functions of a complex variable z.

2. Equations to curves in the xy-plane are often very conveniently expressed in terms of z and z^*. Here are some examples of familiar curves, thus expressed:

(a) In terms of the complex variable z, write down the equation of the ellipse in the xy-plane whose foci are at the points $\pm x_0$ on the x-axis, and whose semi-major axis is equal to a (where $a>x_0$).

(b) Find the locus of points in the z-plane given by the following conditions:

(i) $\big|\,|z-1|-|z+1|\,\big|=1$ (ii) $\big|(z-1)/(z-2)\big|=1$ (iii) $\big|e^{z^2}\big|=e^4$.

The Riemann sphere and stereographic projection: We know that the number line extends from $-\infty$ on the left to $+\infty$ on the right. In the complex plane, z may tend to infinity along any of an infinite number of directions. It is convenient to 'compactify' the plane by bringing together all these points at infinity and 'glueing' them together into a single point. This can be done by a mapping between the complex plane and the surface of a sphere. Consider a sphere of unit radius with its center at the origin

of coordinates. The coordinates of any point on the surface are given by $(\xi_1\,,\,\xi_2\,,\,\xi_3)$, where

$$\xi_1^2+\xi_2^2+\xi_3^2=1.$$

In terms of spherical polar coordinates on the unit sphere, we have of course

$$\xi_1=\sin\,\theta\,\cos\,\varphi\,,\;\;\xi_2=\sin\,\theta\,\sin\,\varphi\,,\;\;\xi_3=\cos\,\theta\,.$$

The coordinates of the 'north pole' N are $(0,0,1)$, while those of the 'south pole' S are $(0,0,-1)$.

Now consider the equatorial plane of the sphere. Let the Cartesian coordinates on this plane be $(x\,,\,y)$, with the same origin as the sphere, and the x and y axes running along the ξ_1 and ξ_2 axes, respectively. The **stereographic projection** of any point P on the sphere onto the equatorial plane is obtained by joining N and P by a straight line. The projection of P is the point P$'$ where the line cuts the equatorial plane. This plane is regarded as the complex plane, while the sphere is called the **Riemann sphere**. The coordinates x and y of P$'$ in the complex plane are related to the coordinates $(\xi_1\,,\,\xi_2\,,\,\xi_3)$ on the Riemann sphere by

$$x=\frac{\xi_1}{1-\xi_3}=\cot\,\tfrac{1}{2}\theta\,\cos\,\varphi\quad\text{and}\quad y=\frac{\xi_2}{1-\xi_3}=\cot\,\tfrac{1}{2}\theta\,\sin\,\varphi.$$

Therefore

$$z=x+iy=\frac{\xi_1+i\xi_2}{1-\xi_3}=\cot\,\tfrac{1}{2}\theta\,e^{i\varphi}.$$

It is evident that points of the northern hemisphere are mapped to points outside the unit circle in the complex plane; points in the southern hemisphere are mapped to points inside the unit circle; and points on the equator are mapped to themselves as points on the unit circle $|z|=1$.

As P gets closer to the point of projection N, it is clear that the point P$'$ moves farther and farther away from the origin in the complex plane. The **point at infinity** in the complex plane is defined as the image of the north pole under the projection, and denoted by ∞. The finite part of the complex plane, together with the point at infinity, is called the **extended complex plane**. In what follows, by the term 'complex plane' we shall generally mean the extended complex plane. It is very advantageous to have 'infinity' identified with a single point in this fashion, and to be able to specify ∞ as a specific value for the complex number z.

2. Some of the properties of this stereographic projection are the following.

(a) The mapping from the Riemann sphere to the complex plane, is *invertible*. Show that the inverse relations expressing $(\xi_1\,,\,\xi_2\,,\,\xi_3)$ in terms of z and z^* are

$$\xi_1=\frac{z+z^*}{|z|^2+1}\,,\quad\xi_2=\frac{z-z^*}{i(|z|^2+1)}\,,\quad\xi_3=\frac{|z|^2-1}{|z|^2+1}\,.$$

(b) What is the projection in the complex plane of a circle of latitude on the Riemann sphere? Of a meridian of longitude?

- In general, circles on the Riemann sphere are mapped into circles or straight lines in the complex plane by stereographic projection.

(c) Let P and Q be any two points on the Riemann sphere, whose projections on the complex plane are given by z_1 and z_2. Show that the *chordal distance* between P and Q (i.e., the length of a chord drawn from P to Q) is given by

$$d(z_1,\, z_2) = \frac{2\,|z_1 - z_2|}{\sqrt{(|z_1|^2+1)(|z_2|^2+1)}}\,.$$

It is obvious that $d(z_1,\, z_2) = d(z_2,\, z_1)$ and that $d(z_1,\, z_2) = 0$ iff $z_1 = z_2$. Further, for any three points z_1, z_2 and z_3, we have

$$d(z_1,\, z_3) \le d(z_1,\, z_2) + d(z_2,\, z_3).$$

With this definition of the distance, the exceptional role played by the point at infinity causes no trouble, as it does in the case of the ordinary Euclidean definition $|z_1 - z_2|$. We have

$$d(z,\, \infty) = \frac{2}{\sqrt{|z|^2+1}}\,.$$

Clearly, $d(z_1,\, z_2) \le 2$ for any two points z_1 and z_2 in the extended complex plane.

Analytic functions of z; the Cauchy-Riemann conditions: The function $f(z)$ is said to be analytic (more precisely, **holomorphic**) in some region of the complex z-plane if its real and imaginary parts satisfy the Cauchy-Riemann conditions at every point in this region. Let $u(x\,,\, y)$ and $v(x\,,\, y)$ be the real and imaginary parts of such a function. The Cauchy-Riemann conditions are

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}\,.$$

The region in which these conditions are satisfied is the region of analyticity, or **domain of holomorphy**, of the function concerned. If a function $f(z)$ is holomorphic in the whole of the finite part of the complex plane (that is, for all $|z| < \infty$), then it is called an **entire function**. There exists a theorem (**Liouville's Theorem**) that states that

- the only function analytic at every point in the *extended* complex plane (that is, at all points in the finite part of the complex plane *as well as* the point at infinity) is a constant.

- Hence an entire function that is not just a constant cannot be analytic at the point at infinity.[1]

One way to understand what is meant by an analytic function of z is as follows. Recall that, just as x and y are linearly independent coordinates on the plane, so are the linear combinations $z = x + iy$ and $z^* = x - iy$. Hence any function of x and y can equally well be written as a function of z and z^*. An analytic function of z is then a sufficiently well-behaved function[2] of x and y that depends on the combination $x + iy$ alone, and that does not involve the other combination, namely, $x - iy$. Thus, f is an analytic function of z if $\partial f/\partial z^* \equiv 0$. But, by the chain rule of differentiation,

$$\frac{\partial f}{\partial z^*} = 0 \quad \Longrightarrow \quad \frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y} = 0.$$

The real and imaginary parts of the last equation yield precisely the Cauchy-Riemann conditions. These conditions may therefore be written compactly as $\partial f/\partial z^* \equiv 0$. Writing any function of x and y in terms of z and z^* enables us to identify functions that *cannot* be analytic functions of z, merely by checking whether z^* makes its appearance in the function. If it does so, the function cannot be an analytic function of z.

4. Identify which among the following are analytic functions of z in some region, and which are not. In the former case, identify the region of analyticity.

(a) x	(b) $ix - y$	(c) r
(d) $e^{i\theta}$	(e) $x^2 - iy^2$	(f) $x - iy$
(g) $(x + iy)^2$	(h) $x^2 - y^2 - 2ixy$	(i) $i\,\tan^{-1}(y/x)$
(j) $(iy + x^2 + y^2)$	(k) $[(x + i)^2 - y^2]^{1/2}$	(l) $(x - iy)/(x^2 + y^2)$
(m) $x^4 + 2ix^2y^2 - y^4$	(n) $i\,e^x \sin y$	(o) $x^2 + x + 1 - y^2 + iy\,(2x + 1)$.

The real and imaginary parts of an analytic function: Let u and v be the real and imaginary parts of an analytic function $f(z)$. It follows from the Cauchy-Riemann conditions that both u and v individually satisfy Laplace's equation in two dimensions, i.e., $\nabla^2 u = 0$ and $\nabla^2 v = 0$.

- The real and imaginary parts of an analytic function are thus *harmonic* functions. The combination $u + iv$ constitutes an analytic function in the intersection of the regions in which they are individually harmonic.

- It is also evident from the Cauchy-Riemann conditions that an analytic function cannot be identically equal to a purely real or purely imaginary function, except in the trivial case when the function is just a constant.

[1] In other words, it must be *singular* at $z = \infty$. We will discuss singularities subsequently.

[2] By the phrase 'sufficiently well-behaved', we mean that the real and imaginary parts of the function have continuous partial derivatives with respect to x and y.

- An analytic function that is real when its argument is real, i.e., a function such that $f(x) = u(x, 0)$ is real (or $v(x, 0)$ vanishes identically), is called a **real analytic** function.

- It follows from the Cauchy-Riemann conditions that the scalar product $(\nabla u \cdot \nabla v) = 0$. The curves $u =$ constant and $v =$ constant thus constitute two mutually orthogonal families of curves in the complex plane, for *any* analytic function.

5. Given u (or v) as indicated below, find the corresponding v (respectively, u), either by inspection or by integrating the Cauchy-Riemann conditions. Indicate also the region of analyticity of $f(z)$ in each case.

(a) $u = x^2 - y^2$	(b) $u = e^x \cos y$	(c) $u = \ln (x^2 + y^2)^{1/2}$
(d) $u = \cos x \ \cosh y$	(e) $v = e^{2x} \sin (2y)$	(f) $v = 2xy$
(g) $v = \sin x \ \sinh y$	(h) $v = \tan^{-1}(y/x)$	(i) $v = e^{x^2-y^2} \sin (2xy)$.

6. Cauchy-Riemann conditions in polar form: Let $z = r\, e^{i\theta}$ and $f(z) = R\, e^{i\psi}$. Thus $R = |f(z)|$, while $\psi = \arg f(z)$. Show that the Cauchy-Riemann conditions now read

$$\frac{\partial \ln R}{\partial \ln r} = \frac{\partial \psi}{\partial \theta}\,, \quad \frac{\partial \ln R}{\partial \theta} = -\frac{\partial \psi}{\partial \ln r}\,.$$

The derivative of an analytic function: The condition of analyticity is so strong that an analytic function is guaranteed to have a derivative that is, moreover, itself an analytic function. It then follows at once that it has derivatives of arbitrarily high order that are also analytic functions. Clearly, this is in marked contrast to the case of functions of a real variable, where a function may be once differentiable but not twice differentiable, or, in general, differentiable r times, but not $(r+1)$ times. On the other hand,

- any analytic function of a complex variable is infinitely differentiable.

Defining the derivative of an analytic function helps us understand the Cauchy-Riemann conditions and the meaning of analyticity in yet another (albeit related) way. In analogy with the definition of the derivative of a function of a real variable, we may define the derivative as

$$\frac{df(z)}{dz} = \lim_{\delta z \to 0} \frac{f(z + \delta z) - f(z)}{\delta z},$$

where δz is an infinitesimal quantity of magnitude ϵ. The question that arises is: in what *direction* should the point $z + \delta z$ be taken, relative to the point z? That is, what should the argument (or phase angle) of δz be? Suppose the complex number δz has an argument α, i.e., $\delta z = \epsilon\, e^{i\alpha}$. Then, provided the real and imaginary parts of $f(z)$

have continuous partial derivatives, the definition of the derivative given above yields, in the limit $\epsilon \to 0$,

$$\frac{df(z)}{dz} = e^{-i\alpha} \left[\left(\frac{\partial u}{\partial x} \cos\,\alpha + i \frac{\partial v}{\partial y} \sin\,\alpha \right) + i \left(\frac{\partial v}{\partial x} \cos\,\alpha - i \frac{\partial u}{\partial y} \sin\,\alpha \right) \right] .$$

The remarkable fact is that this expression becomes completely independent of α *if and only if*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} ,$$

which are precisely the Cauchy-Riemann conditions.

- Analyticity may therefore be understood as equivalent to the requirement that the limit defining the derivative of a complex function be *unique*, independent of the direction along which $z + \delta z$ approaches z in the complex plane.

7. In its region of analyticity, and analytic function $f(z)$ may also be regarded as a **map** from (a region of) the complex plane to (a region of) the complex plane.

(a) Show that (the determinant of) the Jacobian of the transformation $(x\,,\,y) \rightarrowtail (u\,,\,v)$ is just $|f\,'(z)|^2$, where $f\,'(z)$ denotes the derivative of $f(z)$.

(b) Show that $\nabla^2(|f(z)|^2) = 4\,|f\,'(z)|^2$.

Power series as analytic functions: Let $f(z)$ be analytic inside a region $\mathcal{R}$. If z_0 is a point in the region, then there generally exists a neighborhood of z_0 such that $f(z)$ can be expanded in an absolutely convergent power series in $(z - z_0)$, called a **Taylor series**:

$$f(z) = \sum_{n=0}^{\infty} a_n\,(z - z_0)^n .$$

Moreover, this neighborhood is the interior of a circle centered at z_0, i.e., it is given by $|z - z_0| < R$, where R is called the **radius of convergence** of the series. The circle $|z - z_0| = R$ is called the **circle of convergence** of the series. It lies within $\mathcal{R}$, the region of analyticity of $f(z)$. (Its boundary may coincide with that of $\mathcal{R}$ at one or more points.) Absolute convergence means that the sum of the *magnitudes* of the terms of the series is also finite. Loosely speaking, such convergence permits us to manipulate the series terms by term, as long as z remains inside the circle of convergence: that is, the series may be re-arranged without affecting the sum of the series, or integrated or differentiated term by term, and so on.

If $f(z)$ and z_0 are specified, the coefficients a_n in the Taylor series above are uniquely determined. We have

$$a_n = \frac{1}{n!} \left[\frac{d^n f(z)}{dz^n} \right]_{z=z_0} ,$$

the existence of derivatives of all orders being guaranteed by the analyticity of $f(z)$. The radius of convergence itself is determined by the coefficients $\{a_n\}$ —more precisely, by the asymptotic (or $n \to \infty$) behavior of a_n . The radius of convergence is given by

$$R = \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right|,$$

provided the limit concerned exists. More generally,

$$R = \limsup |a_n|^{1/n},$$

where lim sup (or supremum) stands for the *least upper bound* of the quantity concerned. It is important to keep in mind that

- a power series is a *representation* of an analytic function inside the circle of convergence of the series. The function may have other representations as well, each with its own region of validity.

8. Convergence of power series: Find the region of absolute convergence of each of the following power series:

(a) $\displaystyle\sum_{n=0}^{\infty} \frac{z^n}{(n+1)^3}$ (b) $\displaystyle\sum_{n=0}^{\infty} \frac{(z-3)^n}{\sqrt{n!}}$ (c) $\displaystyle\sum_{n=1}^{\infty} \frac{(z/2)^n}{n(n+1)}$ (d) $\displaystyle\sum_{n=1}^{\infty} \frac{z^n}{\ln(n+1)}$

(e) $\displaystyle\sum_{n=0}^{\infty} \frac{z^n}{(n+1)^n}$ (f) $\displaystyle\sum_{n=0}^{\infty} \frac{(z-2)^{2n}}{(2n)!}$ (g) $\displaystyle\sum_{n=1}^{\infty} \frac{z^n}{\ln\ln(n+1)}$ (h) $\displaystyle\sum_{n=0}^{\infty} z^{n^2}$

(i) $\displaystyle\sum_{n=0}^{\infty} z^{n!}$ (j) $\displaystyle\sum_{n=0}^{\infty} \frac{z^n}{(n!)^{1/6}}$ (k) $\displaystyle\sum_{n=0}^{\infty} \frac{(z-i\sqrt{2})^n}{n!}$ (l) $\displaystyle\sum_{n=0}^{\infty} n^7\, z^n$

(m) $\displaystyle\sum_{n=0}^{\infty} (z)^{2^n}$ (n) $\displaystyle\sum_{n=1}^{\infty} \frac{(\ln n)\, z^n}{n}$ (o) $\displaystyle\sum_{n=2}^{\infty} \frac{(\ln\ln n)\, z^n}{\ln n}$ (p) $\displaystyle\sum_{n=0}^{\infty} \left(\frac{1-z}{1+z}\right)^n$

(q) $\displaystyle\sum_{n=0}^{\infty} \frac{(2z)^n}{n^2+1}$ (r) $\displaystyle\sum_{n=0}^{\infty} \left(\frac{z^2-1}{z^2+1}\right)^n$ (s) $\displaystyle\sum_{n=0}^{\infty} \frac{(z/4)^n}{n+2}$ (t) $\displaystyle\sum_{n=1}^{\infty} \frac{\Gamma(n-\frac{1}{2})}{\Gamma(n)}\, z^n.$

If R is infinite for a power series centered about any finite point z_0 , this means that the function represented by the power series is an entire function. A power series *diverges* outside its circle of convergence. *On* the circle of convergence itself, its behavior can be quite complicated, and change from one point on the circle to another. The series may diverge at one or more points on the circle; or it may converge at some

points, diverge at others, and oscillate at others; it may even converge at all points. What is guaranteed, however, is that the function that the power series represents inside the circle of convergence has at least one singularity *on* that circle. Remarkably enough, this remains true even if the power series itself converges absolutely at all points on the circle of convergence. An example of this phenomenon is provided by the series $\sum_{n=1}^{\infty} z^n/n^2$. This series converges absolutely at *all* points on its circle of convergence, which is just the unit circle $|z| = 1$. (It attains its largest value on the circle at the point $z = 1$, where the sum is equal to $\zeta(2) = \pi^2/6$.) And yet, the function represented by the series is singular at $z = 1$. At the other extreme, we have cases in which the function represented by the series has a dense set of singularities everywhere on the circle of convergence. Examples of such series are $\sum_{n=0}^{\infty} z^{n!}$ and $\sum_{n=1}^{\infty}(z)^{2^n}$, for each of which the unit circle $|z| = 1$ is the circle of convergence. We shall return to some of these aspects later on.

2 Calculus of residues

Cauchy's integral theorem: As you know, the property of analyticity of a function $f(z)$ implies that the function is infinitely differentiable, and that each derivative is also an analytic function. What does analyticity imply for the *integral* of a function? Once again, the Cauchy-Riemann condition can be used to arrive at the following conclusion: Suppose $f(z)$ is analytic in some region $\mathcal{R}$. Then, if z_1 and z_2 are two points lying in $\mathcal{R}$, the **line integral**

$$\int_{z_1}^{z_2} dz\, f(z)$$

of the function is *independent of the actual path* between the end-points, provided of course that the path lies entirely in $\mathcal{R}$. The value of the line integral is dependent only on the end points.[3] As a consequence, the path connecting the end-point z_1 and z_2 can be distorted like a flexible and stretchable piece of string, and the value of the integral does not change.

Cauchy's integral theorem follows at once: Let C be an oriented *closed* contour lying entirely in $\mathcal{R}$. Then

$$\oint_C dz\, f(z) = 0.$$

It is evident that the contour C may be distorted like a rubber band without changing the property above, as long as it does not leave $\mathcal{R}$. These properties are responsible for much of the power of contour integration in evaluating integrals, and hence in solving a variety of problems that can be reduced to the evaluation of such integrals.

Singularities: We now come to the *singularities* of an analytic function. These are the most interesting features of analytic functions of one or more complex variables. Even in the case of a function of a single complex variable, singularities come in a remarkable number of varieties. But we shall only be concerned with the simplest (and most frequently encountered kinds) of singularities.

Removable singularity: The case of a **removable singularity** may be disposed of first, with the help of an example. The function

$$f(z) = \frac{\sin\, z}{z}$$

[3] This should remind you of the corresponding property of a conservative vector field $\mathbf{u}(\mathbf{r})$, namely, one that can be expressed as the gradient of a scalar field $\nabla\,\phi(\mathbf{r})$. In that case, too, the value of the line integral of the vector field from $\mathbf{r}_1$ to $\mathbf{r}_2$ is independent of the actual path between the two points, and is given by

$$\int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{u}\cdot d\mathbf{r} = \int_{\mathbf{r}_1}^{\mathbf{r}_2} \nabla\,\phi\cdot d\mathbf{r} = \phi(\mathbf{r}_2) - \phi(\mathbf{r}_1).$$

is not defined at $z = 0$, and appears to have a singularity at that point. However, $\lim_{z\to 0} f(z) = 1$, and we may *define* $f(0) = 1$ so as to preserve continuity. This gets rid of the 'removable' singularity and ensures the continuity of $f(z)$ at $z = 0$. In all that follows, it will be assumed that functions are defined at such removable singularities in accord with the limiting values concerned, so that continuity is guaranteed.

Simple pole: The simplest nontrivial singularity is a **simple pole**. $f(z)$ has a simple pole at the point $z = a$ if, in a neighborhood of that point, it can be expressed in the form

$$f(z) = \underbrace{\frac{c_{-1}}{(z-a)}}_{\text{singular part}} + \underbrace{\sum_{n=0}^{\infty} c_n\,(z-a)^n}_{\text{regular part}}\ .$$

The first term on the right is called the 'singular part' of $f(z)$ at $z = a$. It becomes infinite at $z = a$. The coefficient c_{-1} is a non-zero, finite constant (in general, a complex number) that is called the **residue** of $f(z)$ at $z = a$. The notation c_{-1} is used to remind us that it is the coefficient of $(z-a)^{-1}$ in the expansion of $f(z)$ in powers of $(z-a)$. The 'regular part' stands for a function that is analytic in the neighborhood concerned, and hence is expandable in a convergent power series in (non-negative) powers of $(z-a)$, in the neighborhood concerned. In some cases the series may, of course, terminate after a finite number of terms.

The crucial point is that, at a simple pole, the singular part comprises just the single term $c_1\,(z-a)^{-1}$. This term may or may not be *explicit*. Consider, for example, the function $(\sin z)/z^2$. In the neighborhood of the origin, it has the expansion

$$\frac{\sin z}{z^2} = \frac{1}{z^2}\sum_{n=0}^{\infty}\frac{(-1)^n\,z^{2n+1}}{(2n+1)!} = \underbrace{\frac{1}{z}}_{\text{sing. part}} + \underbrace{\Big(-\frac{z^2}{3!}+\frac{z^4}{5!}-\ldots\Big)}_{\text{reg. part}}\ .$$

This function therefore has a simple pole at $z = 0$, with residue equal to 1. Incidentally, it has no other singularities in the finite part of the complex plane (remember that $\sin z$ is an entire function). As a second example, consider the function $\operatorname{cosec} z = 1/(\sin z)$. Since $\sin z$ has simple zeroes at all integer multiples of π, i.e., at $z = n\pi$ where $n \in \mathbb{Z}$, it follows that $\operatorname{cosec} z$ has simple poles at those points. Using the fact that $\sin n\pi = 0$ and $\cos n\pi = (-1)^n$, the Taylor expansion of $\sin z$ near $z = n\pi$ is given by

$$\sin z = (-1)^n\,(z-n\pi) - \tfrac{1}{6}\,(-1)^n\,(z-n\pi)^3 + \ldots\ .$$

Hence

$$\operatorname{cosec} z = \frac{(-1)^n}{(z-n\pi)} + \tfrac{1}{6}\,(-1)^n\,(z-n\pi) + \ldots = \frac{(-1)^n}{(z-n\pi)} + \text{ regular part.}$$

Thus the residue of $\operatorname{cosec} z$ at its simple pole at $z = n\pi$ is $(-1)^n$.

In general, if $f(z)$ has a simple pole at $z = a$, then its residue at the pole is given by

$$c_{-1} = \operatorname{Res} f(z)\Big|_{z=a} = \lim_{z\to a} [(z-a)\, f(z)].$$

If $f(z)$ is of the form $g(z)/h(z)$ where $g(z)$ and $h(z)$ are analytic at $z = a$, and $h(z)$ has a simple zero at $z = a$ while $g(a) \neq 0$, then $f(z)$ has a simple pole at $z = a$. We have

$$f(z) = \frac{g(z)}{h(z)} \quad \Longrightarrow \quad \operatorname{Res} f(z)\Big|_{z=a} = \frac{g(a)}{h\,'(a)}\,,$$

where the prime denotes the derivative of the function concerned.

1. Verify the following: for every $n \in \mathbb{Z}$,

(a) cosech z has a simple pole at $z = in\pi$, with residue $(-1)^n$.

(b) sec z has a simple pole at $z = (n + \frac{1}{2})\pi$, with residue $(-1)^{n+1}$.

(c) sech z has a simple pole at $z = i(n + \frac{1}{2})\pi$, with residue $i\,(-1)^{n+1}$.

(d) tan z has a simple pole at $z = (n + \frac{1}{2})\pi$, with residue -1.

(e) tanh z has a simple pole at $z = i(n + \frac{1}{2})\pi$, with residue $+1$.

(f) cot z has a simple pole at $z = n\pi$, with residue $+1$.

(g) coth z has a simple pole at $z = in\pi$, with residue $+1$.

Note, in particular, how the residue becomes independent of n in the last four cases. This property will be useful when we deal with the summation of certain series.

Multiple pole: The next case is that of a pole of higher order, or a **multiple pole**. The function $f(z)$ has a pole of order m at $z = a$ if, in the neighborhood of that point, it can be expressed in the form

$$f(z) = \underbrace{\frac{c_{-m}}{(z-a)^m} + \cdots + \frac{c_{-1}}{(z-a)}}_{\text{singular part}} + \underbrace{\sum_{n=0}^{\infty} c_n\,(z-a)^n}_{\text{regular part}}\,,$$

where m is a positive integer. The residue of $f(z)$ at the multiple pole is still the coefficient c_{-1}. The significance of the residue will become clear shortly, when we consider Cauchy's integral theorem. For the present, note that the residue at a multiple pole of order m is given by

$$c_{-1} = \operatorname{Res} f(z)\Big|_{z=a} = \lim_{z\to a} \frac{1}{(m-1)!}\,\frac{d^{m-1}}{dz^{m-1}}\,[(z-a)^m\, f(z)].$$

This formula is a generalization of that for the residue at a simple pole.

Essential singularity; Laurent series: Going a step further, the singular part may involve *all* negative integral powers of $(z-a)$. The function $f(z)$ has an (isolated) **essential singularity** at $z=a$ if, in the neighborhood of that point (see below), it can be expressed in the form

$$f(z) = \underbrace{\sum_{n=1}^{\infty} \frac{c_{-n}}{(z-a)^n}}_{\text{singular part}} + \underbrace{\sum_{n=0}^{\infty} c_n\,(z-a)^n}_{\text{regular part}} \ .$$

Once again, the coefficient c_{-1} is the residue of $f(z)$ at the singularity. A representation such as the foregoing in positive as well as negative powers of $(z-a)$ is called a **Laurent series**. The question now arises as to the region in which such a representation is valid. The regular part, as we know, is convergent *inside* some circle of convergence centered at $z=a$, and of radius r_1, say. By setting $w=(z-a)^{-1}$, it follows that the singular part is convergent inside some circle of convergence in the complex w-plane that is centered at the origin. In other words, it is convergent *outside* some circle centered at $z=a$, of radius r_2, say. Then, *provided* $r_2 < r_1$, *there is an annular overlap region* of inner radius r_2 and outer radius r_1, in which both the infinite series are absolutely convergent. This annulus is the region in which the Laurent series is valid. It may so happen, of course, that $r_2 = 0$. Then the Laurent series is valid in the *punctured disc* of radius r_1 centered at $z=a$, with only the point $z=a$ left out. For example, the function

$$e^{1/z} = \underbrace{\sum_{n=1}^{\infty} \frac{1}{n!\,z^n}}_{\text{sing. part}} + \underbrace{1}_{\text{reg. part}}$$

has an essential singularity at the origin $z=0$. In this case $r_1=\infty$ while $r_2=0$. Hence the Laurent series on the right-hand side of the equation above is valid for all $z\neq 0$, including the point at infinity in the z-plane. That is, it is valid in the region $|z|>0$. In general, *a Laurent series is convergent (i.e., provides a valid representation of an analytic function) in some annular region.*

Singularity at infinity: What about possible singularities at the point at infinity, i.e., at $z=\infty$? The nature of such a singularity, if any, can be deduced by first changing variables to $w=1/z$. This maps the point $z=\infty$ to the point $w=0$. The singularity of the corresponding transformed function $f(1/w)\equiv\phi(w)$ at the origin $w=0$ then determines the nature of the singularity of $f(z)$ at $z=\infty$. For example, if $f(z)=z$ itself, then $\phi(w)=1/w$, which has a simple pole at $w=0$. Hence $f(z)=z$ has a simple pole at $z=\infty$. Likewise, it follows that:

— A polynomial of order n in z has a pole of order n at $z=\infty$.

— The function e^z has an essential singularity at $z = \infty$. Hence each of the functions $\cos z$, $\sin z$, $\cosh z$ and $\sinh z$ also has an essential singularity at ∞.

The function

$$e^{1/z} + e^z = \sum_{n=1}^{\infty} \frac{1}{n!\, z^n} + 1 + \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

has essential singularities at $z = 0$ as well as $z = \infty$. The Laurent series on the right-hand side of the equation above is valid in the annular region $0 < |z| < \infty$.

The **residue at infinity** of a function $f(z)$, however, is a little different from the usual idea of a residue at a singularity at any point in the finite part of the complex plane. I'll return to this question shortly.

Accumulation point of poles: We have seen that the function $\operatorname{cosec} z$ has simple poles at all integer multiples of π, i.e., at the set of points $\{z = n\pi,\ n = 0, \pm 1, \pm 2, \ldots\}$. All these points lie on the x-axis. What is the nature of the singularity, if any, of $\operatorname{cosec} z$ at $z = \infty$? The answer is provided by considering the Riemann sphere. Recall that the mapping from the complex plane to the Riemann sphere is given by

$$\xi_1 = \frac{z + z^*}{|z|^2 + 1}\,, \quad \xi_2 = \frac{z - z^*}{i(|z|^2 + 1)}\,, \quad \xi_3 = \frac{|z|^2 - 1}{|z|^2 + 1}\,.$$

The positive and negative segments of the x-axis on the complex plane map onto the lines of longitude 0° and 180°, respectively, on the Riemann sphere. Hence the poles of $\operatorname{cosec} z$ lie on these lines of longitude, at the coordinates

$$\xi_1 = \frac{2n\pi}{n^2\pi^2 + 1}\,, \quad \xi_2 = 0\,, \quad \xi_3 = \frac{n^2\pi^2 - 1}{n^2\pi^2 + 1}\,, \text{ where } n \in \mathbb{Z}.$$

Except for the south pole (corresponding to $n = 0$), all the poles lie in the northern hemisphere (since π is larger than unity). It is evident that they get more and more crowded together as $n \to \pm\infty$, and approach the point $(0\,,\, 0\,,\, 1)$, i.e., the north pole. They become dense on the two longitude lines in an infinitesimal neighborhood of the north pole. On the extended complex plane, $z = \infty$ is therefore an **accumulation point** of poles. The singularity is no longer an *isolated* singularity.

Meromorphic function: An analytic function whose *only* singularities (if any) in the *finite* part of the complex plane are isolated poles is called a **meromorphic function**. Entire functions are trivially also meromorphic (as they have no singularities at all in the finite part of the complex plane). Any rational function of z, i.e., a ratio of two polynomials in z, is a meromorphic function. So are the functions $\operatorname{cosec} z$, $\sec z$, $\operatorname{cosech} z$ and $\operatorname{sech} z$. The function $\operatorname{cosec}(1/z)$ is not meromorphic, because it has an accumulation point of poles at $z = 0$.

2. Identify the regions of the extended complex plane in which the functions listed below are holomorphic. Also find the location, order and residue of each pole or essential singularity of the functions, including the nature of the singularity (if any) at $z = \infty$.

(a) e^{-z^2}	(b) $(1 - z^{100})/(1 - z)$	(c) $(2^z - 1)/z$
(d) $\coth z - z^{-1}$	(e) $(e)^{e^z}$	(f) $(1 - \cos z)/z^2$
(g) $1/(e^z + 1)$	(h) $z \operatorname{cosec} z$	(i) $z \operatorname{cosec}^2 z$
(j) $(\sin z - z \cos z)/z^4$	(k) $1/(2^z - 1)$	(l) $\operatorname{cosec}(z^2)$
(m) $1/(\sin z - \cos z)$	(n) $1/(\cos z - \cosh z)$	(o) $1/(3^z - 2^z)$
(p) $1/(e^{z^2} - 1)$	(q) $1/(e^{2z} - e^z)$	(r) $\exp[1/(z^2 - 1)]$
(s) $\cos(1/z)$	(t) $(e^z - 1)^{-2}$	(u) $\sum_{n=0}^{\infty} e^{-nz}$
(v) $\sum_{n=1}^{\infty} z^n/n^n$	(w) $\sum_{n=0}^{\infty} z^n/(n!)^2$	(x) $\sum_{n=0}^{\infty} [(1 - z)/(1 + z)]^n$

3. Obtain the Laurent series expansions of the following functions, valid in the regions indicated. In each case, the expansion is understood to be in powers of $(z - a)$, where a is the center of the annular region concerned.

(a) $(z + z^{-1})^{1000} \quad (0 < \lvert z \rvert < \infty)$	(b) $z^2\, e^{1/z} \quad (0 < \lvert z \rvert < \infty)$
(c) $z^{-1}\, e^{-1/z^2} \quad (0 < \lvert z \rvert < \infty)$	(d) $(z^2 + z^{-2})^{100} \quad (0 < \lvert z \rvert < \infty)$
(e) $e^{1/z}\,(1 - z)^{-1} \quad (0 < \lvert z \rvert < 1)$	(f) $z^2\, e^{1/(z-2)} \quad (0 < \lvert z - 2 \rvert < \infty)$
(g) $2^{1/(z-1)}\, z^{-2} \quad (0 < \lvert z - 1 \rvert < 1)$	(h) $(z - 1)^{-1}(z - 2)^{-2} \quad (0 < \lvert z - 1 \rvert < 1).$

Cauchy's integral formula: Consider the contour integral $\oint_C z^n\, dz$ where $n = 0, 1, \ldots$ and C is a simple closed contour that encircles the origin once in the positive sense. As z^n is an entire function, C may be distorted to a circle of arbitrary radius r centered at the origin, and encircling it once in the positive sense. Since $dz = r\, e^{i\theta}\, i\, d\theta$ on this circle, we have

$$\oint_C z^n\, dz = i\, r^n \int_0^{2\pi} d\theta\, e^{(n+1)i\theta} = 0.$$

The vanishing of the contour integral is in any case immediately evident, even without writing it out in the form above: the integrand z^n is holomorphic everywhere inside and on the contour C, so that C can be distorted till it shrinks to a point (i.e., you

can let $r \to 0$), and the integral vanishes.

Now consider the contour integral $\oint_C dz/z^{n+1}$ where $n = 0, 1, \ldots$ and C is again a simple closed contour that encircles the origin once in the positive sense. As the integrand z^{-n-1} has a pole of order $(n+1)$ at the origin, and no other singularities, C can again be deformed to a circle of some radius r centered at the origin. Then

$$\oint_C \frac{dz}{z^{n+1}} = \frac{i}{r^n} \int_0^{2\pi} d\theta \, e^{-ni\theta}.$$

As before, this integral vanishes identically an all cases except one: namely, when $n = 0$. (Alternatively, we can now *expand* the circle outwards, and pass to the limit $r \to \infty$.) Only when $n = 0$, and in that case alone, do we get the nonzero value $\oint_C dz/z = 2\pi i$. Hence we have, for *all* integers $n \in \mathbb{Z}$, $(2\pi i)^{-1} \oint_C dz/z^{n+1} = \delta_{n\,,0}$. A little more generally, if a is any complex number and C is a closed contour encircling a once in the positive sense, then

$$\frac{1}{2\pi i} \oint_C \frac{dz}{(z-a)^{n+1}} = \delta_{n\,,0}\,.$$

This is a most useful representation of the Kronecker delta. Cauchy's Integral Theorem and the whole of the so-called **calculus of residues** is based upon it.

Cauchy's integral formula follows from the result above. Omitting the details, the theorem states: Let C be a closed contour lying entirely in a region in which the function $f(z)$ is analytic, except for isolated poles and essential singularities at the points $z_k\,, k = 1, 2, \ldots$. No singularities must lie on C itself. Then, if C encircles the singularities once in the positive sense,

$$\oint_C dz\, f(z) = (2\pi i) \sum_{\{\text{poles } z_k\}} \text{Res}\, f(z).$$

The sign of the right-hand side is reversed if C encircles the singularities in the negative or clockwise sense. If a singularity is encircled r times, the contribution to the integral is $\pm r$ times the residue at that singularity, the negative sign being applicable when the contour winds around the singularity in the negative sense. (r is called the **winding number** of the contour at the singularity.)

4. Contour integration: Evaluate the following integrals by contour integration. Here $a\,, b\,,$ etc. denote positive constants, and n is a non-negative integer. (The values of the integrals have also been given.)

(a) $$\int_{-\infty}^{\infty} \frac{dx}{(x^2+a^2)(x^2+b^2)} = \frac{\pi}{ab(a+b)}\,.$$

Consider the integral from $-R$ to R on the x-axis, and attach a semicircle of radius R to close the contour either in the upper or lower half-plane. You now have the integral of the function $[(z^2+a^2)(z^2+b^2)]^{-1}$ over a closed contour C. By Cauchy's theorem, this integral is (apart from a factor $\pm 2\pi i$) the sum of the residues of the function at the poles enclosed by C. Let $R \to \infty$. The contribution from the semi-circle vanishes as $R \to \infty$, while the integral along the x-axis becomes the integral you want to evaluate.

(b) $$\int_0^\infty \frac{dx\, x^2}{(x^2+a^2)^3} = \frac{\pi}{16a^3}\,.$$

Write the integral as $\frac{1}{2}$ times the integral from $-\infty$ to ∞ along the x-axis, and proceed as in the preceding case. Note that the pole enclosed by C is a multiple pole in this case.

(c) $$\int_{-\infty}^\infty \frac{dx\, x\, \sin\, x}{(x^2+a^2)} = \pi\, e^{-a}\,.$$

In this case, the contour can't be closed in either the upper or the lower half-plane because $\sin\, z$ is a linear combination of e^{iz} as well as e^{-iz}, and one of the exponentials will diverge as $|z| \to \infty$ in each of the two half-planes. Therefore first write $\sin\, x$ as the imaginary part of e^{ix}, and consider the integral $\int_{-\infty}^\infty dx\, x\, e^{ix}/(x^2+a^2)$. It is clear that you can now close the contour in the upper half-plane. After evaluating the integral, take its imaginary part. Alternatively, retain $\sin\, x$ in the integrand, and note that you have to close the contour in different half-planes for the two exponential factors. Do the integral both ways to check that the final results match.

(d) $$\int_{-\infty}^\infty \frac{dx\, (x^2-x+2)}{x^4+10x^2+9} = \frac{5\pi}{12}\,.$$

(e) $$\int_0^\infty \frac{dx\, \sin\, x}{x} = \frac{\pi}{2}\,.$$

Extend the integral to one from $-\infty$ to ∞, and consider the integrand e^{iz}/z over a contour that is closed in the upper half-plane. The pole at the origin must be avoided by a small semi-circular indentation around $z = 0$. The rest of the integral along the real axis is now a **Cauchy principal value** integral. After evaluation of the integral, equating imaginary parts yields the result sought. Note that the principal value restriction can be dropped at this stage, because $(\sin\, x)/x$ does not have a singularity at $x = 0$.

(f) $$\int_0^{2\pi} \frac{d\theta}{a - b\, \cos\, \theta} = \frac{2\pi}{\sqrt{a^2-b^2}} \quad (a > b)\,.$$

Let $z = e^{i\theta}$. Then $\cos\,\theta = \frac{1}{2}(z + z^{-1})$. The integration over θ becomes an integral over z around the unit circle traversed once in the positive sense. Pick up the residue at the pole of the integrand that lies inside the unit circle.

$$\text{(g)}\ \int_0^{2\pi} d\theta\, e^{\cos\,\theta}\, \cos\,(n\theta - \sin\,\theta) = \frac{2\pi}{n!} \quad (n = 0,\, 1,\, \ldots)\,.$$

The integral required is the real part of $\int_0^{2\pi} d\theta\, e^{\cos\,\theta}\, e^{i(n\theta - \sin\,\theta)}$. Set $e^{i\theta} = z$ to convert the integral to one around the unit circle in the z-plane. Pick up the residue of the integrand at the singularity enclosed by the contour.

Residue at infinity: The residue of an analytic function $f(z)$ at infinity is defined via a contour integral in the same way that it is defined at any finite point $z = a$, namely, as $(2\pi i)^{-1}$ times the contour integral of the function over a sufficiently small circle enclosing (in the positive sense) no singularities other than the possible one at $z = a$. In the case of the point at infinity, we must do this on the Riemann sphere: the contour is a sufficiently small circle going around the point at infinity once in the *negative* sense. On the complex plane, this contour looks like a circle C of sufficiently *large* radius R traversed in the negative sense. On the other hand, this circle now encloses all the singularities of $f(z)$ located in the *finite* part of the plane. Therefore

$$\operatorname*{Res}_{z=\infty} f(z) \stackrel{\text{def.}}{=} \frac{1}{2\pi i} \oint_C dz\, f(z) = -\sum_j \operatorname*{Res}_{z=a_j} f(z),$$

where the sum runs over the singularities at the points a_j in the finite part of the complex plane. Two noteworthy points follow at once.

(i) $f(z)$ *may* have a nonzero residue at $z = \infty$ even if it is *not* singular at the point at infinity!

For instance, the function $(z-1)^{-1} + (z-2)^{-1}$ has simple poles at $z = 1$ (residue $= 1$) and $z = 2$ (residue $= 1$), and is regular at $z = \infty$. And yet it has a nonzero residue at infinity that is equal to the negative of the sum of its residues at $z = 1$ and $z = 2$, namely, -2.

(ii) On the other hand, an entire function *must* have zero residue at infinity, even though it *is*, in general, *singular* at that point!

Thus, a polynomial of order n, which has a pole of order n at $z = \infty$, must have zero residue at that point. Likewise, an exponential such as e^z, which has an essential singularity at ∞, must have zero residue at that point.

There is an equivalent way of finding the residue at infinity. Changing variables from z to $w = 1/z$ converts the contour integral over the circle C of radius R to a

circle c around the origin in the w-plane, with an infinitesimal radius $1/R$, traversed in the *positive* sense. Remembering that $dz = -dw/w^2$, this yields the formula

$$\operatorname*{Res}_{z=\infty} f(z) = -\frac{1}{2\pi i} \oint_c \frac{dw}{w^2}\, f(1/w).$$

It follows that the residue at infnity of $f(z)$ is the coefficient of w^{-1} in the Laurent expansion of $-(1/w^2)f(1/w)$ in powers of w. It is now easily checked that the residue at infinity of $(z-1)^{-1}+(z-2)^{-1}$ is indeed equal to -2, as found earlier. Similarly, it is easy to verify that there is no term proportional to w^{-1} in the power series expansion of $-(1/w^2)f(1/w)$ when $f(z)$ is any polynomial or exponential like e^z.

5. A practical use of the concept of the residue at infinity is in the evaluation of certain contour integrals involving rational functions. Let $p_n(z) = a_n\, z^n + \ldots + a_0$ be a polynomial of order n, and let $q_{n+1}(z) = b_{n+1}z^{n+1} + \ldots + b_0$ be a polynomial of order $(n+1)$. Let C be a simple closed contour that encloses all the roots of $q_{n+1}(z) = 0$ once in the *positive* sense. Show that

$$\oint_C dz\, \frac{p_n(z)}{q_{n+1}(z)} = \frac{2\pi i a_n}{b_{n+1}}\,.$$

Note that this result does not depend on the details of the locations or the multiplicities of the zeroes of $q_{n+1}(z)$.

6. Summation of series using contour integration: The fact that $\pi \cot \pi z$ has a simple pole at each integer $z = n$, with residue equal to $+1$, can be used to sum certain infinite series. Similarly, the fact that $\pi \operatorname{cosec} \pi z$ has simple poles at all integers $z = n$, with residue equal to $(-1)^n$, can be used to sum certain series in which the signs of the terms alternate. The method works when the summand is an even function of n, so that the series can be written as a sum over both positive and negative integers. Evaluate the following infinite sums by the contour integration method. (The values of the sums are also given, so that you can check your answers.)

$$\text{(a)}\;\; S(a) = \sum_{n=1}^{\infty} \frac{1}{n^2 + a^2} = \frac{\pi}{2a}\left(\coth \pi a - \frac{1}{\pi a}\right) \quad (a > 0).$$

First write the sum required as $\frac{1}{2}$ times the sum over positive as well as negative integers n. Each terms in the sum is just the residue of the function $f(z) = \pi \cot(\pi z)/(z^2+a^2)$ at its simple pole at $z = n$. Hence it is $(2\pi i)^{-1}$ times the contour integral of $f(z)$ over a small circle encircling $z = n$ in the positive sense. Merge all these little circles together, to get two 'hairpin' contours straddling the x-axis from $-\infty$ to -1 and from 1 to ∞, respectively. These can be made part of a single closed contour by attaching large semi-circles in the upper and lower half-planes. This contour now encircles just three

poles of $f(z)$, namely, those at $z = 0$ and $z = \pm ia$, but in the clockwise or negative sense. Pick up the residues at these poles and simplify the resulting expression to obtain the final result given above.

(b) $$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}\,.$$

This sum is equal to $\zeta(2)$, where $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$ is the **Riemann zeta function**. The same method as that used in part (a) above will lead to the result quoted. Note that there is a pole of order 3 at $z = 0$ in this instance.

Verify the result above by passing to the limit $a \to 0$ in the sum $S(a)$ found in part (a). You will need to use the fact that $\cot z = z^{-1} + \frac{1}{3}z + \mathcal{O}(z^3)$ in the neighborhood of $z = 0$.

(c) Use the fact that $\zeta(2) = \frac{1}{6}\,\pi^2$ to show (by elementary means) that

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^2} = \frac{\pi^2}{12} \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{1}{(2n-1)^2} = \frac{\pi^2}{8}\,.$$

(d) $$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^2 + a^2} = \frac{\pi}{2a}\left(\frac{1}{\pi a} - \frac{1}{\sinh\,\pi a}\right).$$

Repeat the procedure used in part (a), but with $\operatorname{cosec}(\pi z)$ instead of $\cot(\pi z)$ in the definition of the integrand $f(z)$. Once again, verify the result you obtain by noting that

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^2 + a^2} = S(a) - \tfrac{1}{2}\,S(\tfrac{1}{2}a) \quad \text{where} \quad S(a) = \sum_{n=1}^{\infty} \frac{1}{n^2 + a^2}\,.$$

Finally, pass to the limit $a \to 0$ to check that $\sum_{n=1}^{\infty}(-1)^{n-1}/n^2 = \frac{1}{12}\pi^2$.

(d) $$\sum_{n=1}^{\infty} \frac{1}{(n^2 + a^2)^2} = \frac{\pi}{4a^3}\left(\coth\,\pi a + \frac{\pi a}{\sinh^2 \pi a} - \frac{2}{\pi a}\right).$$

Use the same procedure as in part (a) to establish this result. Verify the answer directly from $S(a)$ by observing that

$$\sum_{n=1}^{\infty} \frac{1}{(n^2 + a^2)^2} = -\frac{1}{2a}\,\frac{dS(a)}{da}\,.$$

(e) Pass to the limit $a \to 0$ in the foregoing result to establish that

$$\zeta(4) = \sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}\,, \quad \text{and hence} \quad \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^4} = \frac{7\pi^4}{720}\,.$$

(f) $$\sum_{n=1}^{\infty} \frac{1}{(n^2+a^2)(n^2+b^2)} = \frac{\pi}{2ab(b^2-a^2)}\,(b\,\coth\,\pi a - a\,\coth\,\pi b) - \frac{1}{2a^2b^2}\,.$$

Verify the result by noting that the sum required is $[S(a) - S(b)]/(b^2 - a^2)$.

(g) Find the sum $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^4 + \alpha^4}$ $(\alpha > 0)$.

It is evident that we must now consider the contour integral of the function $\pi\,\cot\,(\pi z)/(z^4 + \alpha^4)$. Check your answer by writing $(n^4 + \alpha^4)$ as $(n^2 + a^2)(n^2 + b^2)$, where $a = \alpha\, e^{i\pi/4}$ and $b = \alpha\, e^{3\pi i/4} = -\alpha\, e^{-i\pi/4}$, and using (the analytic continuation of) the result of part (f).

7. Solving difference equations using contour integration: Solve the following recursion relations (that is, find c_n as a function of n). In each case, first define the corresponding generating function $f(z) = \sum_{n=0}^{\infty} c_n\, z^n$. Use the recursion relation to obtain the generating function in closed form. It will, in general, turn out to be a rational function in the cases to be considered. The infinite series defining $f(z)$ converges absolutely inside its circle of convergence, and defines an analytic function of z in that region. The inversion formula is of course $c_n = (1/n!)[d^n f/dz^n]_{z=0}$. But what we need here is the more convenient formula

$$c_n = \frac{1}{2\pi i} \oint_C \frac{f(z)\,dz}{z^{n+1}}\,,$$

where C is a simple closed contour enclosing (in the positive sense) only the pole of the integrand at the origin. Evaluating the residue at this $(n+1)^{\text{th}}$ order pole gets you back to the original formula for c_n. But you can also open out the contour to infinity in all directions, picking up the residues at the poles of $f(z)$ (with a minus sign). This will yield the expression sought for c_n much more easily.

(a) $c_{n+2} - c_{n+1} - c_n = 0\,, \quad n \geq 0\,, \quad c_0 = 1\,,\, c_1 = 1.$

The set $\{c_n\}$ in this case is a **Fibonacci sequence** of numbers. (Each member of the sequence is the sum of the preceding two numbers.) The answer for c_n is, in this case,

$$c_n = \frac{1}{2^n\sqrt{5}}\left[(\sqrt{5}+1)^n - (-1)^n(\sqrt{5}-1)^n\right], \qquad n = 0,\, 1,\, \ldots\,.$$

As you know, the irrational number $\tau = \frac{1}{2}(\sqrt{5}+1)$ is called the **golden mean**, and has a large number of interesting properties. So do the Fibonacci numbers, since $c_n = [\tau^n - (-\tau)^{-n}]/\sqrt{5}$. The result above shows that

- the asymptotic (large-n) growth of the Fibonacci sequence is exponential, $\sim e^{\lambda n}$, where $\lambda = \ln \tau \simeq 0.4812$.

(b) $c_{n+2} - c_{n+1} - 2c_n = 0\,, \quad n \geq 0\,, \quad c_0 = 1\,,\, c_1 = 2.$

(c) $c_{n+2} - 2c_{n+1} + c_n = 0\,, \quad n \geq 0\,, \quad c_0 = 1\,,\, c_1 = 1.$ (This recursion relation can be solved by inspection!)

(d) $c_n = c_0 c_{n-1} + c_1 c_{n-2} + c_2 c_{n-3} + \cdots + c_{n-1} c_0\,, \quad c_0 = \frac{1}{4}\,.$

Unlike the preceding examples, the last case is a *nonlinear* difference equation. However, it can be solved quite easily by elementary means. Define the generating function for c_n and obtain a functional equation for it (by examining its square!) Use this to show that

$$c_n = \frac{(2n)!}{4^{n+1}\,(n+1)\,(n!)^2}\,.$$

When $c_0 = 1$, we find

$$c_n = \frac{(2n)!}{(n+1)\,(n!)^2} = \frac{1}{n+1}\binom{2n}{n}.$$

Hence $c_0 = 1$, $c_1 = 1$, $c_2 = 2$, $c_3 = 5$, $c_4 = 14$, $c_5 = 42, \ldots$. These are called **Catalan numbers**. They appear in an enormous number of combinatorial problems.

Mittag-Leffler expansions of meromorphic functions: We have seen that analytic functions can be represented by Taylor series in their regions of holomorphy. In the neighborhood of poles and isolated essential singularities, Taylor series are replaced by Laurent series.

A further generalization is possible for meromorphic functions. Recall that a meromorphic function is one that has, at best, only poles in the finite part of the complex plane. Suppose we are given the locations of all the poles of such a function, and the singular parts at these poles. Can we find a representation for the function that involves a 'sum over poles'? The **Mittag-Leffler expansion** of a meromorphic function provides such a representation. In general, suppose the meromorphic function $f(z)$ has poles at the points a_j $(j = 1, 2, \ldots)$, with singular parts $s_j(z)$. For example, if a_j is a pole of order m, then the singular part[4] is

$$s_j(z) = \frac{c_{-m}^{(j)}}{(z-a_j)^m} + \cdots + \frac{c_{-1}^{(j)}}{(z-a_j)}\,.$$

[4]The singular part is also called the **principal part**, but I prefer the former term because it is more suggestive.

The Mittag-Leffler expansion of $f(z)$ is then of the form

$$f(z) = \sum_j s_j(z) + g(z),$$

where $g(z)$ is an *entire* function. When the number of poles is finite, the sum over j presents no problem. This is the case when $f(z)$ is a **rational function**, i.e., the ratio of two polynomials. In such cases the Mittag-Leffler expansion is nothing but the resolution of the function into partial fractions. For instance, the Mittag-Leffler expansion of the function $f(z) = (z-a)/(z-b)^2$ is just

$$\frac{z-a}{(z-b)^2} = \frac{b-a}{(z-b)^2} + \frac{1}{z-b}\,.$$

When the summation over j involves an *infinite* number of terms, however, the question of the convergence of the sum arises.

Mittag-Leffler expansion of $\cot \pi z$: We have already encountered an example of a Mittag-Leffler expansion that involves an infinite number of terms, or at least an expansion that is related to such an expansion. Recall that

$$S(a) = \sum_{n=1}^{\infty} \frac{1}{n^2+a^2} = \frac{\pi}{2a}\Big(\coth\, \pi a - \frac{1}{\pi a}\Big).$$

Although the infinite series was originally summed for real positive values of a, we may regard the equality above as a relation between two analytic functions of the complex variable a, by virtue of the principle of **analytic continuation**. Set $a = iz$ and note that $\coth\,(i\pi z) = -i\,\cot\,(\pi z)$. The relation above can be written as

$$\pi\,\cot\,(\pi z) = \frac{1}{z} + 2\sum_{n=1}^{\infty} \frac{z}{z^2-n^2} = \sum_{n=-\infty}^{\infty} \frac{z}{z^2-n^2}\,.$$

We're almost there, because this expansion seems to express the meromorphic function $\pi\,\cot\,(\pi z)$ as a sum over the singular parts at its poles (which are located at all integer values of z). But the poles of $\cot\,(\pi z)$ are simple poles, whereas the summand in the last equation involves a quadratic function of z in the denominator. It is a simple matter to write

$$\sum_{n=1}^{\infty} \frac{2z}{z^2-n^2} = \sum_{n=1}^{\infty} \left(\frac{1}{z-n} + \frac{1}{z+n}\right).$$

But we *cannot* go on to split the right-hand side into two sums and write it as

$$\sum_{n=1}^{\infty} \frac{1}{z-n} + \sum_{n=1}^{\infty} \frac{1}{z+n} = \sum_{n=1}^{\infty} \frac{1}{z-n} + \sum_{n=-\infty}^{-1} \frac{1}{z-n}\,,$$

because each of the two individual infinite series in the last line above *diverges*! To find the correct Mittag-Leffler expansion, let's go back to

$$\pi \cot(\pi z) = \sum_{n=-\infty}^{\infty} \frac{z}{z^2 - n^2} = \frac{1}{z} + \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{z}{z^2 - n^2}\,.$$

But

$$\begin{aligned}
\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{z}{z^2 - n^2} &= \tfrac{1}{2} \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{2z}{z^2 - n^2} = \tfrac{1}{2} \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \Big(\frac{1}{z-n} + \frac{1}{z+n}\Big) \\
&= \tfrac{1}{2} \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \Big[\Big(\frac{1}{z-n} + \frac{1}{n}\Big) + \Big(\frac{1}{z+n} - \frac{1}{n}\Big)\Big] \\
&= \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \Big(\frac{1}{z-n} + \frac{1}{n}\Big),
\end{aligned}$$

where we have used (in writing the last line) the fact that the sum over all nonzero integers remains unchanged if $n \to -n$. Hence we obtain

$$\pi \cot(\pi z) = \frac{1}{z} + \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \Big(\frac{1}{z-n} + \frac{1}{n}\Big).$$

This is the Mittag-Leffler expansion of the meromorphic function $\pi \cot(\pi z)$. Note that the term $1/n$ inside the large brackets in the summand on the right-hand side is necessary. Without it, the sum over n of $1/(z-n)$ would diverge. With the $1/n$ term present, the summand is sufficiently well-behaved[5] to guarantee convergence of the series.

Moreover, the series can be differentiated term by term with respect to z. The result is the Mittag-Leffler expansion of the function $\mathrm{cosec}^2(\pi z)$, which reads

$$\frac{\pi^2}{\sin^2 \pi z} = \sum_{n=-\infty}^{\infty} \frac{1}{(z-n)^2}\,.$$

Now step back and admire this remarkably beautiful formula!

[5]It falls off like $1/n^2$ as $|n| \to \infty$.

3 Linear response; dispersion relations

Causal, linear, retarded response: Under very general conditions, the physical response of a system to an applied time-dependent stimulus satisfies certain basic criteria. Let us denote the stimulus (or 'force') by $F(t)$, and the response by $R(t)$. Examples of such stimulus-response pairs are : mechanical force and displacement; electric field and polarization; magnetic field and magnetization; stress and strain; electromotive force and current; and so on. For notational simplicity, I have suppressed the appropriate indices in the stimulus and the response when these are vectors, tensors, etc. in what follows.

Under fairly general conditions, the response at any instant of time t may be assumed to meet three basic requirements:

(i) Causality: The response at time t depends on the force history at all *earlier* instants of time, but *not* later ones. This is the principle of causality, which says that *the effect cannot precede the cause.*

(ii) Linearity: The response is assumed to be *linear* in the applied force. This implies at once that the superposition principle holds good.

(iii) Retarded response: The effect of a force $F(t')$ applied at the instant t' on the response $R(t)$ at a later instant t depends only on the *elapsed* time interval $(t - t')$, rather than on both t and t' individually. This means that the location of the actual origin of the time coordinate, the particular instant at which we set $t = 0$, does not matter.

Taking the force to be applied from $t = -\infty$ onward (this will automatically incorporate all other starting times), the most general linear functional of $F(t)$ subject to these requirements is given by

$$R(t) = \int_{-\infty}^{t} dt'\, \phi(t - t')\, F(t').$$

Note that the upper limit of integration over t' is t, in accordance with the requirement of causality. The quantity $\phi(t - t')$ is called the **response function**. The fact that it is a function of the time difference $(t - t')$ indicates that we are dealing with a *retarded* response. The quantity $\phi(t)$ represents the 'weight' with which a force applied at $t = 0$ contributes to the response $R(t)$ at any later time t. We may expect it to be a decreasing (or at least a non-increasing) function of its argument, such as a decaying exponential. But it could oscillate as a function of t, although we would expect it to do so with decreasing amplitude, in most instances. These are not strict requirements, of course, but they seem to be plausible on physical grounds.

Dynamic susceptibility: It is natural to decompose general time-dependent functions such as $F(t)$ and $R(t)$ into their Fourier (or frequency) components. We have

$$F(t) = \int_{-\infty}^{\infty} d\omega\, e^{-i\omega t}\, \widetilde{F}(\omega) \quad \text{and} \quad R(t) = \int_{-\infty}^{\infty} d\omega\, e^{-i\omega t}\, \widetilde{R}(\omega).$$

The inverse transforms are

$$\widetilde{F}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt\, e^{i\omega t}\, F(t) \quad \text{and} \quad \widetilde{R}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt\, e^{i\omega t}\, R(t).$$

1. Use these relations in the foregoing relation between R and F to show that

$$\widetilde{R}(\omega) = \chi(\omega)\, \widetilde{F}(\omega), \quad \text{where} \quad \chi(\omega) = \int_{0}^{\infty} dt\, e^{i\omega t}\, \phi(t).$$

The function $\chi(\omega)$ is called the **frequency-dependent susceptibility** or **dynamic susceptibility** (corresponding to this particular stimulus-response pair). It measures the response of the system to a sinusoidally varying 'force' of unit amplitude and frequency ω. Observe that $\chi(\omega)$ is *not* the Fourier transform of $\phi(t)$. The lower limit of integration in its definition is $t = 0$ rather than $-\infty$. It is important to note that this lower limit is a direct consequence of the fact that t is the upper limit of integration in the basic expression for $R(t)$. In turn, this follows directly from causality. Thus, the dynamic susceptibility is the *one-sided* Fourier transform (sometimes also called the Fourier-Laplace transform) of the response function $\phi(t)$ because of causality.

Physical examples of the dynamic susceptibility include the AC susceptibility (or frequency-dependent susceptibility) of a magnetic substance, the electric polarizability of a dielectric material, the dynamic mobility of a fluid, the elastic compliances (inverses of the elastic moduli) of a solid, and so on.

Some symmetry properties of the dynamic susceptibility: An important property of the dynamic susceptibility can be extracted from its defining equation above: since $\phi(t)$ is a real quantity, we find that

$$\chi(-\omega) = \chi^*(\omega) \text{ for all } \textit{real } \omega,$$

where $*$ denotes the complex conjugate. I have added the phrase 'for all real ω' has been added because it makes sense, as you'll see shortly, to regard of $\chi(\omega)$ as an analytic function of the *complex* variable ω, although the physically accessible values of the frequency are of course real and non-negative. It follows at once that, for all real ω,

$$\text{Re } \chi(-\omega) = \text{Re } \chi(\omega) \quad \text{and} \quad \text{Im } \chi(-\omega) = -\text{Im } \chi(\omega).$$

The *real* part of the dynamic susceptibility is an *even* function of the frequency, while the *imaginary* part is an *odd* function of the frequency. You can, of course, write down

these properties directly by setting $e^{i\omega t} = (\cos \omega t + i \sin \omega t)$ in the definition of $\chi(\omega)$.

Consider the formula for $\chi(\omega)$ as the one-sided Fourier transform of $\phi(t)$. We have assumed that the integral exists for physical (i.e., real, non-negative) values of the frequency ω. It is obvious that it will continue to do so even if ω is complex, provided Im ω is *positive*. This is because the factor $e^{i\omega t}$ then leads to an extra damping factor $e^{-(\mathrm{Im}\ \omega)\, t}$. Such a factor can only *improve* the convergence of the integral, since the integration over t is restricted to non-negative values. We may therefore conclude that the dynamic susceptibility can be *analytically continued* to the upper half-plane[6] (UHP for short) in ω. The defining expression for $\chi(\omega)$ provides a representation for this analytic function for all $\mathrm{Im}\, \omega \geq 0$.

Given that the dynamic susceptibility is analytic in the UHP, we can extend the relation $\chi(-\omega) = \chi^*(\omega)$ (that holds good for real ω) to complex values of ω. We have

$$\chi(-\omega^*) = \chi^*(\omega), \quad \mathrm{Im}\ \omega \geq 0.$$

Note that if ω lies in the UHP, so does $-\omega^*$. Hence the arguments of the functions on both sides of the equation above do lie in the region in which we are guaranteed that the dynamic susceptibility is analytic. On general grounds, and without further input, we really cannot say very much about its possible behavior in the lower half of the complex ω-plane. But we know that an analytic function of a complex variable cannot be holomorphic at *all* points of the extended complex plane unless it is just a constant, as we know from Liouville's Theorem. In general, therefore, *the dynamic susceptibility will have singularities in the lower half-plane in* ω.

Dispersion relations: The analyticity of $\chi(\omega)$ in the UHP enables us to derive certain relations between its real and imaginary parts. Let us assume that $|\chi(\omega)| \to 0$ as $\omega \to \infty$ along any direction in the UHP. This is a physically plausible assumption to make in most circumstances, for the following reason. Even for *real* ω, we expect the susceptibility to vanish as the frequency becomes very large, because the inertia present in any system will not permit it to respond to a sinusoidal applied force oscillating at a frequency much higher than all the natural frequencies present in the system. And when ω is a complex number with a positive imaginary part, the factor $e^{i\omega t}$ in the formula for the susceptibility lends an extra damping factor $e^{-\,(\mathrm{Im}\ \omega)\, t}$. The assumption is thus a reasonable one to make. However, it is not absolutely essential,

[6] Causality ensures that the Fourier-Laplace transform of a causal response function is analytic in a half-plane in ω. Whether this is the upper half-plane (UHP) or lower half-plane (LHP) depends on the Fourier transform convention chosen. With our particular convention in which a function $f(t)$ of the time is expanded as $\int_{-\infty}^{\infty} d\omega\, \widetilde{f}(\omega)\, e^{-i\omega t}$, it is the *upper* half of the ω-plane in which the susceptibility is analytic. Had we chosen the opposite sign convention and used the factor $e^{+i\omega t}$ in the expansion above, the region of analyticity would have been the *lower* half-plane in ω. The convention I have adopted is the more commonly used one, at least in physics.

and can be relaxed, as you'll see further on.

Let ω be a fixed, real, positive frequency. Consider the quantity

$$f(\omega') = \frac{\chi(\omega')}{\omega' - \omega}$$

as a function of the complex variable ω'. This function is analytic everywhere in the upper half-plane in ω', as well on the real axis in that variable, except for a simple pole at $\omega' = \omega$ located on the real axis. By Cauchy's integral theorem its integral over any closed contour C' lying entirely in the upper half-plane is identically equal to zero Without changing the value of the integral (namely, zero), the contour C' can be expanded to the contour C that comprises the following:

(i) a large semicircle of radius R in the UHP,

(ii) a line integral on the real axis running from $-R$ to $\omega - \epsilon$,

(iii) a small semicircle, from $\omega - \epsilon$ to $\omega + \epsilon$, lying in the UHP so as to avoid the simple pole of the integrand, and, finally,

(iv) a line integral from $\omega + \epsilon$ to R.

Thus

$$\oint_{C'} d\omega'\, f(\omega') = \oint_{C} d\omega'\, \frac{\chi(\omega')}{\omega' - \omega} = 0.$$

In the limit $R \to \infty$, the contribution from the large semicircle vanishes, because $f(\omega')$ vanishes faster than $1/\omega'$ as $\omega' \to \infty$ along all directions in the UHP: this is ensured by the fact that $\chi(\omega') \to 0$ as $|\omega'| \to \infty$. On the small semicircle, we have $\omega' = \omega + \epsilon\, e^{i\theta}$, where θ runs from π to 0. Therefore, in the limit $\epsilon \to 0$, the contribution from the small semicircle tends to $-i\pi\chi(\omega)$. Hence we get

$$\lim_{\epsilon \to 0} \left\{ \int_{-\infty}^{\omega - \epsilon} d\omega'\, \frac{\chi(\omega')}{\omega' - \omega} + \int_{\omega + \epsilon}^{\infty} d\omega'\, \frac{\chi(\omega')}{\omega' - \omega} \right\} - i\pi\chi(\omega) = 0.$$

But the limit on the left-hand side of this equation is just the Cauchy principal value integral of the integrand $\chi(\omega')/(\omega' - \omega)$. Recall that the Cauchy principal value is a specific prescription for avoiding the divergence or infinity that would otherwise arise, owing to the singularity (a simple pole) of the integrand at $\omega' = \omega$. Denoting this principal value integral by $\mathsf{P}\int(\cdots)$, we get

$$\chi(\omega) = -\frac{i}{\pi}\, \mathrm{P}\!\int_{-\infty}^{\infty} d\omega'\, \frac{\chi(\omega')}{\omega' - \omega}\,.$$

Note that this equation expresses the susceptibility at a *real* frequency as a certain weighted sum of the susceptibility over all other *real* frequencies. No complex frequencies appear anywhere in this formula. We made an excursion into the upper half of

the complex ω-plane. This was made possible by the analyticity properties of the susceptibility. But we have returned to the real axis, bringing back the last equation with us. Equating the respective real and imaginary parts of the two sides of this equation, we get

$$\text{Re } \chi(\omega) = \frac{1}{\pi}\,\text{P}\!\int_{-\infty}^{\infty} d\omega'\,\frac{\text{Im } \chi(\omega')}{\omega' - \omega} \quad \text{and} \quad \text{Im } \chi(\omega) = -\frac{1}{\pi}\,\text{P}\!\int_{-\infty}^{\infty} d\omega'\,\frac{\text{Re } \chi(\omega')}{\omega' - \omega}\,.$$

These formulas are called **dispersion relations** (the origin of the term is given below). They imply that the two real functions of a real argument, Re $\chi(\omega)$ and Im $\chi(\omega)$, form a **Hilbert transform pair**.

It should also be evident that the dynamic susceptibility *cannot*, in general, be a *real analytic*[7] function of ω. For, if that were so, the imaginary part Im $\chi(\omega)$ would be zero for all real frequencies, and then so would the real part Re $\chi(\omega)$. Thus $\chi(\omega)$ would itself have to vanish identically.

2. The dispersions relations above still involve integrals over negative as well as positive frequencies, whereas physically accessible frequencies are non-negative. But the symmetry properties of Re $\chi(\omega)$ and Im $\chi(\omega)$ can be used to restrict the range of integration to physically accessible frequencies, namely, $0 \leq \omega' < \infty$. Show that

$$\text{Re } \chi(\omega) = \frac{2}{\pi}\,\mathsf{P}\!\int_{0}^{\infty} d\omega'\,\frac{\omega'\,\text{Im } \chi(\omega')}{\omega'^{\,2} - \omega^2} \quad \text{and} \quad \text{Im } \chi(\omega) = -\frac{2\omega}{\pi}\,\mathsf{P}\!\int_{0}^{\infty} d\omega'\,\frac{\text{Re } \chi(\omega')}{\omega'^{\,2} - \omega^2}\,.$$

The term 'dispersion relation' originates from the fact that they were first derived in the context of optics : the real and imaginary parts of the frequency-dependent refractive index of an optical medium are, respectively, measures of the **dispersion** and **absorption** of radiation by the medium. These are not independent quantities, but are related to each other via dispersion relations, which are also called **Kramers-Kronig relations** in physics. As I have already mentioned, they are a direct consequence of the principle of causality as applied to response functions.

Admittance of an LCR circuit : Consider a series LCR circuit. If the applied emf is $V(t)$, the current $I(t)$ is given by the equation

$$L\frac{dI}{dt} + RI + \frac{1}{C}\int_{-\infty}^{t} I(t')\,dt' = V(t).$$

Now use the Fourier expansions

$$V(t) = \int_{-\infty}^{\infty} d\omega\, e^{-i\omega t}\,\widetilde{V}(\omega) \quad \text{and} \quad I(t) = \int_{-\infty}^{\infty} d\omega\, e^{-i\omega t}\,\widetilde{I}(\omega).$$

[7] A function $g(z)$ is a **real analytic function** of $z = x + iy$ if it is (i) an analytic function of z, and (ii) real when its argument is real, i.e., $g(x)$ is real.

Inserting these in the equation for $I(t)$, we find that the last term on the left-hand side becomes (after a change of integration variable from t' to $\tau = t - t'$)

$$\frac{1}{C}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega t}\,\widetilde{I}(\omega)\,\int_{0}^{\infty} d\tau\, e^{i\omega\tau}.$$

The problem is that the last integral above diverges, i.e., it is formally infinite. We should really have used Laplace transforms instead of Fourier transforms to tackle the problem under consideration! Had we done so, the integral $\int_0^\infty d\tau\, e^{i\omega\tau}$ would have been replaced by $\int_0^\infty d\tau\, e^{-s\tau}$, with s in a region such that $\mathrm{Re}\, s$ large enough to guarantee the convergence of the integral involved. In this case this is simply $\mathrm{Re}\, s > 0$, and the value of the integral is just $1/s$. Let's suppose that this has been done, so that we may interpret $\int_0^\infty d\tau\, e^{i\omega\tau}$ as standing for the *analytic continuation* to $s = -i\omega$ of the corresponding Laplace transform. This gives the value $1/(-i\omega) = i/\omega$ for the apparently divergent integral. With this small technicality taken care of, the Fourier transforms of the current and the voltage are related to each other according to

$$\left(-i\omega L + R + \frac{i}{\omega C}\right)\widetilde{I}(\omega) = \widetilde{V}(\omega).$$

The quantity in the brackets on the left-hand side is of course the **complex impedance** of the circuit, usually denoted by $Z(\omega)$. The corresponding dynamic susceptibility is nothing but the reciprocal of the impedance, namely, the **complex admittance**, customarily denoted by $Y(\omega)$. We have

$$\widetilde{I}(\omega) = Y(\omega)\,\widetilde{V}(\omega),$$

where

$$Y(\omega) = \left(-i\omega L + R + \frac{i}{\omega C}\right)^{-1} = \frac{i\omega}{L\,(\omega^2 + i\gamma\,\omega - \omega_0^2)}\,.$$

You will recognize $\gamma^{-1} = L/R$ as the time constant of an LR circuit, and $\omega_0 = 1/\sqrt{LC}$ as the natural frequency of an undamped LC circuit. Observe that $Y(\omega)$ is analytic in the upper half-plane in ω, as required. Moreover, in this simple case we know its form explicitly. It has two simple poles in the lower half-plane, at

$$\omega_\pm = -\tfrac{1}{2}i\gamma \pm \sqrt{\omega_0^2 - \tfrac{1}{4}\gamma^2}$$

respectively. Further, $Y(\omega)$ vanishes like ω^{-2} as $|\omega| \to \infty$ in the upper half-plane. Its real and imaginary parts must therefore satisfy (unsubtracted) dispersion relations.

3. Verify explicitly that the functions

$$\mathrm{Re}\, Y(\omega) = \frac{\gamma\omega^2}{(\omega^2 - \omega_0^2)^2 + \omega^2\, L^2} \quad \text{and} \quad \mathrm{Im}\, Y(\omega) = \frac{i\omega(\omega^2 - \omega_0^2)}{(\omega^2 - \omega_0^2)^2 + \omega^2\, L^2}$$

satisfy the dispersions relations

$$\text{Re } Y(\omega) = \frac{2}{\pi}\,\mathsf{P}\!\int_0^\infty d\omega'\,\frac{\omega'\,\text{Im } Y(\omega')}{\omega'^{\,2}-\omega^2}\,, \quad \text{Im } Y(\omega) = -\frac{2\omega}{\pi}\,\mathsf{P}\!\int_0^\infty d\omega'\,\frac{\text{Re } Y(\omega')}{\omega'^{\,2}-\omega^2}\,.$$

4. Subtracted dispersion relations: It may so happen that a dynamic susceptibility $\chi(\omega)$ does not vanish as $\omega \to \infty$. Instead, it may tend to a constant as $\omega \to \infty$ along some direction or directions in the region Im $\omega \geq 0$. (It is clear that this is most likely to happen as $\omega \to \infty$ along the real axis itself.) In this case, when we try to derive dispersion relations for the real and imaginary parts of $\chi(\omega)$, we find that the contribution from the large semicircle of radius R no longer vanishes as $R \to \infty$. If $\chi(\omega)$ tends *uniformly* to a constant as $|\omega| \to \infty$ along the real axis and along all directions in the upper half-plane, we could go ahead by including $i\pi\chi_\infty$ as the extra contribution from the large semicircle. But there is no guarantee that this will be the case. In fact, it does not happen, in general.

To get over the problem, we assume that the value of $\chi(\omega)$ at some particular real value of ω_0 of the frequency is known. Then, instead of the function $\chi(\omega')/(\omega'-\omega)$, consider

$$f(\omega') = \frac{\chi(\omega')-\chi(\omega_0)}{(\omega'-\omega_0)(\omega'-\omega)}\,.$$

It is evident that this function does vanish faster than $1/\omega'$ as $|\omega'| \to \infty$, owing to the extra factor $(\omega'-\omega_0)^{-1}$. Moreover, it is analytic everywhere in the upper half ω'-plane and on the real axis, except for a simple pole at $\omega' = \omega$, as before. Observe that it does *not* have any singularity at $\omega' = \omega_0$. This is the reason for subtracting $\chi(\omega_0)$ from $\chi(\omega')$ in the numerator. Repeat the earlier derivation to show that

$$\chi(\omega) = \chi(\omega_0) - \frac{i}{\pi}\,(\omega-\omega_0)\,\text{P}\!\int_{-\infty}^\infty d\omega'\,\frac{\chi(\omega')-\chi(\omega_0)}{(\omega'-\omega_0)(\omega'-\omega)}\,.$$

Hence obtain the dispersion relations satisfied by the real and imaginary parts of $\chi(\omega)$ in this case. These are called **once-subtracted dispersion relations**, and ω_0 is the 'point of subtraction'. From the mathematical point of view, it should be evident that the procedure given above can be extended to cover situations when the analytic function $\chi(\omega)$ actually diverges as $\omega \to \infty$ in the upper half-plane. For instance, if $\chi(\omega) \sim \omega^{n-1}$ asymptotically, where n is a positive integer, we can write down an n-fold subtracted dispersion relation for it. The latter will require n constants as inputs. These could be, for instance, the values of $\chi(\omega)$ at n different frequencies or points of subtraction.

Hilbert transform pairs: Given a real function $g(x)$ of the real variable x, its Hilbert transform is defined as

$$h(x) = \frac{1}{\pi}\,\text{P}\!\int_{-\infty}^\infty dx'\,\frac{g(x')}{x'-x}\,.$$

The inverse transform is given by

$$g(x) = -\frac{1}{\pi}\,\mathrm{P}\!\int_{-\infty}^{\infty} dx'\,\frac{h(x')}{x'-x}\,.$$

Eliminating h, we have

$$g(x) = -\frac{1}{\pi^2}\mathrm{P}\!\int_{-\infty}^{\infty} dx'\;\mathrm{P}\!\int_{-\infty}^{\infty} dx''\frac{g(x'')}{(x'-x)(x''-x')}$$

This may be re-written as

$$g(x) = \int_{-\infty}^{\infty} dx''\,\left\{\frac{1}{\pi^2}\mathrm{P}\!\int_{-\infty}^{\infty}\frac{dx'}{(x'-x)(x'-x'')}\right\}\,g(x'').$$

Hence the quantity in the curly brackets in the equation above must be the unit operator in function space, i.e., we must have

$$\frac{1}{\pi^2}\mathrm{P}\!\int_{-\infty}^{\infty}\frac{dx'}{(x'-x)(x'-x'')} = \delta(x-x'').$$

5. Verify this identity explicitly. First consider $x \neq x''$, and show that the integral on the left-hand side vanishes identically. Next, for any sufficiently well-behaved function $h(x)$, show that

$$\int_{-\infty}^{\infty} dx''\,\left\{\frac{1}{\pi^2}\mathrm{P}\!\int_{-\infty}^{\infty}\frac{dx'}{(x'-x)(x'-x'')}\right\}\,h(x'') = h(x),$$

thereby establishing the required representation of the δ-function.

6. Discrete relaxation spectrum: As we have seen, the real and imaginary parts of the dynamic susceptibility constitute a Hilbert transform pair. Given that

$$\mathrm{Re}\,\chi(\omega) = \sum_{j=1}^{n}\frac{\sigma_j}{1+\omega^2\,\tau_j^2}$$

where σ_j and τ_j are positive constants, use the dispersion relation to show that

$$\mathrm{Im}\,\chi(\omega) = \sum_{j=1}^{n}\frac{\omega\,\sigma_j\,\tau_j}{1+\omega^2\,\tau_j^2}\,.$$

Hence

$$\chi(\omega) = \sum_{j=1}^{n}\frac{\sigma_j}{1-i\omega\,\tau_j}\,.$$

Thus $\chi(\omega)$ has a set of poles in the lower half-plane in ω, at $\omega = -i/\tau_j$, $1 \leq j \leq n$. Physically, this corresponds to a **discrete relaxation spectrum** comprising the relaxation times τ_j , contributing with corresponding weights σ_j . The response function $\phi(t)$ itself is a weighted sum of decaying exponentials e^{-t/τ_j}. In a physical situation in which there is essentially just a single relaxation time τ (or when it is a good approximation to assume that this is so), we have the well-known case of **Debye relaxation**. In this instance, the so-called **Cole-Cole plot** or the Argand diagram of $\chi(\omega)$, i.e., a plot of (the suitably scaled variables) $\mathrm{Im}\,\chi$ versus $\mathrm{Re}\,\chi$, obtained by eliminating ω, is the upper half of semi-circle of radius $\frac{1}{2}$ centered at $(\frac{1}{2}, 0)$. Departures of the Cole-Cole plot from this curve serve as an indication that the relaxation is not of the single-exponential form.

7. Continuous relaxation spectrum: If the number of contributing relaxation modes is infinite, or if there is a continuum of such modes, much more complex dynamical behavior may result. I will not go into these aspects here. Given that

$$\mathrm{Re}\,\chi(\omega) = \int_{\tau_{\min}}^{\tau_{\max}} \frac{\sigma(\tau)\,d\tau}{1+\omega^2\,\tau^2}\,,$$

show that

$$\mathrm{Im}\,\chi(\omega) = \int_{\tau_{\min}}^{\tau_{\max}} \frac{\omega\tau\sigma(\tau)\,d\tau}{1+\omega^2\,\tau^2}\,.$$

Hence

$$\chi(\omega) = \int_{\tau_{\min}}^{\tau_{\max}} \frac{\sigma(\tau)\,d\tau}{1-i\omega\,\tau}\,.$$

The point is that the singularity structure of $\chi(\omega)$ in the lower half-plane in ω may now be quite intricate. (It continues to remain holomorphic in the UHP, of course.) In general, a functional form such as that given by the last equation above will have logarithmic branch points at $\omega = -i/\tau_{\max}$ and $\omega = -i/\tau_{\min}$. (For instance, take $\sigma(\tau)$ to be a constant independent of τ.) Even more intricate possibilities arise when $(\tau_{\min}\,,\,\tau_{\max}) = (0\,,\,\infty)$. These, in turn, are associated with physically interesting kinds of **non-Debye relaxation** of response functions, including 'glassy' or 'slow dynamics', 'stretched-exponential decay', etc., in various condensed matter systems, notably those possessing some kind of frozen or 'quenched' disorder.

4 Analytic continuation and the gamma function

The gamma function: In practice, there are numerous methods of analytically continuing a function from a region in which a local representation of the function is available, to a larger region. Some of these are: chains of rearranged power series; the Schwarz reflection principle (for real analytic functions); the use of the 'permanence' of functional equations; summability methods for power series; integral representations for functions in terms of contour integrals, using the distortability of the contours; and so on.

A good illustration is provided by the gamma function, which may be defined (to start with) by the integral

$$\Gamma(z) = \int_0^\infty dt\, t^{z-1}\, e^{-t}, \quad \mathrm{Re}\, z > 0.$$

As you know, $\Gamma(n+1) = n!$ for positive integer values of n. The integral above serves also to define 0! as unity. The question is, how do we analytically continue the gamma function to the region $\mathrm{Re}\, z \leq 0$ in the complex z-plane?

It is obvious that the problem with the convergence of the integral above for $\mathrm{Re}\, z \leq 0$ arises from the lower limit of integration in t. The behavior of the integrand at this end-point can be improved by integration by parts. *Keeping* $\mathrm{Re}\, z > 0$*, integrate by parts* to obtain

$$\begin{aligned}\Gamma(z) &= \int_0^\infty dt\, t^{z-1}\, e^{-t} \quad (\mathrm{Re}\, z > 0) \\ &= \underbrace{\left[\frac{t^z\, e^{-t}}{z}\right]_{t=0}^{t=\infty}}_{=0 \text{ because } \mathrm{Re}\, z>0} + \frac{1}{z}\int_0^\infty dt\, t^z\, e^{-t} \quad (\mathrm{Re}\, z > 0)\end{aligned}$$

But the final integral on the right-hand side converges in the extended region $\mathrm{Re}\, z > -1$, while the factor $1/z$ multiplying it suggests that the gamma function itself has a simple pole at $z = 0$. And indeed it does, because the integral $\int_0^\infty dt\, t^z\, e^{-t}$ is finite and nonzero at $z = 0$. It is equal to 1 when $z = 0$, so that we may conclude that $\Gamma(z)$ has a simple pole at $z = 0$ with residue equal to unity. We have thus extended the region in which we have an explicit representation of the gamma function by the extra strip $-1 < \mathrm{Re}\, z \leq 0$, and we may write

$$\Gamma(z) = \frac{1}{z}\int_0^\infty dt\, t^z\, e^{-t}, \quad \mathrm{Re}\, z > -1.$$

Continuing this procedure of integrating by parts, it is easy to see that we have

$$\Gamma(z) = \frac{1}{z(z+1)\cdots(z+n)}\int_0^\infty dt\, t^{z+n}\, e^{-t}, \quad \mathrm{Re}\, z > -n-1.$$

The integral on the right-hand side converges for $\mathrm{Re}\, z > -n-1$, for any $n = 0, 1, 2, \ldots$, while the factor multiplying it makes the poles at $z = 0, 1, \ldots, -n$ explicit. We have thus achieved an analytic continuation of the gamma function, strip by strip, to the region $\mathrm{Re}\, z > -n-1$, starting with a representation for the function that was only valid in the region $\mathrm{Re}\, z > 0$. In principle, this can be done for arbitrarily large value of n. The underlying mechanism is summarized in the *functional equation* satisfied by the gamma function, namely,

$$\Gamma(z+1) = z\,\Gamma(z) \quad \text{or} \quad \Gamma(z) = \frac{1}{z}\,\Gamma(z+1).$$

The 'permanence' of this equation (i.e., its validity for all z) can be used to analytically continue the gamma function arbitrarily far to the left of the imaginary axis in the complex z-plane.

1. Connection with the Gaussian integral: The gamma function helps us express a number of very useful integrals in closed form.

(a) Show, by a simple change of variables, that

$$\int_0^\infty du\, u^n\, e^{-au^2} = \tfrac{1}{2}\Gamma\left(\tfrac{1}{2}(n+1)\right)\, a^{-(n+1)/2}\,, \quad a > 0,\ n > -1.$$

Note, in particular, that n need not be an integer in this formula. In fact, as you can readily guess, the formula continues to be valid for complex values of n and a, provided $\mathrm{Re}\, n > -1$ and $\mathrm{Re}\, a > 0$. Setting $n = 0$ and $a = 1$ in the formula, we may turn it around to obtain

$$\Gamma\left(\tfrac{1}{2}\right) = 2\int_0^\infty du\; e^{-u^2} = \sqrt{\pi},$$

where we have made use of the known result for the basic Gaussian integral.

(b) It follows at once that the value of the gamma function of a half-odd-integer can be written down explicitly. Use the functional equation for the gamma function to show that

$$\Gamma\left(n+\tfrac{1}{2}\right) = \frac{(2n)!\,\sqrt{\pi}}{2^{2n}\,n!} \quad \text{and} \quad \Gamma\left(-n+\tfrac{1}{2}\right) = (-1)^n \frac{2^{2n}\,n!\,\sqrt{\pi}}{(2n)!}\,, \quad n = 0, 1, 2, \ldots$$

Analytic properties of the gamma function: It is clear from the analytic continued form of the gamma function (obtained either by integration by parts, or from the functional equation) that $\Gamma(z)$ has simple poles at all the non-positive integers, $n = 0, -1, -2, \ldots$. Its residue at $z = -n$ is equal to $(-1)^n/n!$. Other than these

simple poles, it has no other singularities in the finite part of the complex plane. It is therefore a meromorphic function. It is then natural to ask for the Mittag-Leffler expansion of $\Gamma(z)$—namely, a representation that explicitly separates out the contribution from all the poles of the function from the part that is an entire function.

2. Mittag-Leffler expansion of $\Gamma(z)$: Note that the pole at $z = -n$ arises from the behavior of the factor t^{z+n-1} at the lower limit of integration, $t = 0$, in the integral representation of $\Gamma(z)$. Write

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

The second integral on the right-hand side in the last equation above does not have any singularities in the region $|z| < \infty$, and is therefore an entire function of z. In the integral from $t = 0$ to $t = 1$, expand e^{-t} in powers of t and integrate term by term, to obtain the Mittag-Leffler expansion of the gamma function:

$$\Gamma(z) = \underbrace{\sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,(z+n)}}_{\text{sum over pole terms}} + \underbrace{\int_1^\infty dt\, t^{z-1}\, e^{-t}}_{\text{entire function}}.$$

3. The logarithmic derivative of $\Gamma(z)$ is also known as the **digamma function**, and is denoted by $\psi(z)$. We have

$$\psi(z) = \frac{d}{dz}\Gamma(z) = \frac{\Gamma\,'(z)}{\Gamma(z)}\,.$$

The functional relation $\Gamma(z+1) = z\,\Gamma(z)$ leads at once to the difference equation

$$\psi(z+1) - \psi(z) = \frac{1}{z}\,.$$

From the fact that

$$\Gamma(z) = \frac{(-1)^n}{n!\,(z+n)} + \text{regular part}$$

in the neighborhood of its simple pole at $z = -n$, it follows that $\Gamma\,'(z)$ has a double pole at $z = -n$, and has the behavior

$$\Gamma\,'(z) = -\frac{(-1)^n}{n!\,(z+n)^2} + \text{regular part.}$$

Hence show that $\psi(z)$ has simple poles at $n = 0, -1, -2, \ldots$, with residue equal to -1 at each pole.

Infinite product representation for the gamma function: Recall that the harmonic series $\sum_{n=1}^{N} n^{-1}$ diverges logarithmically as $N \to \infty$. That is, the sum has a leading asymptotic behavior $\sim \ln N$ for large values of N. The question that arises naturally is: does the *difference* $\left(\sum_{n=1}^{N} n^{-1}\right) - \ln N$ tend to a finite limit as $N \to \infty$? It turns out that

$$\lim_{N\to\infty} \left\{ \left(\sum_{n=1}^{N} n^{-1} \right) - \ln N \right\} = \gamma,$$

where γ is a number (conjectured to be irrational[8]) called the **Euler-Mascheroni constant**. Its numerical value is $0.5772\ldots$. This constant has many close links with the gamma function. For instance, we know that $\Gamma(z)$ has a simple pole at $z = 0$, with residue equal to 1. If we subtract out this singular part, and then we pass to the limit $z \to 0$, we get

$$\lim_{z\to 0} \left\{ \Gamma(z) - \frac{1}{z} \right\} = -\gamma.$$

In other words, the behavior of the gamma function near in the neighborhood of $z = 0$ is given by

$$\Gamma(z) = \frac{1}{z} - \gamma + \mathcal{O}(z).$$

Another relationship is

$$\psi(1) = \Gamma'(1) = -\gamma.$$

The constant γ occurs explicitly in the following representation of the gamma function. $\Gamma(z)$ is a meromorphic function of z. Moreover, it has no zeroes. Hence its reciprocal is an entire function of z. These properties are explicit in the infinite product representations

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

and hence

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

Taking logarithms and differentiating with respect to z, this gives

$$\psi(z) = -\frac{1}{z} - \gamma + \sum_{n=1}^{\infty} \frac{z}{n(n+z)}\,.$$

Setting $z = 1$ in this expression, we recover the result $\psi(1) = -\gamma$.

The beta function $B(m, n)$ is defined as

$$B(m,n) = \int_0^1 dt\, t^{m-1} (1-t)^{n-1},$$

[8] In fact, a transcendental number, rather than an algebraic irrational.

where m and n are positive integers. It is clear, however, that the same integral representation can be used to define the beta function as a function of *two* complex variables z and w, say, according to

$$B(z,w) = \int_0^1 dt\, t^{z-1}\,(1-t)^{w-1}, \quad \mathrm{Re}\, z > 0 \quad \text{and} \quad \mathrm{Re}\, w > 0.$$

$B(z,w)$ is an analytic function of its arguments in the region indicated. As in the case of the gamma function, we can try to extend the region of analyticity to the left in the z-plane by integrating by parts with respect to t. However, while integrating the factor t^{z-1} will improve matters in the z-plane, it will also worsen matters in the w-plane, because the factor $(1-t)^{w-1}$ will then get differentiated, and lead to the factor $(1-t)^{w-2}$. The opposite will happen if we integrate the factor $(1-t)^{w-1}$ and differentiate the factor t^{z-1}. In fact, a change of variables from t to $(1-t)$ easily establishes the symmetry property

$$B(z,w) = B(w,z).$$

The actual structure of the beta function is clarified by relating it to the gamma function as follows.

4. Consider the double integral

$$I(z,w) = \int_0^\infty du \int_0^\infty dv\, e^{-(u^2+v^2)}\, u^{2z-1}\, v^{2w-1},$$

which converges in the region $\mathrm{Re}\, z > 0$ and $\mathrm{Re}\, w > 0$. Using the formula connecting the Gaussian integral to the gamma function,

$$I(z,w) = \tfrac{1}{4}\,\Gamma(z)\,\Gamma(w).$$

Now go over to plane polar coordinates $u = r\,\cos\theta$ and $v = r\,\sin\theta$, to get

$$I(z,w) = \tfrac{1}{2}\Gamma(z+w) \int_0^{\pi/2} d\theta\,(\cos\theta)^{2z-1}\,(\sin\theta)^{2w-1}.$$

Set $t = \cos^2\theta$ to show that

$$\int_0^{\pi/2} d\theta\,(\cos\theta)^{2z-1}\,(\sin\theta)^{2w-1} = \tfrac{1}{2}\,B(z,w).$$

Hence, equating the two expressions for $I(z,w)$, arrive at the basic relationship

$$B(z,w) = \frac{\Gamma(z)\,\Gamma(w)}{\Gamma(z+w)}\,.$$

As always, this equation must be regarded as one between two analytic functions, valid for all values of the arguments z and w. It leads to a number of important properties

of the gamma function.

Reflection formula for $\Gamma(z)$: Set $w = 1 - z$ in the foregoing relation. Hence $\Gamma(z)\,\Gamma(1-z) = B(z, 1-z)$. In the strip region given by $0 < \mathrm{Re}\, z < 1$, we may use the defining representation of the beta function to write

$$\Gamma(z)\,\Gamma(1-z) = \int_0^1 dt\, \frac{t^{z-1}}{(1-t)^z} \quad (0 < \mathrm{Re}\, z < 1).$$

Changing variables to $u = t/(1-t)$, this becomes

$$\Gamma(z)\,\Gamma(1-z) = \int_0^\infty du\, \frac{u^{z-1}}{u+1} \quad (0 < \mathrm{Re}\, z < 1).$$

The integral on the right-hand side can be evaluated by contour integration.[9] The result is

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

This is the formula sought. Once again, it is a relation between analytic functions. Therefore, by analytic continuation, it is valid for all z. Note that $\mathrm{cosec}\,(\pi z)$ has a simple pole at every integer value of z. On the left-hand side, $\Gamma(z)$ supplies the poles at zero and the negative integers, while $\Gamma(1-z)$ has poles at the positive integers. Taking logarithmic derivatives on both sides, it follows that

$$\psi(1-z) - \psi(z) = \pi\,\cot\,(\pi z).$$

5. Using the fact that $\Gamma(z)$ is a real analytic function of z, i.e., it is real when z is real, show that

$$\left|\Gamma\left(\tfrac{1}{2} + iy\right)\right|^2 = \pi\,\mathrm{sech}\,(\pi y).$$

6. Legendre's duplication formula for $\Gamma(z)$ is a very useful identity that expresses $\Gamma(2z)$ as a product of $\Gamma(z)$ and $\Gamma\left(z + \frac{1}{2}\right)$. Start with

$$B(z, z) = \int_0^1 dt\,[t\,(1-t)]^{z-1}, \quad \mathrm{Re}\, z > 0.$$

But the integrand in the above is symmetric about the mid-point of integration, $t = \frac{1}{2}$. Hence

$$B(z, z) = 2\int_0^{1/2} dt\,[t\,(1-t)]^{z-1}.$$

[9] You'll see how this can be done after we discuss multivalued functions and branch points.

Change variables to $u = 4t(1-t)$, to get

$$B(z,z) = 2^{1-2z}\, B\left(z, \tfrac{1}{2}\right).$$

Now use the relation between the beta and gamma functions to obtain Legendre's duplication formula, namely,

$$\Gamma(2z) = \frac{2^{2z-1}}{\sqrt{\pi}}\, \Gamma(z)\, \Gamma\left(z+\tfrac{1}{2}\right).$$

The formula is actually a special case of a **multiplication theorem** for the gamma function, which reads

$$\Gamma(nz) = (2\pi)^{(1-n)/2}\, n^{nz-\frac{1}{2}} \prod_{j=0}^{n-1} \Gamma\Big(z + \frac{j}{k}\Big), \quad n = 1, 2, \ldots .$$

7. Taking logarithmic derivatives in the duplication formula, we get

$$\psi(2z) = \ln\, 2 + \tfrac{1}{2}\psi(z) + \tfrac{1}{2}\psi\left(z+\tfrac{1}{2}\right).$$

Use this relation, and the difference equation for $\psi(z)$, to establish the following:

(a) $\psi(z+n) = \psi(z) + \displaystyle\sum_{j=1}^{n} \frac{1}{(z+j-1)}\,.$

(b) $\psi(n+1) = -\gamma + \displaystyle\sum_{j=1}^{n} \frac{1}{j}\,.$

(c) $\psi\left(\tfrac{1}{2}\right) = -\gamma - 2\, \ln\, 2\,.$

(d) $\psi\left(n+\tfrac{1}{2}\right) = -\gamma - 2\, \ln\, 2 + 2\displaystyle\sum_{j=1}^{n} \frac{1}{(2j-1)}\,.$

5 Möbius transformations

Conformal mapping: Every analytic function $f(z)$ is a map of some region $\mathcal{R}$ of the complex plane to some region $\mathcal{R}'$ of the complex plane. In other words, given a complex number $z \in \mathcal{R}$, the map $f : z \mapsto w$ produces another complex number $w \in \mathcal{R}'$. Under such a mapping, points in $\mathcal{R}$ map to points in $\mathcal{R}'$, curves to curves, and areas to areas, in general.

- The special feature of an *analytic* function $f(z)$ is that this map is *conformal*. That is, it preserves the angles between curves.

Conformal mapping is a standard and very well-studied topic in complex analysis. It is also one that has a number of practical uses—for instance, in fluid dynamics and aerodynamics. We will not go into conformal mapping *per se* here. Instead, in this chapter, the focus is on one specific kind of conformal map, because it is a very special one in a sense that will become clear shortly. It is also of fundamental importance in many areas of mathematical physics.

A natural question that arises is whether the mapping is *one-to-one*: namely, is there a *unique* $w = f(z)$ for every given z, and *vice versa*? If the function $f(z)$ is single-valued, then by definition there is a unique w for every z. The converse is not necessarily true, of course. It is clear that a *nonlinear* function such as $f(z) = z^2$ does not satisfy this requirement. While there is a unique $w = z^2$ for every z, we cannot find z uniquely if we are given w. We can only do so up to an overall sign. Or else, as you know, we agree to map the complex w-plane onto a *two*-sheeted Riemann surface, on which $z = w^{1/2}$ is single-valued. On the other hand, it may well be possible to map of some specific, restricted part $\mathcal{R}$ of the complex plane to another such part $\mathcal{R}'$, in a one-to-one manner. Indeed, in most of the standard applications of conformal mapping, this is all that is sought.

The more general question of whether there is a one-to-one conformal mapping $\mathbb{C} \mapsto \mathbb{C}$ has a simple and rather trivial answer: The most general one-to-one mapping that takes the whole of the *finite* part of the complex plane to itself is the linear map

$$f(z) = az + b,$$

where a and b are complex constants. Further, the point at ∞ is mapped to itself under this map, i.e., ∞ is a **fixed point** of this map. In terms of the real and imaginary parts of $z = x + iy$, this map amounts to the linear transformations

$$x \mapsto a_1\, x - a_2\, y + b_1\,, \quad y \mapsto a_2\, x + a_1\, y + b_2\,,$$

where $a = a_1 + ia_2$ and $b = b_1 + ib_2$. In geometrical terms, this map is made up (in general) of a rotation, dilation and shear in the xy-plane, followed by a shift of the origin. Interesting as it is, the map is quite simple, and the matter appears to end

here. But this is not quite so, as you will see below.

Definition of a Möbius transformation: Things become much more interesting when we consider the *extended* complex plane $\hat{\mathbb{C}}$ (that is, we include the point at infinity). The linear transformation just discussed leaves the point at infinity unaffected, i.e., it maps it to itself. What if we consider transformations that are not restricted in this manner? Remember that $\hat{\mathbb{C}}$ is essentially equivalent to the Riemann sphere $\mathcal{S}$. A nontrivial result emerges now.

- The most general one-to-one, conformal map $\hat{\mathbb{C}} \mapsto \hat{\mathbb{C}}$ of the extended complex plane to itself (or of the Riemann sphere, $\mathcal{S} \to \mathcal{S}$) is of the form

$$w = f(z) = \frac{az+b}{cz+d}, \quad (ad - bc \neq 0)$$

  where a, b, c, d are finite, complex constants. This is called a **fractional linear transformation** or a **Möbius transformation**.[10]

Note that the condition $ad - bc \neq 0$ is required. (If $ad - bc = 0$, then $w = f(z)$ trivially reduces to a constant for every z, which is not very interesting.) It is obvious that we can divide all four of the four constants a, b, c and d by any constant without affecting the ratio $(az + b)/(cz + d)$ that defines the transformation. In particular, we can divide through by $(ad - bc)^{1/2}$. In effect, therefore, $(ad - bc)$ can always be taken to be equal to 1, without loss of generality. We'll assume henceforth that this has been done, unless specified otherwise.

The *inverse* of a Möbius transformation is also a Möbius transformation. We have

$$z = f^{-1}(w) = \frac{dw - b}{-cw + a}.$$

The same determinant condition, $ad-bc = 1$, remains valid for this transformation too.

Every Möbius transformation maps a certain point in the z-plane to ∞. If $c \neq 0$, this point is $z = -d/c$. Similarly, the point at infinity maps to a/c. That is,

$$z = -\frac{d}{c} \mapsto w = \infty \quad \text{and} \quad z = \infty \mapsto w = \frac{a}{c}.$$

If $c = 0$, we have just a linear transformation, and $z = \infty$ maps to $w = \infty$.

The fixed points of a Möbius transformation, i.e., points that map to themselves under the transformation, are of fundamental importance. They help classify the

[10]The latter name is more commonly used when the transformation is regarded as a mapping of the Riemann sphere, and I will use this term. There are several other names (!) for this transformation—*linear fractional transformation*, *homographic transformation*, *projective transformtion*, and *bilinear transformation*.

transformations into distinct types, as you'll see later. The fixed points are the roots of the equation $f(z) = z$, or

$$cz^2 + (d - a)z - b = 0.$$

It is immediately clear that a Möbius transformation can have at most *two* fixed points, which I'll denote by ξ_1 and ξ_2. (The only exception is the identity transformation, which of course maps every point to itself.) Two cases arise, each with a sub-case, so that there are four possibilities:

(i) ***Finite, distinct fixed points*:** In the general case, when $c \neq 0$ and $a + d \neq \pm 2$, there are two distinct, finite, fixed points given by

$$\xi_{1,2} = \frac{a - d \pm \sqrt{(a + d)^2 - 4}}{2c},$$

using the fact that $ad - bc = 1$.

(ii) ***Finite, coincident fixed points*:** If $c \neq 0$ but $a + d = \pm 2$, the two fixed points coincide, and we have

$$\xi_1 = \xi_2 = \frac{a - d}{2c} = \frac{a \mp 1}{c}.$$

(iii) ***One fixed point at ∞*:** If $c = 0$, the Möbius transformation reduces to the *linear* transformation

$$w = \frac{az + b}{d} = a^2 z + ab,$$

since $ad = 1$ in this case. The fixed points are then at

$$\xi_1 = \frac{b}{d - a} = \frac{ab}{1 - a^2} \quad \text{and} \quad \xi_2 = \infty.$$

(iv) ***Both fixed points at ∞*:** If $c = 0$ and further $a = d$, then the linear transformation is merely a shift of the origin,

$$w = z + \frac{b}{a} = z \pm b,$$

since $a^2 = 1$ in this case. The fixed points then coincide at infinity,

$$\xi_1 = \xi_2 = \infty.$$

We'll return to the role played by these fixed points shortly.

The cross-ratio of four points is a basic concept in geometry. It plays a fundamental role in the subject of *projective geometry*. It is extremely useful in the context of Möbius transformations, as many results can be established quite easily with the help of its properties.

Let z_1 , z_2 , z_3 and z_4 be an ordered set of four distinct points in the complex plane. The cross-ratio of these points is defined as

$$[z_1\ ,\ z_2\ ;\ z_3\ ,\ z_4] \stackrel{\text{def.}}{=} \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)}\,.$$

(Draw a figure showing the distances involved in the cross-ratio, for a general configuration of points.) It is obvious that the cross-ratio of four points depends on the order in which they are specified. Hence there are $4! = 24$ possible cross-ratios that can be associated with an unordered set of four points. However, there are several obvious symmetries among these, and the 24 cross-ratios are related to each other. Let's write $[z_i\ ,\ z_j\ ;\ z_k\ ,\ z_l]$ as $[ij;kl]$, for brevity. It is easy to see that

$$[ij;kl] = 1/[ij;lk] = 1/[ji;kl] = [ji;lk] = [kl;ij].$$

The identities

$$[ik;jl] = 1 - [ij;kl] \quad \text{and} \quad [il;jk] = 1 - 1/[ij;kl]$$

are also easily established. Using these, it is possible to express all 24 cross-ratios among any set of 4 points in terms of any one cross-ratio (see below). If one of the four points in a cross-ratio is the point at infinity, the cross-ratio is defined by a limiting process. For instance, if $z_1 = \infty$, we have

$$[\infty\ ,\ z_2\ ;\ z_3\ ,\ z_4] = \lim_{z_1 \to \infty} [z_1\ ,\ z_2\ ;\ z_3\ ,\ z_4] = \frac{(z_2 - z_4)}{(z_2 - z_3)}\,.$$

Numerous results can be derived from the following fundamental property:

- A Möbius transformation leaves the cross-ratio of any four arbitrary points unaltered in value.

1. The properties of the cross-ratio stated above are easily established.

(a) Find the relations between the 24 cross-ratios associated with any four distinct points $(z_1\ ,\ z_2\ ,\ z_3\ ,\ z_4) \in \mathbb{C}$.

(b) Verify that $[z_1\ ,\ z_2\ ;\ z_3\ ,\ z_4] = [w_1\ ,\ w_2\ ;\ w_3\ ,\ w_4]$, where $w_i = (az_i{+}b)/(cz_i{+}d),\ i = 1, 2, 3, 4$.

From the invariance of the cross-ratio under a Möbius transformation, the following property can be established:

- Given two sets of three distinct points $(z_1\,,\,z_2\,,\,z_3)$ and $(w_1\,,\,w_2\,,\,w_3)$ on the Riemann sphere, there exists a unique Möbius transformation that maps z_i to w_i for $i = 1, 2$ and 3.

There are several ways to prove this assertion. We won't go into the complete proof here, but it is possible to *construct* the Möbius transformation concerned quite easily:

2. Assume that such a Möbius transformation exists. Construct it explicitly.

Hint: Consider the cross-ratio $[z\,,\,z_1\,;\,z_2\,,\,z_3]$. If $z \mapsto w$ under the transformation, then this cross-ratio must be equal to $[w\,,\,w_1\,;\,w_2\,,\,w_3]$. Solve this (simple) equation for w, to show that it is in the form of a Möbius transformation of z.

And now we come to a crucial geometrical property of Möbius transformations. Since a circle can be drawn through any three distinct points, the result stated above has the following implication:

- A Möbius transformation maps circles to circles on the Riemann sphere.
- In the complex plane, a Möbius transformation maps circles and straight lines to circles and straight lines. (Recall that a straight line in the complex plane is a circle passing through the point of projection on the Riemann sphere.)

3. Let $z_1\,,\,z_2\,,\,z_3$ be three arbitrary finite points in the complex plane. Show that the Möbius transformation that maps the circle passing through these three points (or the straight line on which they lie, if they are collinear) to the real axis such that $z_1 \mapsto 0,\ z_2 \mapsto 1,\ z_3 \mapsto \infty$ is

$$w = \frac{(z - z_1)(z_3 - z_2)}{(z - z_3)(z_1 - z_2)}\,.$$

Hint: Use the fact that we must have $[z\,,\,z_1\,;\,z_2\,,\,z_3] = [w\,,\,0\,;\,1\,,\,\infty]$ under the transformation sought. Solve this equation to find w as a function of z.

Normal form of a Möbius transformation: The original form of a Möbius transformation, $z \mapsto w = (az + b)/(cz + d)$ with $ad - bc = 1$, is not the most convenient way to express the transformation for some purposes. The so-called *normal form* provides a better representation. It involves the fixed points of the transformation, comprising the four cases listed earlier.

(i) ***Finite, distinct fixed points*** ξ_1 ***and*** ξ_2**:** Consider the four points z, ∞, ξ_1 and ξ_2, where z is variable. Under the Möbius transformation, these points are mapped according to

$$z \mapsto w, \quad \infty \mapsto \frac{a}{c}, \quad \xi_1 \mapsto \xi_1, \quad \xi_2 \mapsto \xi_2 .$$

Since the cross-ratio does not change under the mapping, we have

$$[z, \infty ; \xi_1 , \xi_2] = [w, a/c ; \xi_1 , \xi_2].$$

Writing this out explicitly, we have the following result. Define the constant

$$K \stackrel{\text{def.}}{=} \frac{a - c\xi_1}{a - c\xi_2} = \frac{a + d - \sqrt{(a+d)^2 - 4}}{a + d + \sqrt{(a+d)^2 - 4}} .$$

Then the transformation can be expressed in the form

$$\frac{w - \xi_1}{w - \xi_2} = K \left(\frac{z - \xi_1}{z - \xi_2} \right).$$

This is the normal form of the Möbius transformation. The constant K is called the *multiplier* of the Möbius transformation. Observe that it depends solely on the sum $(a + d)$. The importance of this fact will become clear subsequently.

(ii) ***Finite, coincident fixed points*****:** Recall that this case corresponds to $c \neq 0$ and $(a + d) = \pm 2$. The fixed point is given by $\xi_1 = \xi_2 = \xi = (a - d)/(2c)$. The multiplier reduces to $K = 1$. Starting with $w = (az + b)/(cz + d)$, a little algebra leads to the normal form in this case, which is

$$\frac{1}{w - \xi} = \frac{1}{z - \xi} \pm c, \quad \text{depending on whether } a + d = \pm 2.$$

(iii) ***One fixed point at*** ∞**:** This is the case $c = 0,\ a \neq d$. As you know, the Möbius transformation now reduces to the linear transformation $z \mapsto w = (az + b)/d$, with fixed points at $\xi_1 = b/(d - a)$ and $\xi_2 = \infty$. It is easy to see that we now have the normal form

$$w - \xi_1 = K(z - \xi_1) \quad \text{where} \quad K = a/d.$$

(iv) ***Both fixed points at*** ∞**:** Now $c = 0$ and $a = d$. The transformation reduces to a shift of the origin,

$$w = z \pm b, \quad \text{depending on whether } a = d = \pm 1.$$

Again, this is the normal form in the present case. The multiplier K remains equal to unity, of course.

4. Verify that the normal forms in the four cases above are as given above.

Iterates of a Möbius transformation: It is easy to check that the result of making two Möbius transformations in succession is again a Möbius transformation. Hence the effect of an arbitrary number of them performed in succession is also a Möbius transformation. In particular, we may ask for the result of a given transformation, $z \mapsto w = (az + b)/cz + d)$, iterated an arbitrary number of times. Let $z \mapsto z^{(n)}$ after n repeated applications of a Möbius transformation, i.e.,

$$z^{(n)} = \underbrace{f(f(\cdots f}_{n-\text{fold iteration}}(z))\cdots) \quad \text{where} \quad f(z) = \frac{az+b}{cz+d}\,.$$

In its original form, the final expression for $z^{(n)}$ will obviously be quite complicated. But if you use the normal form, you can write down the answer by inspection, in each of the four cases listed above.

(i) ***Finite, distinct fixed points*:** In this case,

$$\frac{z^{(n)} - \xi_1}{z^{(n)} - \xi_2} = K^n \left(\frac{z - \xi_1}{z - \xi_2}\right).$$

(ii) ***Finite, coincident fixed points*:** In this case,

$$\frac{1}{z^{(n)} - \xi} = \frac{1}{z - \xi} \pm n\,c, \quad \text{for } a + d = \pm 2.$$

(iii) ***One fixed point at ∞*:** In this case,

$$z^{(n)} - \xi_1 = K^n\,(z - \xi_1), \quad \text{where} \quad K = a/d.$$

(iv) ***Both fixed points at ∞*:** In this case,

$$z^{(n)} = z \pm n\,b, \quad \text{for } a = d = \pm 1.$$

The relations above are trivially valid for $n = 0$ as well, with $z^{(0)} \equiv z$. Interestingly, they are also valid for *negative* integer values of n, with $z^{(-1)}$ interpreted as the inverse map of z, and $z^{(-n)}$ as the n-fold iterate of the inverse map of z:

$$z^{(-1)} = f^{-1}(z) = \frac{dz - b}{-cz + a}\,, \quad z^{(-n)} = f^{-1}\big(z^{-(n-1)}\big).$$

It is often useful to re-write the formula for $z^{(n)}$ in the general case (i) in the form

$$z^{(n)} = \frac{(\xi_1 - K^n\,\xi_2)\,z + \xi_1\xi_2\,(K^n - 1)}{(1 - K^n)\,z + (K^n\,\xi_1 - \xi_2)}\,.$$

This form shows explicitly that the result of an n-fold iteration of a Möbius transformation is again such a transformation.[11] The relation above is easily inverted to express $z \equiv z^{(0)}$ in terms of $z^{(n)}$. All you have to do is to replace K^n by K^{-n}. It is easy to see that this merely entails an interchange of ξ_1 and ξ_2. Thus,

$$z^{(0)} = \frac{(\xi_2 - K^n\,\xi_1)\,z^{(n)} + \xi_1\xi_2\,(K^n - 1)}{(1 - K^n)\,z^{(n)} + (K^n\,\xi_2 - \xi_1)}\,.$$

Classification of Möbius transformations: We are ready, now, to consider the systematic classification of Möbius transformations. The discussion will be restricted to some introductory remarks on the topic.

For any value of z, the set of points $\{z^{(n)}\,|n \in \mathbb{Z}\}$ is an *orbit* under the map represented by a Möbius transformation. Such an orbit is analogous to the orbit of a point in the phase space of a dynamical system in discrete time, n playing the role of time. The collection of orbits corresponding to different values of $z = z^{(0)}$ (or the initial conditions, in the case of a dynamical system) gives us a *flow* in the complex plane $\hat{\mathbb{C}}$ (analogous to a phase portrait in phase space). The nature of this flow is essentially determined by the fixed points, which are the analogs of the *critical points* or equilibrium points in phase space. This is the starting point of the classification of Möbius transformations.

Let's consider the general case in which $c \neq 0$ (so that the transformation does not reduce to a mere linear transformation), and there are two distinct, finite fixed points. It is convenient to write $z^{(n)}$ in the form

$$z^{(n)} = \left\{\xi_1 - \xi_2\,K^n\left(\frac{z^{(0)} - \xi_1}{z^{(0)} - \xi_2}\right)\right\}\Big/\left\{1 - K^n\left(\frac{z^{(0)} - \xi_1}{z^{(0)} - \xi_2}\right)\right\}.$$

It is now easy to see, at least in most cases, what happens as $n \to \infty$.

(i) If $|K| > 1$, the factor K^n will grow in magnitude with increasing n. Therefore $z^{(n)} \to \xi_2$ as $n \to \infty$, for all initial points $z^{(0)}$ (other than ξ_1, of course). In the spirit of dynamical systems, we may regard ξ_2 as an asymptotically *stable* fixed point, or *attractor*; the flow is generally *toward* this point. On the other hand, ξ_1 acts like an

[11] Incidentally, this expression should suggest to you the possibility of a simple 'formula' for the n^{th} power of a general (2×2) matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with unit determinant. Check this out.

unstable fixed point, or *repellor*; the flow is generally *away* from this point.

(ii) If $|K| < 1$, the factor K^n tends to zero with increasing n. Hence $z^{(n)} \to \xi_1$ as $n \to \infty$, for all initial points $z^{(0)}$ (other than ξ_2). The attractor is now ξ_1 , and the flow is toward this point. The repellor is ξ_2 , and the flow leads away from this point.

(iii) When $|K| = 1$, we have the analog of marginal fixed points, and the flow is neither toward nor away from the fixed points.

(iv) The case when the two fixed points coincide at $\xi = (a-d)/(2c)$ occurs when $(a+d)^2 = 4$, and corresponds to $K = 1$. This is a special case of $|K| = 1$.

These statements can be corroborated as follows. Under the transformation from z to $w = f(z)$, we have $dw = f'(z)dz$. The Jacobian of the transformation is therefore $f'(z)$. Its magnitude $|f'(z)|$ is the local *stretch factor* (or contraction factor). It is this quantity that essentially characterizes the local flow in the neighborhood of any point. (Remember that we are considering the nontrivial case in which $c \neq 0$.) Now,

$$f(z) = \frac{az+b}{cz+d} \quad \Rightarrow \quad |f'(z)| = \frac{1}{|cz+d|^2}\,,$$

on using the fact that $ad - bc = 1$. At the fixed points ξ_1 and ξ_2 , the respective stretch factors become

$$\left.\begin{aligned} |f'(\xi_1)| &= \frac{4}{\left|\, a+d+\sqrt{(a+d)^2-4}\,\right|^2} &&= |K|^2, \\ |f'(\xi_2)| &= \frac{4}{\left|\, a+d-\sqrt{(a+d)^2-4}\,\right|^2} &&= \frac{1}{|K|^2}\,. \end{aligned}\right\}$$

It can be shown that a fixed point is

— unstable if the stretch factor at that point is greater than unity;

— asymptotically stable if it is less than unity; and

— marginal when it is equal to unity.

The statements made above then follow.

★ **5.** Verify the foregoing.

The crucial point is that, in all cases, K depends only on the trace $T \equiv (a+d)$ of the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ made up of the coefficients of the transformation. The multiplier is given in terms of T by

$$K = \frac{T - \sqrt{T^2-4}}{T + \sqrt{T^2-4}}\,.$$

This is the starting point for the classification of Möbius transformations into different types. Four types of transformations are possible, depending solely on the value of T. I merely list them here, without going into further detail. The first three types correspond to real values of the trace T.

Type 1: Elliptic transformation, when T is real and $-2 < T < 2$. Then $|K| = 1$, but $K \neq 1$.

Type 2: Parabolic transformation, when $T = \pm 2$. The two fixed points now coincide, and we have $K = 1$.

Type 3: Hyperbolic transformation, when T is real and $|T| > 2$. In this case K is a positive number other than unity (i.e., $K \neq 1$).

Type 4: Loxodromic transformation, when T is any complex number such that $T \notin [-2, 2]$. A *loxodrome* is a curve drawn on the surface of a sphere that intersects all lines of longitude at a given, constant angle ψ. It is obvious that $\psi = 0$ and $\psi = \frac{1}{2}\pi$ correspond, respectively, to a line of longitude and a latitude, respectively. For intermediate values of ψ, the loxodrome is a spiral. As you might expect, loxodromes first arose in the context of navigation on the high seas. Sailing on a path that makes a constant angle with the northern direction takes you on a loxodrome. Hyperbolic transformations are a special case of this type, corresponding to real values of T.

The isometric circle: As stated already, the local stretch or contraction factor associated with the mapping of an infinitesimal neighborhood of any point z by the Möbius transformation $z \mapsto (az + b)/(cz + d)$ is $|f'(z)| = 1/|cz + d|^2$. It follows that there is *no* distortion at all in the map of the circle given by

$$|cz + d| = 1, \quad \text{or} \quad |z + (d/c)| = 1/|c|.$$

(Recall, once again, that we are considering the case $c \neq 0$.) This circle with center at $-d/c$ and radius equal to $1/|c|$ is called the *isometric circle* corresponding to the transformation concerned. It plays an important role in the theory of Möbius transformations.

Points inside the isometric circle satisfy $|cz + d| < 1$, and hence infinitesimal area elements inside the circle are *magnified* by the mapping, by a factor $|cz + d|^{-4}$. Similarly, points outside the isometric circle satisfy $|cz + d| > 1$. Hence infinitesimal area elements in this region are *shrunk* by the transformation, by a factor $|cz + d|^{-4}$.

6. Consider the Möbius transformation $z \mapsto w = (az + b)/(cz + d)$ (where $c \neq 0$).

(a) Show that the transformation maps its isometric circle to the isometric circle of the inverse transformation $w \mapsto z = (dw - b)/(-cw + a)$. Further, the interior

(respectively, exterior) of the isometric circle in the z-plane is mapped to the exterior (respectively, interior) of its image in the w-plane. (As you know, the center $-d/c$ of the isometric circle is mapped to ∞, while ∞ is mapped to the center a/c of the image of the isometric circle under the mapping.)

(b) Consider the case of finite, distinct fixed points ξ_1 and ξ_2, and (i) $|K| > 1$ and (ii) $|K| < 1$, respectively. What is the isometric circle of the n^{th} iterate of the transformation? What happens as $n \to \infty$?

Group properties; the Möbius group: An important feature of Möbius transformations or maps $f : z \mapsto w$ is the fact that they form a *group*, the so-called **Möbius group** $\text{Möb}\,(2, \mathbb{C})$. A few salient properties of this group will be discussed here.

Möbius transformations are easily seen to satisfy the axioms that define a group:

(a) The successive application of two Möbius transformations is another Möbius transformation. More formally, if f and g are maps, so is their composition $f \circ g$.

(b) The composition of transformations is associative. If f, g and h are maps, then $f \circ (g \circ h) = (f \circ g) \circ h$.

c) There is an identity transformation under which every z is mapped to itself.

(d) Each transformation $f : z \mapsto w$ has an inverse $f^{-1} : w \mapsto z$.[12]

Here's how the transformation obtained by composing two successive transformations can be read off easily:

- The transformation $z \mapsto w = (az + b)/(cz + d)$ can be associated with the matrix of coefficients, $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$.
- The composition of transformations then corresponds to the product of the associated matrices (in the right order).

Thus, if

$$z \mapsto z' = \frac{a_1 z + b_1}{c_1 z + d_1} \quad \text{is followed by} \quad z' \mapsto z'' = \frac{a_2 z + b_2}{c_2 z' + d_2}\,,$$

then

$$z'' = \frac{(a_2 a_1 + b_2 c_1) z + a_2 b_1 + b_2 d_1}{(c_2 a_1 + d_2 c_1) z + c_2 b_1 + d_2 d_1}\,.$$

But the set of all nonsingular (2×2) matrices with complex elements constitutes a group. The group composition law is just matrix multiplication. The group is called the general linear group over the complex numbers in 2 dimensions, and is denoted by $GL(2, \mathbb{C})$. As you know, however, the coefficients of a Möbius transformation can

[12] That is, if $w = f(z) = (az + b)/(cz + d)$, then $z = f^{-1}(w) = (dw - b)/(-cw + a)$.

always be chosen such that $ad - bc = 1$. What is relevant here, therefore, is the set of (2×2) matrices with complex elements and determinant equal to unity. These matrices form a subgroup of $GL(2,\mathbb{C})$: the unimodular or *special* linear group over the complex numbers, denoted by $SL(2,\mathbb{C})$. We might therefore expect Möb $(2,\mathbb{C})$ to be essentially the same as $SL(2,\mathbb{C})$. But there is one more point to be taken into account.

Changing the sign of all the four coefficients in a Möbius transformation does not alter it, because $(az+b)/(cz+d) \equiv (-az-b)/(-cz-d)$. Hence, to each Möbius transformation

$$z \mapsto w = \frac{az+b}{cz+d} \quad \text{with} \quad ad-bc=1,$$

there correspond the *two* matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ and } \begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix} \in SL(2,\mathbb{C}).$$

The identity transformation $z \mapsto z$ thus corresponds to the (2×2) identity matrix I as well as its negative, $-I$. Hence:

- There is a two-to-one correspondence, or *homomorphism*, from the group $SL(2,\mathbb{C})$ to the group Möb $(2,\mathbb{C})$.
- The matrices I and $-I$ form the *kernel* of the homomorphism. That is, $\{I,-I\}$ is the set of matrices in $SL(2,\mathbb{C})$ whose image in Möb $(2,\mathbb{C})$ is the identity transformation.

But the two matrices I and $-I$ themselves form a group, under matrix multiplication. This group is just the cyclic group of order 2, denoted by $\mathbb{Z}_2$. It is the same as the group formed by the integers 0 and 1 under binary addition, i.e., addition *modulo* 2. One now forms the *quotient group* $SL(2,\mathbb{C})/\mathbb{Z}_2$. This is read as '$SL(2,\mathbb{C})$ modulo $\mathbb{Z}_2$', and refers, in effect, to the group of unimodular 2×2 matrices *up to the overall sign* of the matrices. In other words, it is a group of these matrices, such that any two matrices differing only in sign are *identified* with each other and regarded as a single element of the group. It is this group, also known as the *projective* special linear group $PSL(2,\mathbb{C})$, with which the group of Möbius transformations is in one-to-one correspondence.

- Möb $(2,\mathbb{C})$ is *isomorphic* to $SL(2,\mathbb{C})/\mathbb{Z}_2$. This isomorphism is written as

$$\text{Möb}\,(2,\mathbb{C}) \cong SL(2,\mathbb{C})/\mathbb{Z}_2 \equiv PSL(2,\mathbb{C}).$$

- The group $SL(2,\mathbb{C})$ is the so-called **universal covering group** of Möb $(2,\mathbb{C})$. The latter is a subgroup of $SL(2,\mathbb{C})$.

Finally, I mention the following remarkable fact that connects Möbius transformations to special relativity. The special linear group $SL(2,\mathbb{C})$ is also the so-called universal

covering group of the group of homogeneous, proper, orthochronous Lorentz transformations, $SO(3,1)$, in the usual spacetime comprising three spatial dimensions and one time dimension. There is again a 2-to-1 homomorphism between $SL(2,\mathbb{C})$ and $SO(3,1)$, and we have

$$\text{Möb}\,(2,\mathbb{C}) \cong SL(2,\mathbb{C})/\mathbb{Z}_2 \cong SO(3,1).$$

That is, the group of Möbius transformations is isomorphic to the Lorentz group!

The Möbius group over the reals: We've seen that the Möbius group $\text{Möb}\,(2,\mathbb{C})$ is isomorphic to certain other important groups such as the projective linear group $PSL(2,\mathbb{C})$ and the homogeneous Lorentz group $SO(3,1)$. Further, it is a subgroup of the special linear group $SL(2,\mathbb{C})$, and hence of the general linear group $GL(2,\mathbb{C})$. In turn, the Möbius group itself has some important and interesting subgroups.

Möbius transformations with real parameters, in which the coefficients a, b, c and d are restricted to *real* numbers, comprise a group on their own, $\text{Möb}\,(2,\mathbb{R})$. This group is isomorphic to the projective linear group $PSL(2,\mathbb{R})$ over the reals, a subgroup of $PSL(2,\mathbb{C})$. It is also a quotient group, being the special linear group $SL(2,\mathbb{R})$ modulo $\mathbb{Z}_2$: we have

$$\text{Möb}\,(2,\mathbb{R}) \cong SL(2,\mathbb{R})/\mathbb{Z}_2 \equiv PSL(2,\mathbb{R}).$$

Möbius transformations with real parameters have some additional properties that are of importance. It is easy to show that

- a real Möbius transformation maps the upper (respectively, lower) half of the complex plane to the upper (respectively, lower) half-plane.

7. Verify the foregoing statement. *Hint*: Consider the Möbius transformation $z \mapsto w = (az+b)/(cz+d)$ where a, b, c and d are real numbers, and $ad - bc = 1$. Let $z = x + iy$ and $w = u + iv$. Show that $v = y/|cz+d|^2$, so that $\text{Im}\, w \gtrless 0$ according as $\text{Im}\, z \gtrless 0$.

The modular group: Among Möbius transformations with real parameters (a, b, c, d) where $ad - bc = 1$, those with *integer* values of the parameters form a group by themselves! This is the group $PSL(2,\mathbb{Z})$, called the **modular group** Similarly, the set of Möbius transformations in which (a, b, c, d) are *Gaussian integers*, i.e., each parameter is of the form $m + ni$ where m and n are integers, also forms a group. The modular group has numerous remarkable properties, and it plays a role in many areas of mathematics including hyperbolic geometry, number theory, elliptic functions, etc. It also appears in certain topics in theoretical physics such as conformal field theory.

The invariance group of the unit circle: Here is another important example of a continuous subgroup of the Möbius group. Consider all Möbius transformations that

leave the *unit circle* unchanged, i.e., those that map the circle $|z| = 1$ in the z-plane to the circle $|w| = 1$ in the w-plane. These transformations are either of the form

$$z \mapsto w = \frac{az + b}{b^* z + a^*}, \quad \text{where} \quad \det \begin{pmatrix} a & b \\ b^* & a^* \end{pmatrix} = |a|^2 - |b|^2 = 1,$$

or of the form

$$z \mapsto w = \frac{az + b}{-b^* z - a^*}, \quad \text{where} \quad \det \begin{pmatrix} a & b \\ -b^* & -a^* \end{pmatrix} = -|a|^2 + |b|^2 = 1.$$

Let's call these type (i) and type (ii) transformations, for convenience. They form a subgroup of the group Möb $(2, \mathbb{C})$ of Möbius transformations. Transformations of types (i) and (ii) are distinguished by the following property:

— A type (i) transformation maps the interior of the unit circle in the z-plane to its interior in the w-plane; and the exterior of the unit circle in the z-plane to the exterior of the unit circle in the w-plane.

— A type (ii) transformation maps the interior of the unit circle in the z-plane to its exterior in the w-plane, and vice versa.

8. Show that

(a) the general form of a Möbius transformation $z \mapsto w$ that maps the unit circle to the unit circle is as given above;

(b) all such transformations form a group;

(c) $|z| \lesseqgtr 1 \;\Rightarrow\; |w| \lesseqgtr 1$ for type (i) transformations, and $|z| \lesseqgtr 1 \;\Rightarrow\; |w| \gtreqless 1$ for type (ii) transformations.

Connection with the pseudo-unitary group $SU(1,1)$**:** The matrices corresponding to Möbius transformations of types (i) and (ii) above are, respectively, of the general form

$$U_+ = \begin{pmatrix} a & b \\ b^* & a^* \end{pmatrix} \quad \text{where} \quad \det U_+ = |a|^2 - |b|^2 = 1,$$

and

$$U_- = \begin{pmatrix} a & b \\ -b^* & -a^* \end{pmatrix} \quad \text{where} \quad \det U_- = -|a|^2 + |b|^2 = 1.$$

But these are precisely the general forms of the matrices belonging to the **indefinite unitary group** (or pseudo-unitary group) $SU(1,1)$. This is the group of (2×2) matrices with complex entries satisfying the following conditions: if $U \in SU(1,1)$, then

$$\det U = +1 \;\text{ and }\; U^\dagger \, g \, U = g, \;\text{ where }\; g = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Now, $SU(1,1)$ turns out to be isomorphic to the special linear group $SL(2,\mathbb{R})$ of unimodular (2×2) matrices over the real numbers. Further, the group $SL(2,\mathbb{R})$ is also isomorphic to the symplectic group $Sp(2,\mathbb{R})$, which is the group of canonical transformations of a Hamiltonian system with one degree of freedom! We have

$$SU(1,1) \cong SL(2,\mathbb{R}) \cong Sp(2,\mathbb{R}),$$

and the group of Möbius transformations that leaves the unit circle invariant is therefore isomorphic to $SL(2,\mathbb{R})/\mathbb{Z}_2$, namely, Möb $(2,\mathbb{R})$ once again.

The group of cross-ratios: An interesting *discrete* subgroup of Möb $(2,\mathbb{C})$ is made up of the following set of six Möbius transformations, which I will denote by e_i where $i = 1, 2, \ldots, 6$:

$$\left.\begin{array}{lll} e_1 : & z \mapsto z & \text{(the identity transformation)} \\ e_2 : & z \mapsto 1/z & \text{(inversion)} \\ e_3 : & z \mapsto 1-z & \text{(rotation about the point } \tfrac{1}{2} \text{ through an angle } \pi) \\ e_4 : & z \mapsto 1-(1/z) & \text{(inversion followed by rotation)} \\ e_5 : & z \mapsto 1/(1-z) & \text{(rotation followed by inversion)} \\ e_6 : & z \mapsto z/(z-1) & \text{(inversion-rotation-inversion).} \end{array}\right\}$$

The name 'group of cross-ratios' arises from the fact that these transformations are precisely the ones leading to the identities (already listed) satisfied by the cross-ratio of any four distinct points in $\hat{\mathbb{C}}$.

The transformation e_3 can also be regarded as successive reflections about the line $\mathrm{Re}\, z = \frac{1}{2}$ and about the real axis, performed in either order. Let $e_i\, e_j$ stand for the transformation e_j followed by the transformation e_i. It is obvious that $e_2\, e_2 = e_1 = e_3\, e_3$, so that $e_2^{-1} = e_2$ and $e_3^{-1} = e_3$. Further, $e_4 \equiv e_3\, e_2$ and $e_5 \equiv e_2\, e_3$. The last element $e_6 = e_2\, e_3\, e_2$ (inversion-rotation-inversion) or, alternatively, $e_6 = e_3\, e_2\, e_3$ (rotation-inversion-rotation).

9. Verify that the values of the parameters a, b, c, d corresponding to the foregoing transformations (such that $ad - bc = 1$ in each case), and the 'multiplication table' for the group, are as follows:

	a	b	c	d
e_1	1	0	0	1
e_2	0	$-i$	$-i$	0
e_3	i	$-i$	0	$-i$
e_4	1	-1	1	0
e_5	0	1	-1	1
e_6	i	0	i	$-i$

	e_1	e_2	e_3	e_4	e_5	e_6
e_1	e_1	e_2	e_3	e_4	e_5	e_6
e_2	e_2	e_1	e_5	e_6	e_2	e_4
e_3	e_3	e_4	e_1	e_2	e_6	e_5
e_4	e_4	e_3	e_6	e_5	e_1	e_2
e_5	e_5	e_6	e_2	e_1	e_4	e_3
e_6	e_6	e_5	e_4	e_3	e_2	e_1

Note that this group is *not* an abelian or commutative group, since $e_i\, e_j \neq e_j\, e_i$ in general. Several identities follow at once from the multiplication table above: for instance, $(e_4)^3$, $(e_5)^3$, $(e_6)^2$, $(e_3\, e_4)^2$, $(e_4\, e_3)^2$, $(e_3\, e_5)^2$, $(e_5\, e_3)^2$, ... are all equal to the identity transformation.

6 Multivalued functions; integral representations

Branch points and branch cuts: An analytic function $f(z) = w$ should be regarded as a map $f : z \mapsto w$ of some region $\mathcal{R}$ of the extended complex plane $\hat{\mathbb{C}}$ to some region $\mathcal{R}'$ of $\hat{\mathbb{C}}$. If $f(z)$ is single-valued, this map yields a unique value of w for every value of z. All the functions we have considered so far satisfy this property. In general, however, we have to deal with *multivalued* functions.

A simple example is provided by the function $f(z) = z^{1/2}$. It is obvious that the whole of the complex z-plane is mapped to just the upper half-plane in w, because $z = r\,e^{i\theta}$ maps to $w = r^{1/2}\,e^{i\theta/2}$. In order to cover the whole of the w-plane, the z-plane has to be covered *twice*: the argument θ of z has to run from 0 to 4π, rather than just 2π. As θ increases from 0 to 2π, we obtain the branch $+\sqrt{z}$ of the function $f(z) = z^{1/2}$. As it increases further from 2π to 4π, we obtain the branch $e^{i\pi}\,z^{1/2} = -\sqrt{z}$ of the square root function.

In order to keep track of the correct branch, we therefore need two copies of the z-plane. The two copies may be imagined to be two sheets, one lying below the other, such that we descend to the second sheet by starting just above the positive real axis on the first sheet and traversing a path that encircles the origin once in the positive sense. On encircling 0 once again, this time on the second sheet, we ascend back to the first sheet. The two sheets (labelled I and II, say) are called **Riemann sheets**, and they are supposed to be connected to each other along a slit in each sheet running from 0 to ∞ on the positive real axis. The top sheet, on which the phase of z runs from 0 to 2π, is called the **principal sheet** of the function concerned. The points where the two branches of the two-valued function $z^{1/2}$ coincide in value, namely, $z = 0$ and $z = \infty$ (recall that there is only one point at infinity in $\hat{\mathbb{C}}$), are called **branch points**. They are connected by a **branch cut**. The two sheets pasted together as described above form the **Riemann surface** of the function $z^{1/2}$.

The branch cut joining $z = 0$ and $z = \infty$ may be taken to run along any curve running between these two points. It is most convenient to take it to run along a straight line. We have chosen the positive real axis in the foregoing. But this need not always be so. All that is needed is a specification of the *phase* (or *argument*) of the function concerned just above and just below the branch cut, so that we can calculate the **discontinuity** (or the jump in the value) of the function across the branch cut. Label the function on the sheets I and II as $f_{\rm I}(z)$ and $f_{\rm II}(z)$, respectively. By continuity, the value of the function on sheet I, as you approach the positive real axis from below, is the same as the value of the function on sheet II, as the real axis is approached from above. That is,

$$\lim_{\epsilon\to 0}\ f_{\rm I}(x - i\epsilon) = \lim_{\epsilon\to 0}\ f_{\rm II}(x + i\epsilon), \quad x > 0.$$

The discontinuity across the cut is then easily determined. In the present instance, it

is given by

$$\begin{aligned}\text{disc } f(z)\Big|_{x>0} &= \lim_{\epsilon\to 0}\left[f_{\rm I}(x+i\epsilon) - f_{\rm I}(x-i\epsilon)\right] \\ &= \lim_{\epsilon\to 0}\left[f_{\rm I}(x+i\epsilon) - f_{\rm II}(x+i\epsilon)\right] \\ &= \sqrt{x} - (-\sqrt{x}) = 2\sqrt{x}.\end{aligned}$$

***Algebraic branch point*:** The branch points of $z^{1/2}$ at $z = 0$ and ∞ are *algebraic* branch points. So are those of the function $z^{1/3}$, for instance. In this case the Riemann surface has three sheets. More generally, the function $z^{p/q}$, where p and q are integers with no common factors, has algebraic branch points at $z = 0$ and $z = \infty$. Its Riemann surface comprises q sheets. Each sheet descends smoothly to the one below it as we cross the branch cut; crossing the cut on the lowest sheet brings us back to topmost sheet.

***Winding point*:** The function z^α, where α is not a rational real number, also has branch points at $z = 0$ and $z = \infty$. These are called *winding points*. The Riemann surface of this function has an infinite number of sheets, because $e^{2\pi n i\alpha}$ is never equal to unity for any integer value of n. These sheets are labelled by an integer $n \in \mathbb{Z}$. On the principal sheet, the phase of z runs from 0 to 2π, as usual, and $n = 0$.

***Logarithmic branch point*:** The function $\ln z$ has *logarithmic branch points* at $z = 0$ and $z = \infty$. The Riemann surface is again infinite-sheeted, the sheets being numbered by the full set of integers. On the principal sheet, $\ln z = \ln r + i\theta$, where $0 \le \theta < 2\pi$. On the $n^{\rm th}$ sheet, $\ln z = \ln r + i\theta + 2\pi n i$, where $n \in \mathbb{Z}$ and $0 \le \theta < 2\pi$.

Branch cuts run from one branch point to another. No function can have just one branch point; the smallest number of branch points that a function can have is two. When a branch point is encircled in the z-plane, the function does *not* return to its original value. In order to obtain a *closed* contour that encircles a branch point, the contour must encircle the branch point (or other branch points) as many times as is necessary to return the function to its original value at the starting point.

Consider the functions $(z-a)^{1/2}\,(z-b)^{1/2}$ and $(z-a)^{1/2}/(z-b)^{1/2}$, where a and b are any two finite complex numbers. They have algebraic (square-root) branch points at $z = a$ and $z = b$, but their behavior at $z = \infty$ is regular. Therefore their branch cuts can be chosen to run over the *finite* segments from a to b.

For any arbitrary non-integer value of α, including complex values, the cuts of the function $(z-a)^\alpha/(z-b)^\alpha$ can also be chosen to run over the finite segment from $z = a$ to $z = b$ alone. But when α is not a half-odd-integer, this is not possible for the product function $(z-a)^\alpha\,(z-b)^\alpha$. This function has branch points at $z = a$, $z = b$, *as well as* $z = \infty$, for all α that is not an integer or half-odd-integer. The branch cut structure

of this function must necessarily run up to ∞.

The singularity structure of multivalued functions can be quite intricate and interesting. Here's a simple example. The function

$$f(z) = \frac{1}{z} \ln(1 - z)$$

has logarithmic branch points at $z = 1$ and $z = \infty$. On the *principal* sheet of the logarithm, we have $\ln 1 = 0$; hence $\ln(1 - z) \simeq -z$ in the infinitesimal neighborhood of the origin on this sheet. The simple pole at $z = 0$ of the factor z^{-1} is therefore 'cancelled' by the simple zero of the logarithm, and $f(z)$ only has a removable singularity at $z = 0$. On *every other* sheet of the logarithm, however, $f(z)$ does have a simple pole at $z = 0$, because $\ln 1 = 2\pi ni \neq 0$ on all these sheets.

Contour integrals in the presence of branch points: As you have seen, the evaluation of integrals via contour integration relies, ultimately, on integrating analytic functions over *closed* contours. In the case of multivalued functions, however, if the contour starts at some point z and encircles a branch point before returning to the starting point, the function does not return to its original value. Hence the contour is not really a closed one—the final point is actually on some other Riemann sheet of the function!

In order to apply the theorems pertaining to integrals over closed contours, you must ensure that the function has returned to its starting value (or, equivalently, that z has returned to the starting point on the Riemann surface of the function). In general, this involves encircling more than one branch point, or the same branch point more than once, and so on, as will become clear from the examples that follow.

1. Let a and b be arbitrary real numbers, where $a < b$. Use contour integration to show that the integral

$$I = \int_a^b \frac{dx}{\sqrt{(b - x)(x - a)}} = 2\pi,$$

independent of a and b.

Hint: Consider the function $f(z) = (z - a)^{-1/2} (z - b)^{-1/2}$. Note that the behavior of the function at ∞ is regular. Choose the branch cut to run from a to b, and write down the phases of the function above and below the cut. Relate the line integral I to a contour integral surrounding the branch cut. Now open out the contour to make it a large circle whose radius tends to infinity, and pick up the contribution from the circle to arrive at the result. This method makes it evident why the value of the integral is independent of the actual values of a and b.

2. Use a similar method to show that

$$\int_C \frac{dz}{\sqrt{1+z+z^2}} = 2\pi i,$$

where C is the circle $|z| = 2$ traversed once in the positive sense.

Hint: The branch points of the integrand are at two of the cube roots of unity, given by $z = \omega = e^{2\pi i/3}$ and $z = \omega^2 = e^{4\pi i/3}$. The branch cut of the integrand may be taken to run between these points. As $|z| \to \infty$ along any direction, the integrand tends to $1/z$.

An integral involving a class of rational functions: Consider the definite integral of a rational function of the form

$$I = \int_0^\infty dx\, \frac{p(x)}{q(x)}\,,$$

where $p(x)$ and $q(x)$ are polynomials in x satisfying two conditions:

(i) The degree of $q(x)$ exceeds that of $p(x)$ by at least 2. Hence the integrand decays at least as rapidly as $1/x^2$ as $x \to \infty$, and the convergence of the integral is guaranteed.

(ii) $q(x)$ does not have any zeroes for $x \geq 0$. Hence there is no non-integrable singularity on the path of integration, and the integral exists.

As you know, such an integral can be evaluated by elementary means—for instance, by resolving the integrand into partial fractions.[13] But here is how you can evaluate I by contour integration, using a simple trick. Consider, instead of I, the contour integral

$$I' = -\frac{1}{2\pi i}\int_C dz\, \frac{p(z)\,\ln z}{q(z)}\,,$$

where C is the familiar hairpin contour that comes in from infinity just below the positive real axis, encircles zero from the left, and runs just above the positive real axis. The branch cut of $\ln z$ is taken to run from 0 to ∞ along the positive real axis in the z-plane. Since $\ln z = \ln x$ just above the cut, and $\ln z = \ln x + 2\pi i$ just below the cut, we have

$$I' = -\frac{1}{2\pi i}\int_\infty^0 dx\, \frac{p(x)\,(\ln x + 2\pi i)}{q(x)} - \frac{1}{2\pi i}\int_0^\infty dx\, \frac{p(x)\,\ln x}{q(x)} = I.$$

[13] Obviously, you must then take care to avoid the appearance of any spurious logarithmic singularity owing to the upper limit of integration. Evaluate the integral up to some upper limit L, and take the limit $L \to \infty$ after the different terms are recombined properly.

On the other hand, we can evaluate I' by completing the contour C with the addition of a large circle (that does not cross the cut). The contribution of the large circle vanishes as its radius R of the circle tends to infinity, because the integrand vanishes at least as fast as $(\ln R)/R$ on this circle. But the closed contour can now be *shrunk* to pick up the residues of all the poles of the integrand (located at the zeroes of $q(z)$). The integral I is thus evaluated quite easily.

3. Use the method just described to evaluate the following standard integrals:

(a) $$\int_0^\infty \frac{dx}{x^n+1} = \frac{\pi}{n}\operatorname{cosec}\frac{\pi}{n} \quad (n = 2, 3, \dots).$$

(b) $$\int_0^\infty \frac{dx}{x^{n-1}+x^{n-2}+\cdots+x+1} = \frac{\pi}{n}\operatorname{cosec}\frac{2\pi}{n} \quad (n = 3, 4, \dots).$$

Hint: (a) The poles of the integrand that are enclosed by C are at the roots of $z^n + 1 = 0$, namely, at $z = e^{i\pi/n}\,\omega^r$, where $\omega = e^{2\pi i/n}$ and $r = 0,\, 1,\, \dots\,,\, n-1$.

(b) The poles of the integrand that are enclosed by C are at the roots of unity other than 1 itself, i.e., at $z = \omega^r$, where $r = 1,\, \dots\,,\, n-1$.

Contour integral representations for various special functions are easily constructed using the properties of multivalued functions. Such representations often provide 'master representations' or analytic continuations of the functions concerned that are valid for all complex values of the arguments of these functions. Let's consider some examples.

The gamma function: Recall the defining representation of this function, namely,

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

We've seen that the functional equation $z\,\Gamma(z) = \Gamma(z+1)$ helps us extend the region of analyticity to the left of the region $\operatorname{Re} z > 0$, stripwise. But we are now in a position to write down a single representation for the gamma function that is valid *throughout* the complex plane.

Consider the integrand $t^{z-1}\, e^{-t}$ in the defining representation as an analytic function of t, for any arbitrary complex value of z. The factor t^{z-1} has a branch points at $t = 0$ and $t = \infty$, with a cut running between them. Take the branch cut to run along the positive t-axis. The phase of t^{z-1} just above the cut is 0, while it is $2\pi i z$ just below the cut (the factor $e^{-2\pi i}$ is just unity). Consider the integral $\int_C dt\, t^{z-1}\, e^{-t}$, where the contour C comes in from ∞ to ϵ just below the cut, encircles the origin in the negative sense in an arc of a circle of radius ϵ, and runs just above the cut from ϵ to ∞. As long as $\operatorname{Re} z > 0$, the contribution from the arc of the small circle vanishes as $\epsilon \to 0$.

Moreover, the contributions from the two line segments equal $-e^{2\pi iz}\,\Gamma(z)$ and $\Gamma(z)$, respectively. Hence, in the region $\mathrm{Re}\,z > 0$, we have

$$\Gamma(z) = \frac{1}{(1 - e^{2\pi iz})} \int_C dt\, t^{z-1}\, e^{-t}.$$

But C does not pass through the point $t = 0$, and may be deformed to stay clear of the origin, in the form of a *hairpin-shaped* contour straddling the branch cut of t^{z-1}. The contour integral is thus defined for all finite values of z. On the other hand, the factor $(1 - e^{2\pi iz})^{-1}$ has simple poles at all integer values of z. The product of these two factors is a meromorphic function of z. It represents $\Gamma(z)$ for *all* z. Note that the hairpin contour C can be partially 'straightened out', but we must always ensure that $\mathrm{Re}\,t \to +\infty$ asymptotically at both ends of the open contour, so that the damping factor e^{-t} ensures the convergence of the integral.

4. It can be checked that the known analytic properties of $\Gamma(z)$ follow from the integral representation derived above.

(a) Owing to the factor $1/(1 - e^{2\pi iz})$, it might appear that $\Gamma(z)$ has a simple pole at *every* integer value of z. But we know that $\Gamma(n)$ is finite $(= (n-1)!)$ when n is a positive integer. What happens is that the contour integral $\int_C dt \cdots$ *also* vanishes when $z = 1, 2, \ldots$: the integrand becomes single-valued because t^{n-1} does not have any branch points, and the two straight segments of the contour cancel each other out. Verify that, as z tends to any positive integer n, the limiting value of the right-hand side is precisely $(n-1)!$.

(b) On the other hand, when $z = -n$ where $n = 0, -1, -2, \ldots$, the branch cut disappears, but a *pole* of order $(n+1)$ is left at the origin in the t-plane. The contour integral is then evaluated easily. Verify that $\Gamma(z)$ has a simple pole at $z = -n$ with residue equal to $(-1)^n/n!$.

The beta function: Recall the original definition of the beta function,

$$B(z,w) = \int_0^1 dt\, t^{z-1}\,(1-t)^{w-1}, \quad \text{where } \mathrm{Re}\,z > 0,\ \mathrm{Re}\,w > 0.$$

You have also seen that, in this case, it is not possible to extend the region of analyticity in z and w *simultaneously* by using integration by parts.

Once again, we can continue $B(z,w)$ analytically to all values of z and w by suitably exploiting the discontinuity of the integrand $t^{z-1}\,(1-t)^{w-1}$ across the branch cut running between the branch points at $t = 0$ and $t = 1$. For general values of z and w, there is a branch point at $t = \infty$ as well. The branch cut structure of the integrand is therefore more complicated in this case, because the cut actually runs all the way to infinity. *But all that we need to keep track of is the following fact*: starting on the

principal sheet, the integrand acquires a factor $e^{2\pi iz}$ (respectively, $e^{-2\pi iz}$) whenever the contour encircles the branch point at $t = 0$ in the positive (respectively, negative) sense. Similarly, a factor $e^{2\pi iw}$ or $e^{-2\pi iw}$ results when the branch point at $t = 1$ is encircled in the positive or negative sense.

What is required in order to return to the original value of the function, and hence to close the contour, is a *double* encircling of each of the branch points $t = 0$ and $t = 1$, once in the positive sense and once in the negative sense. The outcome is a *Pochhammer contour* that is written symbolically as $C : (1-, 0-, 1+, 0+)$.[14] The contributions from the infinitesimal circles around 0 and 1 vanish as long as $\mathrm{Re}\, z > 0$ and $\mathrm{Re}\, w > 0$. Therefore

$$B(z,w) = \int_0^1 dt\, t^{z-1}\,(1-t)^{w-1} = \frac{1}{(1-e^{-2\pi iz})(1-e^{-2\pi iw})}\int_C dt\, t^{z-1}\,(1-t)^{w-1}.$$

But the contour of integration C does not pass through the singularities of the integrand at $t = 0$ and $t = 1$. The contour can be distorted away from these points without changing the value of the integral. The contour integral thus provides an analytic continuation of the beta function for *all* finite values of z and w:

$$B(z,w) = \frac{1}{(1-e^{-2\pi iz})(1-e^{-2\pi iw})}\int_C dt\, t^{z-1}\,(1-t)^{w-1}.$$

The Riemann zeta function: Recall that the Riemann zeta function is defined as

$$\zeta(z) = \sum_{n=1}^{\infty}\frac{1}{n^z} \quad (\mathrm{Re}\, z > 1).$$

Once again, an analytic continuation of $\zeta(z)$ to the whole of the complex z-plane is easily derived, in terms of a contour integral. Consider the product

$$\zeta(z)\,\Gamma(z) = \sum_{n=1}^{\infty}\frac{1}{n^z}\int_0^{\infty} du\, u^{z-1}\, e^{-u}, \quad \text{where} \quad \mathrm{Re}\, z > 1.$$

Change variables of integration by setting $u = nt$. Then

$$\begin{aligned}\zeta(z)\,\Gamma(z) &= \sum_{n=1}^{\infty}\int_0^{\infty} dt\, t^{z-1}\, e^{-nt} = \int_0^{\infty} dt\, t^{z-1}\sum_{n=1}^{\infty} e^{-nt} \\ &= \int_0^{\infty} dt\,\frac{t^{z-1}}{(e^t-1)}, \ \mathrm{Re}\, z > 1. \qquad (1)\end{aligned}$$

[14] The term Pochhammer contour is used for the class of such contours that encircle several branch points, each in a specific (positive or negative) sense.

The divergence of the *sum* $\sum_{n=1}^{\infty} n^{-z}$ that occurs when $\mathrm{Re}\, z = 1$ now becomes a divergence of the *integral*: near $t = 0$, the integrand behaves like t^{z-2}, an extra factor of t^{-1} coming from the denominator $(e^t - 1)^{-1}$. Hence the integral diverges unless $\mathrm{Re}\, z > 1$.

It is obvious that we cannot achieve convergence of the integral to the left of $\mathrm{Re}\, z = 1$ by integration by parts, in this case. While $t^z \to t^{z+1}$ upon integration, $(e^t - 1)^{-1} \to (e^t - 1)^{-2}$ upon differentiation. The behavior of the integrand near $t = 0$ does not improve. But we may exploit (once again) the branch cut of t^{z-1} running from $t = 0$ to $t = \infty$ to convert the integral to a contour integral over a hairpin contour C straddling the branch cut, encircling the origin in the negative sense. The contour is the same as the one used for the gamma function. This gives

$$\zeta(z) = \frac{1}{\Gamma(z)\,(1 - e^{2\pi i z})} \int_C dt\, \frac{t^{z-1}}{(e^t - 1)}\,.$$

The path of integration no longer passes through $t = 0$. This integral representation is therefore valid for all z. Using the reflection formula $\Gamma(z)\,\Gamma(1 - z) = \pi\,\mathrm{cosec}\,(\pi z)$, it is convenient to re-write it as

$$\zeta(z) = \frac{i}{2\pi}\, e^{-i\pi z}\, \Gamma(1 - z) \int_C dt\, \frac{t^{z-1}}{(e^t - 1)}\,.$$

A number of interesting properties of the zeta function can now be deduced easily.

5. The contour integral in the formula above is an entire function of z. It follows that the only possible singularities of $\zeta(z)$ in the finite part of the complex plane must come from the poles of the factor $\Gamma(1 - z)$ at $z = 1,\, 2,\, \ldots$.

(a) Show that $\zeta(z)$ has a simple pole at $z = 1$, with residue equal to 1.

(b) Show that $\zeta(z)$ has *no* singularities at $z = n$ where $n = 2,\, 3,\, \ldots$.

Hint: (a) At $z = 1$, the contour integral reduces to $\oint dt/t = -2\pi i$.

(b) The contour integral is an analytic function of z in the neighborhood of $z = n$ (where $n = 2,\, 3,\, \ldots$), and vanishes at those points. To find the coefficient of $(z - n)$, use the Taylor series $t^{z-1} = t^{n-1} + (z - n)\, t^{n-1} \ln t + \cdots$. The limiting value of $\zeta(z)$ as $z \to n$ then works out to be

$$\zeta(n) = \frac{i}{2\pi(n-1)!} \int_C dt\, \frac{t^{n-1} \ln t}{(e^t - 1)}\,, \quad n = 2, 3, \ldots\,.$$

Since the discontinuity of $\ln t$ across the cut on the positive real axis is $2\pi i$, this further simplifies to

$$\zeta(n) = \frac{1}{(n-1)!} \int_0^{\infty} dt\, \frac{t^{n-1}}{(e^t - 1)}\,, \quad n = 2, 3, \ldots\,.$$

Hence:

- $\zeta(z)$ is a meromorphic function of z, with a simple pole at $z = 1$. The residue at this pole is equal to unity.

Connection with Bernoulli numbers: We have seen earlier that $\zeta(2n)$, where n is a positive integer, is π^{2n} multiplied by a rational number less than unity. I have also mentioned there that no such simple closed-form expression exists for $\zeta(2n+1)$. In contrast, the value of $\zeta(z)$ at 0 and at the negative integers can be determined quite easily, as follows.

The function $1/(e^t - 1)$ has a simple pole at $t = 0$, with residue equal to unity. The function $t/(e^t - 1)$ is analytic in the neighborhood of the origin.[15] Its Taylor expansion about $t = 0$ is

$$\frac{t}{(e^t - 1)} = \sum_{n=0}^{\infty} B_n \, \frac{t^n}{n!} \, ,$$

where the constants B_n are certain rational numbers, called the **Bernoulli numbers**. These are *defined* by the expansion above. Thus $t/(e^t - 1)$ is the generating function for the Bernoulli numbers. The first few numbers are found to be

$$B_0 = 1, \; B_1 = -\tfrac{1}{2} \, , \; B_2 = \tfrac{1}{6} \, , \; B_4 = -\tfrac{1}{30} \, , \; B_6 = \tfrac{1}{42} \, , \ldots \, .$$

Interestingly, *all the odd Bernoulli numbers B_{2n+1} are equal to zero, except for B_1 .*

The sequence $|B_2|$, $|B_4|$, $|B_6|$, ... *appears* to be a decreasing sequence. You might therefore think that the series $\sum_{n=0}^{\infty} B_n \, t^n/n!$ ought to converge at least as well as the series for the exponential e^t. The radius of convergence would then be infinite, and the series would represent an entire function. But $t/(e^t - 1)$, the function represented by the series, has poles at $t = 2\pi n i$, where $n = \pm 1, \pm 2, \ldots$. Hence the radius of convergence of the series must be the distance from the origin to the nearest of these singularities, i.e., 2π. After the first few numbers, the Bernoulli numbers B_{2n} actually start *increasing* in magnitude quite rapidly with increasing n.

6. Use the expansion above in the representation of the zeta function to establish the following results:

(a) $\zeta(0) = -\frac{1}{2}$.

(b) $\zeta(-2n) = 0, \quad$ where $\quad n = 1, 2, \ldots$.

(c) $\zeta(1 - 2n) = -\dfrac{B_{2n}}{2n} \, , \quad$ where $\quad n = 1, 2, \ldots$.

[15] The function has a removable singularity at the origin, and tends to unity as $t \to 0$. We take this to be the value of the function at $t = 0$, as usual.

Hint: When z is zero or a negative integer, the factor t^{z-1} in the integrand does not have any branch points. Instead, there is a pole at $t = 0$. The contour C then collapses to a small circle encircling the pole once in the clockwise sense. The integral is easily evaluated by the residue theorem.

The Riemann Hypothesis: The zeroes of $\zeta(z)$ at even negative integer values of z are called the *trivial zeroes* of the zeta function. $\zeta(z)$ also has an infinite number of other zeroes in the strip $0 < \mathrm{Re}\, z < 1$.

- The famous **Riemann hypothesis** asserts that *all of these nontrivial zeroes lie on the so-called critical line* $\mathrm{Re}\, z = \frac{1}{2}$.

It is known that an infinite number of zeroes do lie on the critical line. It is also known that all nontrivial zeroes lie on that line 'almost surely' (in the sense of probability theory, with probability equal to 1). The first 13 billion or so zeroes have indeed been verified explicitly to lie on the critical line.[16] The Riemann conjecture is perhaps the most important unsolved problem in mathematics. It has resisted a rigorous and complete proof for over a century and a half.[17] A very large number of other results in mathematics rest on the validity of the hypothesis. There are also several intriguing connections between the distribution of the zeros of $\zeta(z)$ on the critical line, on the one hand, and physical problems, on the other—for instance, the level-spacing of the energy eigenvalues of quantum mechanical systems whose classical counterparts are chaotic, of the eigenvalues of certain classes of random matrices, and of the energy eigenvalues of complex nuclei. It is clear that the zeta function and its counterparts and analogs in classical and quantum dynamical systems hide many secrets yet to be discovered.

The Legendre functions $P_\nu(z)$ and $Q_\nu(z)$: As another example of integral representations of special functions, let's consider the **Legendre functions of the first and second kinds**, $P_\nu(z)$ and $Q_\nu(z)$, for complex z and ν. These functions appear in the solution of a very large number of physical problems. They are the linearly independent solutions of Legendre's differential equation,

$$\left\{(1 - z^2)\,\frac{d^2}{dz^2} - 2z\,\frac{d}{dz} + \nu(\nu + 1)\right\}\phi(z) = 0.$$

They turn out to have the contour integral representations given below, for complex values of both the argument z and the order ν.

Consider the function $f(t\,;\, z, \nu) = [(t^2 - 1)/2(t - z)]^\nu$ of the complex variable t, for general complex values of z and ν. There are branch points (winding points, in general)

[16] $\zeta(z)$ can be shown to be a real analytic function. Therefore if $\zeta(\frac{1}{2} + iy) = 0$ for any real value of y, it follows that $\zeta(\frac{1}{2} - iy) = 0$.

[17] You might want to give it a try!

at $t = 1$, $t = -1$ and $t = z$. Suppose we start at some point on the principal sheet of $f(t; z, \nu)$, and move in a path that encircles one or more of the branch points. In how many different ways can this be done, such that the function returns to its original value when we return to the starting point? In each case, the path then becomes a closed contour over which $f(t; z, \nu)$ can be integrated.

If either $t = 1$ or $t = -1$ is encircled once in the positive sense, the function acquires a factor $e^{2\pi i\nu}$. If $t = z$ is encircled once in the positive sense, it acquires a factor $e^{-2\pi i\nu}$. This shows that there are essentially three such independent paths. Using the Pochhammer notation, the contours are as follows. (i) $C_1 = (1+, z+)$; (ii) $C_2 = (1+, -1-)$; and (iii) $C_3 = (-1+, z+)$. More complicated paths can be decomposed into linear combinations of these basic paths. (Convince yourself that this is so.) The branch cut structure of $f(t; z, \nu)$ in each of the three cases is implicit in the statement that C_i is a closed contour for the function. In case (i), there is a cut of finite length running between $t = 1$ and $t = z$, and a cut from -1 to $-\infty$ along the negative real axis on the t-plane (say). In case (ii), there is a cut running from each of the three branch points to infinity. Case (iii) is similar to case (i), with the roles of the branch points at 1 and -1 interchanged.

The Legendre functions of the first and second kind are then given by the formulas

$$P_\nu(z) = \frac{1}{2\pi i} \oint_{C_1} \frac{dt}{(t-z)} \left[\frac{(t^2-1)}{2(t-z)}\right]^\nu$$

and

$$Q_\nu(z) = \frac{1}{\pi(e^{2\pi\nu i} - 1)} \oint_{C_2} \frac{dt}{(t-z)} \left[\frac{(t^2-1)}{2(t-z)}\right]^\nu .$$

The analytic properties of $P_\nu(z)$ and $Q_\nu(z)$ can be deduced from these representations. Of the extensive set of these properties, I mention the following:

$P_\nu(z)$ ***for non-integral order*** ν**:** For general non-integer values of the order ν, the function $P_\nu(z)$ *is no longer a polynomial in* z. It has branch points at $z = -1$ and $z = \infty$. It is conventional to choose the branch cut to run from -1 to $-\infty$ on the real axis (the x-axis) in the z-plane. The discontinuity across the cut at any any point x is again proportional to $P_\nu(|x|)$ itself.

$Q_\nu(z)$ ***for non-integral order*** ν**:** Similarly, for general non-integer values of the order ν, the function $Q_\nu(z)$ has branch points at $z = 1$, $z = -1$ and $z = \infty$. It is conventional to choose the branch cut to run from 1 to $-\infty$ on the x-axis. The discontinuity of $Q_\nu(x)$ for $x \in (-1, 1)$ is proportional to $P_\nu(x)$, while the discontinuity at $x < -1$ is proportional to $Q_\nu(|x|)$. Based on these properties, it is possible to write down *dispersion relations* for the functions $P_\nu(x)$ and $Q_\nu(x)$.

$P_l(z)$ ***for integer values of the order*** l**:** Consider $P_\nu(z)$ when $\nu = l$, where $l = 0, 1, \ldots$. The contour C_1 now encloses no branch points, but only a pole of order $(l+1)$ at $t = z$. The integral is then evaluated easily, and the **Rodrigues formula** for $P_l(z)$ is recovered:

$$P_l(z) = \frac{1}{2^l\, l!}\, \frac{d^l}{dt^l}(t^2-1)^l \bigg|_{t=z} \equiv \frac{1}{2^l\, l!}\, \frac{d^l}{dz^l}(z^2-1)^l.$$

You would have encountered this formula in the case when z is a real variable lying in the range $[-1, 1]$. We see now that this formula is also valid for complex values of z. Note that $P_l(z)$ continues to remain a polynomial of order l in the complex variable z.

When ν is a negative integer $-(l+1)$, the contour C_1 encloses a pole of order $(l+1)$ at $t = 1$. Once again, the integral can be evaluated. The symmetry property

$$P_{-l-1}(z) = P_l(z)$$

can then be deduced. This is just a special case of the more general reflection symmetry

$$P_{-\nu-1}(z) = P_\nu(z)$$

that is valid for all complex values of ν.

$Q_l(z)$ ***for integer values of the order*** l**:** Turning to $Q_\nu(z)$, when $\nu = l\,(= 0, 1, \ldots)$, the contour C_2 encloses no singularity at all. Hence the contour integral vanishes. But so does the factor $(e^{2\pi\nu i} - 1)$ in the denominator of the integral representation for $Q_\nu(z)$. Their ratio has a finite limit as $\nu \to l$. The outcome is the function $Q_l(z)$, which turns out to have logarithmic branch points at $z = \pm 1$. For instance,

$$Q_0(z) = \frac{1}{2}\, \ln\left(\frac{1+z}{1-z}\right).$$

In general, for positive integer values of l, we have

$$Q_l(z) = \frac{1}{2}\, P_l(z)\, \ln\left(\frac{1+z}{1-z}\right) + R_{l-1}\,,$$

where R_{l-1} is a polynomial of order $l-1$. There is a simple and very useful formula that connects $Q_l(z)$, for any arbitrary complex value of the argument z, to the Legendre polynomial of order l. It is

$$Q_l(z) = \frac{1}{2} \int_{-1}^{1} dt\, \frac{P_l(t)}{(z-t)}\,, \quad l = 0,\, 1,\, \ldots\,.$$

When ν is a negative integer $-(l+1)$, the contour C_2 encloses poles of order $(l+1)$ at $z=-1$ and $z=1$. The contour integral makes a finite, nonzero contribution. The factor $(e^{2\pi\nu i}-1)$ in the denominator then leads to a simple pole of $Q_\nu(z)$ at $\nu=-(l+1)$. The residue at the pole turns out to be $P_l(z)$ itself.

Finally, I mention that such contour integral representations exist for all other special functions as well. In most cases, they serve as analytic continuations of the corresponding functions to the largest possible domain of their arguments, including parameters such as the order, degree, and so on.

Singularities of functions defined by integrals: As I've just stated, integral representations of functions are most useful for exhibiting their analytic properties. The natural question that arises is the following:

- How does a function defined by an integral become singular, and what are the possible singularities?

As you might expect, the answer to such a general question involves whole branches of mathematics (such as **homology theory**). Let's narrow the question down very considerably, and examine the simplest possibilities. The treatment below is elementary, heuristic, and essentially based on a few simple examples.

Consider functions of the form

$$f(z)=\int_a^b dt\,\phi(t,z),$$

where the path of integration is an open contour running from some point a to some other point b in the complex t-plane. The integrand $\phi(t,z)$ is assumed to be analytic in t and z, with some singularities. In general, when the integral exists, it defines an analytic function of z. Let's start with a value of z in the region in which $f(z)$ is holomorphic, and examine how the function could develop a singularity as z moves out of this region. This is determined by the behavior of the singularities in the t-plane of the integrand $\phi(t,z)$.

These singularities are of two kinds: they could be *fixed* or z-independent, such as a pole at some point $t=c$ (not lying on the original path of integration); or they could be *movable*, or z-dependent, such as a pole at $t=z$. As z changes, one or more of the latter could approach the path of integration in the t-plane. If the contour can be distorted away so as to avoid the singularity (keeping it pinned down at the end-points a and b, of course), we have an analytic continuation of $f(z)$, defined by the integral over the distorted contour. There are two cases in which this simple device will *not* work. Each of them leads to a singularity of $f(z)$.

(i) *End-point singularity*: If a moving singularity approaches one of the end-points of integration, a or b, the contour cannot be moved away and an **end-point singularity** of $f(z)$ ensues. Consider the (extremely!) elementary example

$$f(z) = \int_0^1 \frac{dt}{t-z}\,.$$

The integrand has a moving pole at $t = z$. As it stands, the integral exists for all $z \notin [-1, 1]$. If z approaches the real axis in the t-plane either from above or from below, the contour of integration can be moved away ahead of the pole, and the integral will continue to exist in each case.[18] As $z \to 0$ or 1 (in the z-plane), the pole in the t-plane approaches one of the end-points of the contour, and a singularity of $f(z)$ occurs. This is corroborated by the explicit form for $f(z)$ obtained by carrying out the trivial integral above, to get

$$f(z) = \ln\Big(\frac{z-1}{z}\Big).$$

Clearly, $f(z)$ has logarithmic branch points at $z = 0$ and $z = 1$, confirming our expectation.

A little more generally, consider the function

$$f(z) = \int_0^1 dt\,\frac{\phi(t)}{(t-z)}\,,$$

where $\phi(t)$ is a well-behaved function (e.g., a polynomial in t) such that the integral exists as long as $z \notin [0, 1]$. Once again, $f(z)$ has end-point singularities at $z = 0$ and $z = 1$, and these are again logarithmic branch points. Note the important point that it is *not* necessary to be able to carry out the integral defining $f(z)$ explicitly in order to reach this conclusion! The *discontinuity* of $f(z)$ across the branch cut running from $z = 0$ to $z = 1$ on the real axis in the z-plane is also computed easily. All you have to do is to apply the formula we have already encountered (implicitly) when discussing dispersions relations: namely,

$$\frac{1}{t - x \mp i\epsilon} = \mathrm{P}\frac{1}{t-x} \pm i\pi\,\delta(t-x).$$

Here t and x are real variables, P stands for the Cauchy principal value, and the formula is to be understood in the following sense: multiply both sides by a smooth function of t, and integrate with respect to t over a range that includes the point x. The result is

$$\mathrm{disc}\ f(z)\Big|_{z=x\in(-1,1)} \stackrel{\mathrm{def.}}{=} \lim_{\epsilon\downarrow 0}\ \big[f(x+i\epsilon) - f(x-i\epsilon)\big] = 2\pi i\,\phi(x).$$

7. Apply this result to the formula $Q_l(z) = \frac{1}{2}\int_{-1}^{1} dt\,P_l(t)/(z-t)$, to show that

[18] As you might guess, however, the analytic continuations of $f(z)$ will differ in the two cases, suggesting already that we're dealing with different branches of $f(z)$.

(a) the Legendre function $Q_l(z)$ (where $l = 0,\ 1,\ \ldots$) has logarithmic branch points at $z = \pm 1$ (as stated earlier);

(b) the discontinuity across the branch cut running from -1 to 1 on the real axis is given by

$$\text{disc } Q_l(x)\Big|_{-1<x<1} = \lim_{\epsilon \downarrow 0} \left[Q_l(x + i\epsilon) - Q_l(x - i\epsilon)\right] = -i\pi\, P_l(x).$$

(ii) *Pinch singularity*: If two moving singularities approach the same point on the path of integration, but from *opposite* sides of the path, the contour gets trapped between them, and cannot 'escape'. A **pinch singularity** of $f(z)$ may then occur. The same thing happens if a moving singularity traps the contour between itself and a fixed singularity of the integrand lying on the other side of the contour. Consider the function

$$f(z) = \int_{-1}^{1} \frac{dt}{(z^2 - t^2)}\,.$$

The integral exists and defines an analytic function of z, as long as $z \notin [-1, 1]$. The integrand has moving poles at $t = \pm z$. We may expect end-point singularities of $f(z)$ at $z = \pm 1$. Further, the poles at $t = z$ and $t = -z$ pinch the contour of integration from opposite sides as $z \to 0$, and so a pinch singularity may be expected at $z = 0$. Explicit evaluation of the integral gives

$$f(z) = \frac{1}{z}\, \ln\left(\frac{z+1}{z-1}\right).$$

Thus, $f(z)$ does have singularities (logarithmic branch points) at $z = \pm 1$. It also has a pole at $z = 0$, on every sheet of the logarithm *except* the principal sheet (on which $\ln 1 = 0$).[19] As before, a similar analysis is applicable to the more general integral

$$f(z) = \int_{-1}^{1} dt\, \frac{\phi(t)}{z^2 - t^2}\,,$$

where $\phi(t)$ is a well-behaved function such as a polynomial.

***A slightly more complicated example*:** The nature of an end-point or pinch singularity depends also on the *kind* of moving singularities involved. In the preceding examples, these were simple poles. As an example of what can happen when two branch-points pinch the contour of integration, consider the integal

$$f(z) = \int_{-1}^{1} \frac{dt}{\sqrt{z^2 - t^2}}\,.$$

Again, the integral is very easily evaluated. But let's first list the possible singularities of the integral as it stands. The integrand has square-root branch points at $t = z$ and

[19] Recall that you have already encountered an example of this feature.

$t = -z$. End-point singularities may be expected at $z = \pm 1$. We may expect these to be 'mild' singularities, since $(t \pm z)^{-1/2}$ is an integrable singularity. (It is trivially seen that $f(z)$ does not diverge at $z = \pm 1$, but rather has the finite value π.) Further, as $z \to 0$, the contour of integration is pinched between these branch points. Hence $z = 0$ must also be a singularity of $f(z)$. Note that if we simply set $z = 0$ in the expression for $f(z)$, the integral diverges.

In order to find $f(z)$ explicitly, let's start with z on the positive real axis, $z = x > 1$. Then a simple change of the variable of integration yields

$$f(x) = 2 \sin^{-1}(1/x).$$

This is a multivalued function, and you must be careful about its branch structure when writing down its analytic continuation to the rest of the z-plane. It is convenient to re-write the arcsine function as a logarithmic function, using the identity

$$\sin^{-1} u = -i \ln \left[iu \pm (1 - u^2)^{1/2} \right].$$

In order to choose the right sign before the radical, note the following: When $1 \leq x < +\infty$, the integral representing $f(x)$ is real positive; and it decreases monotonically from π to 0 as $x \to \infty$. It follows that

$$f(x) = 2i \ln \left(\frac{x}{i + \sqrt{x^2 - 1}} \right), \quad 1 \leq x < \infty.$$

In this form, the function is (trivially) analytically continued to

$$f(z) = 2i \ln \left(\frac{z}{i + \sqrt{z^2 - 1}} \right).$$

It is now easy to see that $f(z)$ has

(i) a logarithmic branch point at 0, as well as

(ii) square-root branch points at $z = 1$ and $z = -1$,

corroborating our earlier conclusions.

Singularities of the Legendre functions: Finally, let's apply the foregoing to the integral representations written down earlier for the Legendre functions $P_\nu(z)$ and $Q_\nu(z)$. It should now be obvious that, for general values of the index ν, this is what happens:

- As $z \to -1$, the contour C_1 gets pinched between the moving singularity of the integrand at $t = z$ and the fixed singularity at $t = -1$. Hence $P_\nu(z)$ has a singularity at $z = -1$. The branch cut of $P_\nu(z)$ is customarily taken to run from -1 to $-\infty$ along the negative real axis.

- As $z \to \pm 1$, the contour C_2 gets pinched between the singularities of the integrand at $t = z$ and at $t = \pm 1$, respectively. Hence $Q_\nu(z)$ has a singularities at $z = 1$ and $z = -1$. The branch cut of $Q_\nu(z)$ is customarily taken to run from 1 through -1 to $-\infty$ along the real axis.

8. With a little effort, you can show that the discontinuities of $P_\nu(z)$ and $Q_\nu(z)$ across their branch cuts on the real axis are as follows:

$$\begin{aligned} \text{disc } P_\nu(x) &= 2i \sin(\pi\nu)\, P_\nu(-x) \qquad \text{for} \qquad -\infty < x < -1 \\ \text{disc } Q_\nu(x) &= 2i \sin(\pi\nu)\, Q_\nu(-x) \qquad \text{for} \qquad -\infty < x < -1 \\ \text{disc } Q_\nu(x) &= -i\pi\, P_\nu(x) \qquad\qquad \text{for} \qquad -1 < x < 1. \end{aligned}$$

Hint: In order to find the discontinuity of $P_\nu(z)$ across the cut from -1 to $-\infty$, consider its integral representation in the respective cases when $z = x + i\epsilon$ and $z = x - i\epsilon$, where $x < -1$. Note the configuration of the contour C_1 in the two cases, along with the branch cuts of the integrand. Write down the contour integrals segment by segment, keeping careful track of the phases of the various factors in the integrand. Compare the result with the contour integral for $P_\nu(-x)$ (where $-x > 1$), to arrive at the result quoted. A similar procedure will yield the corresponding results for $Q_\nu(x)$.

As expected, the discontinuity of $P_\nu(z)$ vanishes when $\nu = l$, an integer. $P_l(z)$ is a polynomial in z, and it has no singularity at $z = -1$ or at any other point in $\mathbb{C}$, the finite part of the complex z-plane. Similarly, $Q_\nu(z)$ has no branch cut running from -1 to $-\infty$ when $\nu = l$; the discontinuity across this cut vanishes identically. The discontinuity of $Q_\nu(z)$ between -1 and 1 is nonzero even when $\nu = l$, as you have seen already.

Dispersion relations for the Legendre functions: Finally, based on the discontinuities found above, we can derive dispersions relations for the Legendre functions of both kinds. As you might expect, the asymptotic ($|z| \to \infty$) behaviors of $P_\nu(z)$ and $Q_\nu(z)$ are also required for this purpose. I do not go into this here, but merely quote the relations, for completeness:

$$P_\nu(z) = \frac{\sin \pi\nu}{\pi} \int_1^\infty dt\, \frac{P_\nu(t)}{(z+t)} \quad (\nu \neq \text{integer}),$$

and

$$Q_\nu(z) = \frac{1}{2} \int_{-1}^1 dt\, \frac{P_\nu(t)}{(z-t)} + \frac{\sin \pi\nu}{\pi} \int_1^\infty dt\, \frac{Q_\nu(t)}{(z+t)}\,.$$

The relation $Q_l(z) = \frac{1}{2}\int_{-1}^1 dt\, P_l(t)/(z-t)$ when $\nu = l\,(= 0, 1, \dots)$ is a special case of the last formula above.

7 Laplace transforms

In physical applications, we are often concerned with functions that are only defined on a half-line, say $[0, \infty)$. An example is a causal response function $\phi(t)$, where t denotes the time variable. More generally, the whole class of initial value problems involves functions of this kind. Such functions may vanish, or tend to a constant, or even diverge as $t \to \infty$. It is very helpful to define an **integral transform** of the functions concerned, that turns differentiation with respect to t into multiplication by the variable 'conjugate' to t.

Definition of the Laplace transform: The **Laplace transform** of a function $f(t)$ (where $0 \leq t < \infty$) is defined as

$$\mathcal{L}[f(t)] = \widetilde{f}(s) \stackrel{\text{def.}}{=} \int_0^\infty dt\, e^{-st}\, f(t).$$

It is clear that the factor e^{-st} provides a convergence factor if $\mathrm{Re}\, s > 0$: the integral above converges even if $f(t)$ increases like any arbitrary power of t for large values of t. In fact, even if $f(t)$ increases exponentially with t, so that $f(t) \sim e^{ct}$ (where c is a positive number) as $t \to \infty$, its Laplace transform as given by the integral above is well-defined, as long as s is held in the region $\mathrm{Re}\, s > c$. Thus:

- The Laplace transform $\widetilde{f}(s)$ of a function $f(t)$, as given by its defining integral representation, is an analytic function of s for a sufficiently large positive value of $\mathrm{Re}\, s$, i.e., in some right-half plane in s.

In general, we may then expect to be able to define it to the left of this region by analytic continuation.[20] It is evident that $\widetilde{f}(s)$ would have one or more singularities in the left half-plane, in general.

The Laplace transforms of simple functions are easily written down. Consider, for instance, the function

$$f(t) = t^\alpha\, e^{-at}.$$

It is clear that its Laplace transform

$$\widetilde{f}(s) = \int_0^\infty dt\, t^\alpha\, e^{-(s+a)t}$$

exists as long as $\mathrm{Re}\, \alpha > -1$ (so that the lower limit of integration, $t = 0$, does not pose any problem), and also $\mathrm{Re}\, s > -\mathrm{Re}\, a$ (so that there is no divergence owing to the upper limit of integration, $t = \infty$). The integral is easily seen to be a gamma function. We have

$$f(t) = t^\alpha\, e^{-at}\ (\mathrm{Re}\, \alpha > -1) \quad \Longrightarrow \quad \widetilde{f}(s) = \frac{\Gamma(\alpha+1)}{(s+a)^{\alpha+1}}\,.$$

[20] Note that there do exist functions which do not possess a Laplace transform—e.g., if $f(t) \sim \exp\,(t^{1+\alpha})$ (where $\alpha > 0$) as $t \to \infty$, then $\widetilde{f}(s)$ does not exist for any value of s.

$\widetilde{f}(s)$ may now be continued analytically to all values of s, using the explicit representation above. Observe that $\widetilde{f}(s)$ has a singularity at $s = -a$, in accord with the general statement made earlier.

A number of simpler cases may be read off from this result. For instance,

$$\mathcal{L}[e^{-at}] = \frac{1}{(s+a)}\,,\ \mathcal{L}[t^n] = \frac{n!}{s^{n+1}}\,,\ \mathcal{L}[\cos(at)] = \frac{s}{s^2+a^2}\,,\ \mathcal{L}[\sin(at)] = \frac{a}{s^2+a^2}\,.$$

In general, if $\mathcal{L}[f(t)] = \widetilde{f}(s)$, then $\mathcal{L}[f(t)\,e^{-at}] = \widetilde{f}(s+a)$.

1. It is easy to see that

$$\int_0^\infty \frac{dt\,f(t)}{t} = \int_0^\infty ds\,\widetilde{f}(s),$$

provided both integrals exist. Numerous definite integrals may be evaluated with the help of this identity. Use it to recover the known result for the Dirichlet integral,

$$\int_0^\infty dt\,\frac{\sin(at)}{t} = \tfrac{1}{2}\pi\,\mathrm{sign}\,(a)$$

where a is a real constant.

2. The convolution theorem for Laplace transforms:

(a) If $f(t)$ and $g(t)$ have Laplace transforms $\widetilde{f}(s)$ and $\widetilde{g}(s)$, show that their convolution has the Laplace transform $\widetilde{f}(s)\,\widetilde{g}(s)$. That is,

$$\mathcal{L}\left\{\int_0^t dt'\,f(t')\,g(t-t')\right\} = \mathcal{L}\left\{\int_0^t dt'\,f(t-t')\,g(t')\right\} = \widetilde{f}(s)\,\widetilde{g}(s).$$

Hence note that $\mathcal{L}\left\{\int_0^t dt'\,f(t')\right\} = \widetilde{f}(s)/s$.

(b) Show that

$$\mathcal{L}\left\{\int_0^t dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} dt_1\,f(t-t_n)\,f(t_n-t_{n-1})\,\ldots f(t_2-t_1)\right\} = \left[\widetilde{f}(s)\right]^n.$$

3. Laplace transforms of derivatives: Using integration by parts, it is easy to see that

$$\mathcal{L}[df(t)/dt] = s\,\widetilde{f}(s) - f(0).$$

Let $f^{(n)}(t) \equiv d^n f(t)/dt^n$, with $f^{(0)}(t) \equiv f(t)$. Check that

$$\mathcal{L}[f^{(n)}(t)] = s\,\mathcal{L}[f^{(n-1)}(t)] - f^{(n-1)}(0), \quad n \geq 1.$$

Hence show that

$$\begin{aligned}\mathcal{L}[f^{(n)}(t)] &= s^n\,\widetilde{f}(s) - s^{n-1}\,f(0) - s^{n-2}\,f^{(1)}(0) - \ldots - s\,f^{(n-2)}(0) - f^{(n-1)}(0) \\ &= s^n\,\widetilde{f}(s) - \sum_{j=1}^{n} s^{n-j}\,f^{(j-1)}(0).\end{aligned}$$

Thus the Laplace transform essentially converts differentiation (with respect to t) to multiplication (by s). This fact is of great use in the applications of Laplace transforms—among others, in the solution of linear differential equations with constant coefficients. Note the occurrence of the 'initial data' (the values of the function and its first $(n-1)$ derivatives at $t=0$) in the expression for the transform of $d^n f(t)/dt^n$.

The inverse transform: The inverse Laplace transform that expresses $f(t)$ in terms of $\widetilde{f}(s)$ is given by the **Mellin formula**

$$f(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} ds\, e^{st}\,\widetilde{f}(s),$$

where the contour runs parallel to the imaginary axis in the s-plane, cutting the real axis at a point c such that the contour stays to the *right* of *all* the singularities of $\widetilde{f}(s)$. Hence the contour lies entirely in a region in which $\widetilde{f}(s)$ is analytic, as required to make the integral well-defined. (Recall that $\widetilde{f}(s)$ generally has singularities in a certain left half-plane.) Using the Mellin formula, it is easy to invert a Laplace transform whenever $\widetilde{f}(s)$ is a rational function of s. The contour over s can then be closed by adding a large semi-circle in the left half-plane, because the contribution of the semi-circle to the integral vanishes as its radius tends to infinity: the factor e^{st} in the Mellin formula ensures this. (Remember that $t \geq 0$.) The closed contour now encircles all the poles of $\widetilde{f}(s)$. The residue theorem then gives the value of the contour integral, and hence $f(t)$ is determined.

4. Show that the inverse transform of $\widetilde{f}(s) = (s^2+a^2)^{-2}$ is given by

$$f(t) = \frac{1}{2a^3}\,[\sin(at) - at\,\cos(at)].$$

5. The **Bessel function** of the first kind and of order ν may be defined by the power series

$$J_\nu(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{\Gamma(\nu+n+1)\,n!}\,\left(\tfrac{1}{2}z\right)^{2n+\nu}\,.$$

The order ν does not necessarily have to be an integer, in this definition. $J_\nu(z)$ is an entire function of z. When $\nu = 0$, we have

$$J_0(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(n!)^2}\,\left(\tfrac{1}{2}z\right)^{2n}\,.$$

Now consider the function $\widetilde{f}(s) = (s^2+a^2)^{-1/2}$, where a is a real constant (say). Expand it in a binomial series in inverse powers of s, and invert the transform term by term using the fact that the inverse transform of $1/s^{2n+1}$ is $t^{2n}/(2n)!$ Hence show that

$$\mathcal{L}^{-1}\left[\frac{1}{\sqrt{s^2+a^2}}\right] = J_0(at).$$

Since

$$\int_0^\infty dt\, e^{-st}\, J_0(at) = \frac{1}{\sqrt{s^2+a^2}}\,,$$

it follows that $\int_0^\infty dt\, J_0(at) = 1/a$ for $a > 0$.

6. The iterate of the Laplace transform: Show that, provided the integrals concerned exist,

$$\mathcal{L}^2[f(t)] = \int_0^\infty ds\, e^{-us} \int_0^\infty dt\, e^{-st}\, f(t) = \int_0^\infty dt\, \frac{f(t)}{(t+u)}\,.$$

The last integral is the so-called **Stieltjes transform** of $f(t)$.

LCR series circuit: As another standard example from elementary physics, consider an LCR series circuit under a sinusoidal applied voltage of amplitude V_0 and (angular) frequency ω. Recall that we have already determined the complex admittance of the system, when discussing linear response and dispersions relations.

The differential equation satisfied by the charge $q(t)$ on the capacitor is

$$L\,\ddot{q} + R\,\dot{q} + (1/C)\,q = V_0\,\sin\,\omega t.$$

As you know, this equation is precisely the equation of motion of a sinusoidally forced damped simple harmonic oscillator. The damping constant is $R/L = \gamma$, which is just the reciprocal of the time constant of an LR circuit. The natural frequency of the circuit in the absence of the resistor $1/\sqrt{LC} = \omega_0$. The condition $\omega_0 > \frac{1}{2}\gamma$ corresponds to the underdamped case. Let's consider this case first , for definiteness.[21] It is convenient, in this case, to work in terms of the shifted frequency

$$\omega_u = \left(\omega_0^2 - \tfrac{1}{4}\gamma^2\right)^{1/2}.$$

7. Suppose the initial conditions are such that both the initial charge and the initial current are equal to zero, i.e.,

$$q(0) = 0 \quad \text{and} \quad \dot{q}(0) = 0.$$

[21]The critically damped and overdamped cases will be considered subsequently.

The solution for $q(t)$ can be written as the sum of a *transient part* and a *steady state part*,

$$q(t) = q^{\rm tr}(t) + q^{\rm st}(t).$$

Show that these are given, respectively, by

$$q^{\rm tr}(t) = \left(\frac{V_0\,\omega}{L\,\omega_u}\right)\left\{\frac{(\omega^2-\omega_0^2+\frac{1}{2}\gamma^2)\,\sin(\omega_u t)+\omega_u\,\gamma\,\cos(\omega_u t)}{(\omega^2-\omega_0^2)^2+\omega^2\gamma^2}\right\}e^{-\gamma t/2}$$

and

$$q^{\rm st}(t) = -\left(\frac{V_0}{L}\right)\left\{\frac{(\omega^2-\omega_0^2)\,\sin(\omega t)+\omega\,\gamma\,\cos(\omega t)}{(\omega^2-\omega_0^2)^2+\omega^2\gamma^2}\right\}.$$

Hint: Take the Laplace transform of both sides of the differential equation. With the initial conditions $q(0)=0$ and $\dot{q}(0)=0$, we get

$$\widetilde{q}(s) = \frac{(V_0\,\omega/L)}{(s^2+\omega^2)(s^2+\gamma s+\omega_0^2)}\,.$$

Resolve the right-hand side into partial fractions, and invert the Laplace transform. All you need to use is the fact that the inverse Laplace transform of $(s+a)^{-1}$ is e^{-at}.

Observe that the transient component of the solution, $q^{\rm tr}(t)$, is characterized by the frequency ω_u that depends on the circuit parameters L, C and R. This part decays to zero exponentially in time, owing to the dissipation in the system. The steady state component $q^{\rm st}(t)$, on the other hand, oscillates with the same frequency ω as the applied voltage. These statements apply equally to the current in the circuit, given by $I(t)=\dot{q}(t)$.

8. The remarks just made are, in fact, applicable to all initial conditions. In order to check this out, consider general values $q(0)$ and $\dot{q}(0)=I(0)$, respectively, of the initial charge on the capacitor and the initial current in the circuit. Show that the complete solution for $q(t)$ is now given by the one already found above, plus an extra term added to $q^{\rm tr}(t)$, namely,

$$\left\{\left(\frac{I(0)+\frac{1}{2}\gamma\,q(0)}{\omega_u}\right)\sin(\omega_u t)+q(0)\,\cos(\omega_u t)\right\}e^{-\gamma t/2}.$$

***Complementary function and particular integral*:** Note, incidentally, that the last expression above is precisely the solution to the *homogeneous* differential equation $L\,\ddot{q}+R\,\dot{q}+(1/C)\,q=0$ that is satisfied by the charge on the capacitor in the *absence* of any applied voltage. With reference to the inhomogeneous differential equation for $q(t)$, the full solution given above represents the **particular integral**, while the solution of the homogeneous equation represents the **complementary function**.

- In the solution of an inhomogeneous differential equation, the purpose of adding 'the right amount' of the complementary function to the particular integral is to ensure that the boundary conditions (in this case, the initial conditions) are satisfied by the solution.

The example just considered ought to help you see this quite clearly.

9. By now, it should be obvious to you that the solutions in the critically damped ($\omega_0 = \frac{1}{2}\gamma$) and overdamped ($\omega_0 < \frac{1}{2}\gamma$) cases may be written down by simple analytic continuation of the solution in the underdamped case.[22] Do so.

Laplace transforms and random processes: The master equations for certain Markov processes lead to simple differential equations for the generating functions of these processes. Such equations are solved very easily using Laplace transforms. Here are a few examples.

10. The Poisson process: Radioactive decay of an unstable isotope provides a physical example of a Poisson process. If $P_n(t)$ is the probability that exactly n events of the process take place in a time interval t, and $\lambda > 0$ is the mean rate at which events take place, then

$$\frac{dP_0(t)}{dt} = -\lambda\, P_0(t) \quad \text{and} \quad \frac{dP_n(t)}{dt} = \lambda\left[P_{n-1}(t) - P_n(t)\right], \quad n \geq 1.$$

This coupled set of ordinary differential equations is to be solved with the initial conditions $P_0(0) = 1$ and $P_n(0) = 0$ for $n \geq 1$. It follows at once that $P_0(t) = e^{-\lambda t}$. The generating function $f(z,t) = \sum_{n=1}^{\infty} P_n(t)\, z^n$ satisfies the differential equation[23]

$$\frac{\partial f}{\partial t} + \lambda(1-z)\, f = \lambda z\, P_0 = \lambda z\, e^{-\lambda t}\,,$$

with the initial condition $f(z,0) = 0$. Use Laplace transforms to obtain the solution

$$f(z,t) = e^{-\lambda t}\,\left(e^{\lambda z t} - 1\right).$$

Picking out the coefficient of z^n in the power series expansion of the exponential we get the Poisson distribution

$$P_n(t) = \frac{e^{-\lambda t}\,(\lambda t)^n}{n!}\,.$$

The mean number of events in a time interval t is therefore λt. Every cumulant of the distribution is also equal to λt.

[22] Typically, trigonometric functions will become hyperbolic functions.

[23] Note that the generating function has been defined as a sum from $n = 1$ rather than $n = 0$.

The Poisson process is an example of what is known as a **birth process**, because the random variable $n(t)$ never decreases as t increases. The next example shows what happens when the value of the random variable can both increase as well as decrease as time elapses (a **birth-and-death process**).

11. A biased random walk on a linear lattice: Consider a linear lattice, with its sites labelled by an integer $j \in \mathbb{Z}$. A walker on this lattice jumps from any site to one of the two neighboring sites with a mean jump rate λ. The probability of a jump from the site j to the site $(j+1)$ is p, and that of a jump to the site $(j-1)$ is $q = (1-p)$, where $0 < p < 1$. Successive jumps are statistically independent of each other. Let $P_j(t)$ denote the probability that the walker is at site j at time t. (The random variable in this problem is j). The master equation satisfied by the set of probabilities $\{P_j(t)\}$ is

$$\frac{dP_j(t)}{dt} = \lambda \, \big[\underbrace{p\, P_{j-1}(t) + q\, P_{j+1}(t)}_{\text{'gain' terms}} - \underbrace{P_j(t)}_{\text{'loss' term}} \big], \quad j \in \mathbb{Z}.$$

Without loss of generality,[24] we may take the site 0 to be the starting point of the random walk (RW for short). The initial condition is then

$$P_j(0) = \delta_{j,0}\,.$$

Define the generating function

$$f(z,t) = \sum_{j=-\infty}^{\infty} P_j(t)\, z^j.$$

Note that the summation is over *all* integers j. Hence the equation above is to be regarded as a Laurent series expansion of $f(z,t)$.

(a) From the differential equation satisfied by $P_j(t)$ and its initial condition, show that $f(z,t)$ satisfies the equation

$$\frac{\partial f}{\partial t} = \lambda\,(pz - 1 + qz^{-1})\, f\,,$$

with the initial condition $f(z,0) = 1\,.$

(b) Solve this equation using Laplace transforms, to get

$$f(z,t) = e^{-\lambda t}\, e^{\lambda t\,(pz + qz^{-1})}.$$

[24] Because the lattice is of *infinite* extent in both directions! Finite boundaries would obviously spoil this translation invariance.

(c) The probability we seek, $P_j(t)$, is the coefficient of z^j in the Laurent expansion of $f(z,t)$ in powers of z. It is helpful to write

$$pz + qz^{-1} = \sqrt{pq}\left(z\sqrt{p/q} + z^{-1}\sqrt{q/p}\right).$$

Now expand the second exponential factor in the solution for $f(z,t)$, and pick out the coefficient of z^j. Consider $j \geq 0$ first. Show that

$$P_j(t) = e^{-\lambda t}\,(p/q)^{j/2} \sum_{n=0}^{\infty} \frac{1}{n!\,(n+j)!}\,(\lambda t\sqrt{pq})^{2n+j}\,, \quad j \geq 0.$$

The modified Bessel function of the first kind and of order ν may be defined by means of its power series, namely,

$$I_\nu(z) = \sum_{n=0}^{\infty} \frac{1}{\Gamma(\nu+n+1)\,n!}\,\left(\tfrac{1}{2}\,z\right)^{2n+\nu}.$$

Compare this series with that for the ordinary Bessel function $J_\nu(z)$ given earlier. Apart from the factor $(-1)^n$ in the summand in the case of $J_\nu(z)$, the series are the same.[25] Like the Bessel function $J_\nu(z)$, the modified Bessel function $I_\nu(z)$ is also an entire function of z. The generating function for the modified Bessel function of integer order is[26]

$$e^{t\,(z+z^{-1})/2} = \sum_{j=-\infty}^{\infty} I_j(t)\,z^j\,.$$

Owing to the symmetry of the generating function under the exchange $z \leftrightarrow z^{-1}$, it is obvious that $I_{-j}(z) = I_j(z)$ for every integer value of j.

(d) Compare the series expansion obtained above for $P_j(t)$ with the power series for the modified Bessel function, to obtain

$$P_j(t) = e^{-\lambda t}\,(p/q)^{j/2}\,I_j\,(2\lambda t\sqrt{pq})\,.$$

Verify that exactly the same expression is valid for negative integer values of j as well, using the fact that $I_{-j} = I_j$ for any integer j.

[25] The two functions are related to each other according to $J_\nu(z) = e^{i\pi\nu/2}\,I_\nu(-iz)$, or $I_\nu(z) = e^{-i\pi\nu/2}\,J_\nu(iz)$.

[26] For completeness, I mention that the generating function of the Bessel function J_k is given by

$$e^{t\,(z-z^{-1})/2} = \sum_{j=-\infty}^{\infty} J_j(t)\,z^j\,.$$

An RW in which $p \neq q$ is called a **biased random walk**. When $p > q$ [respectively, $q > p$], there is a bias to the right [left], and the factor $(p/q)^{j/2}$ in the expression for $P_j(t)$ shows that the probability of positive [negative] values of j is enhanced, at any $t > 0$.

In the case of an **unbiased random walk**, $p = q = \frac{1}{2}$. The probability distribution then simplifies to

$$P_j(t) = e^{-\lambda t}\, I_j(\lambda t) \quad \text{(unbiased RW in 1 dimension)}.$$

Since $I_j = I_{-j}$, we have in this case $P_j(t) = P_{-j}(t)$ at any time t, as expected on physical grounds.

The asymptotic or long-time behavior of the RW is of importance. In the problem at hand, the characteristic time scale is set by λ^{-1}. Now, for $\lambda t \gg 1$, the leading asymptotic behavior of the modified Bessel function is given by

$$I_j(\lambda t) \sim \frac{e^{\lambda t}}{\sqrt{2\pi\lambda t}}\,,$$

independent of j. Therefore, in an unbiased RW, $P_j(t)$ has a leading asymptotic time-dependence $\sim t^{-1/2}$, which is characteristic of purely **diffusive behavior**.[27] In marked contrast, the leading asymptotic behavior of $P_j(t)$ for a *biased* random walk is given by $P_j(t) \sim t^{-1/2}\, e^{-\lambda t(1-2\sqrt{pq})}$. Since $1 > 2\sqrt{pq}$ when $p \neq q$, this shows that $P_j(t)$ decays exponentially with time for all j.

12. Unbiased random walk in d dimensions: Consider an unbiased simple RW on an infinite hypercubic lattice in d dimensions. Each site is labelled by a set of d integers, $(j_1, \ldots, j_d) \equiv \mathbf{j}$. The walker jumps from any given site to any one of the $2d$ nearest-neighbor sites with a probability $1/(2d)$, with a mean transition rate λ. Let $P(\mathbf{j}, t)$ be the probability that the walker is at $\mathbf{j}$ at time t, given that the walk started at the origin $\mathbf{0}$ at $t = 0$. The master equation satisfied by $P(\mathbf{j}, t)$ is

$$\frac{dP(\mathbf{j},t)}{dt} = \underbrace{\frac{\lambda}{2d}\sum_{\boldsymbol{\delta}} P(\mathbf{j}+\boldsymbol{\delta},t)}_{\text{gain term}} - \underbrace{\lambda P(\mathbf{j},t)}_{\text{loss term}},$$

where $\boldsymbol{\delta}$ stands for a nearest-neighbor vector of the site $\mathbf{j}$. (That is, $\boldsymbol{\delta}$ has one of its d components equal to ± 1, and all other components equal to 0.) The initial condition is

$$P(\mathbf{j}, 0) = \begin{cases} 1, & \text{if} \quad \mathbf{j} = \mathbf{0} \\ 0, & \text{if} \quad \mathbf{j} \neq \mathbf{0}. \end{cases}$$

[27] We'll study the diffusion equation and its properties later on.

Define the generating function

$$f(z_1\,,\,\ldots\,,\,z_d\,,\,t) = \sum_{\mathbf{j}} P(\mathbf{j},t)\, z_1^{j_1}\, z_2^{j_2}\, \cdots\, z_d^{j_d},$$

where each component of $\mathbf{j}$ is summed over all the integers.

(a) Show that the equation satisfied by this generating function is

$$\frac{\partial f}{\partial t} = \lambda\Big[\,\frac{z_1 + z_1^{-1} + \cdots + z_d + z_d^{-1}}{2d}\,-1\Big]\,f,$$

with the initial condition $f = 1$ at $t = 0$.

(b) The solution to this equation is clearly an exponential that factors into a product of generating functions of modifed Bessel functions. Show that

$$P(\mathbf{j},t) = e^{-\lambda t}\, I_{j_1}(\lambda t/d)\, I_{j_2}(\lambda t/d)\, \cdots\, I_{j_d}(\lambda t/d).$$

The leading asymptotic ($t \to \infty$) behavior of $P(\mathbf{j},t)$ follows from that of the modified Bessel function. We find $P(\mathbf{j},t) \sim t^{-d/2}$. This specific power-law decay of the probability is characteristic of purely diffusive motion.

13. The generalization of the solution above to include any directional bias is straightforward. Let p_i and q_i be the probabilities of a jump in which j_i changes to $j_i + 1$ and $j_i - 1$, respectively. Then $\sum_1^d (p_i + q_i) = 1$.

(a) Write down the master equation satisfied by $P(\mathbf{j},t)$.

(b) Show that the solution for an RW starting at the origin at $t = 0$ is

$$P(\mathbf{j},t) = e^{-\lambda t} \prod_{i=1}^{d} (p_i/q_i)^{j/2}\, I_{j_i}\big(2\lambda t\sqrt{p_i\, q_i}\,\big).$$

As in the one-dimensional case, the long-time behavior of $P(\mathbf{j},t)$ is now a decaying exponential in t, rather than a pure power-law fall-off.

8 Fourier transforms

Fourier integrals: As you know, **periodic functions** can be expanded in **Fourier series**. Recall that, if the fundamental interval of the function $f(x)$ is $(a\,,\,b)$ so that the period is $L = (b-a)$, the expansion and inversion formulas are

$$f(x) = (1/L)\sum_{n=-\infty}^{\infty} f_n\, e^{2\pi nix/L} \quad \text{and} \quad f_n = \int_a^b dx\, f(x)\, e^{-2\pi nix/L}.$$

What happens if the function is not periodic? By letting $a \to -\infty$ and $b \to \infty$, and hence passing to the limit $L \to \infty$, we can extend the idea of the expansion of an arbitrary periodic function in terms of elementary periodic functions (sines and cosines) to functions that need not be periodic. The number of 'harmonics' required now becomes *uncountably* infinite. Therefore, instead of a summation over the integer index n, we require an *integration* over the continuous variable k to which $2\pi n/L$ tends in the limit. The general coefficient in the original expansion, f_n, becomes a function of the continuous variable k. In order to avoid confusion with $f(x)$, we may denote this function by $\widetilde{f}(k)$. The 'dictionary' to go from Fourier series to Fourier integrals is

$$\frac{2\pi n}{L} \longrightarrow k\,, \quad \frac{1}{L}\sum_{n=-\infty}^{\infty} \longrightarrow \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\,, \quad f_n \longrightarrow \widetilde{f}(k)\,.$$

We then have the expansion formula

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{ikx}\, \widetilde{f}(k) \quad \text{(expansion formula)}.$$

The inversion formula is

$$\widetilde{f}(k) = \int_{-\infty}^{\infty} dx\, e^{-ikx}\, f(x) \quad \text{(inversion formula)}.$$

The function $\widetilde{f}(k)$ is obtained from the function $f(x)$ by acting upon the latter with an **integral operator** whose **kernel** is $\exp(-ikx)$. Similarly, the function $f(x)$ is obtained from the function $\widetilde{f}(k)$ by acting upon the latter with an integral operator whose kernel is $(2\pi)^{-1}\exp(ikx)$. The functions $f(x)$ and $\widetilde{f}(k)$ are **Fourier transforms** of each other.[28] A Fourier transform is an example of an **integral transform**. A function $f(x)$ has a Fourier transform if it satisfies conditions analogous to the Dirichlet conditions for periodic functions. Broadly speaking, if

(i) $f(x)$ has at most a finite number of finite discontinuities or jumps, and

[28] Which of the two we call the transform, and which we call the inverse transform, is a matter of convention.

(ii) $f(x)$ is absolutely integrable in $(-\infty\,,\,\infty)$, so that $\int_{-\infty}^{\infty} dx\,|f(x)| < \infty$ (or $f(x) \in \mathcal{L}_1(-\infty\,,\,\infty)$),

then its Fourier transform exists.

Note also the Fourier transform conventions used here:

— When integrating over k, the kernel is $(2\pi)^{-1}e^{+ikx}$.

— When integrating over x, the kernel is e^{-ikx}.

I repeat that these are just matters of convention,[29] but it is quite important to choose some specific convention and stick to it consistently, in order to avoid errors.

Parseval's formula for Fourier transforms: If $f(x)$ is not only integrable but is also square-integrable, i.e., if $f(x) \in \mathcal{L}_2(-\infty\,,\,\infty)$, it follows on applying Parseval's formula that its Fourier transform $\widetilde{f}(k)$ is *also* an element of $\mathcal{L}_2(-\infty\,,\,\infty)$: we have

$$\int_{-\infty}^{\infty} dx\,|f(x)|^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\,|\widetilde{f}(k)|^2.$$

(This relation is just the continuum analog of the discrete version of Parseval's formula for Fourier series.) The connection between $\mathcal{L}_2$ functions and Fourier tranforms should be familiar to you from elementary quantum mechanics. The Fourier transform, as applied to elements of the function space $\mathcal{L}_2$, may be regarded as *a change of basis*. In the context of quantum mechanics, Parseval's formula implies that, if the position-space wave function of a particle is normalizable, then so is its momentum-space wave function, because these two wave functions form a Fourier transform pair.

Fourier transform of the δ-function: The Dirichlet conditions and the conditions stated above are *sufficient* conditions for a function $f(x)$ to have a Fourier series expansion or a Fourier transform, as the case may be. They are by no means *necessary* conditions. For instance, numerous functions that are integrable, but not absolutely integrable (i.e., their absolute values are not integrable) have Fourier transforms. Functions that are more *singular* than permitted by the Dirichlet conditions may also have Fourier representations. The theory of generalized functions or distributions is, in fact, very closely linked with the Fourier transforms of these objects. There is a highly-developed area of mathematics, harmonic analysis, dealing with these matters and their generalizations.

[29] For instance, as you'll see below, the kernels $(2\pi)^{-1/2}e^{ikx}$ and $(2\pi)^{-1/2}e^{-ikx}$ are often used in the expansion formula and the inversion formula, instead of the kernels $(2\pi)^{-1}e^{ikx}$ and e^{-ikx} that we have used.

The Dirac δ-function, which is so useful in applications, also has a Fourier transform. Since we know that

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{ikx}\,,$$

it follows that the Fourier transform of $\delta(x)$ is just unity, i.e., $\widetilde{\delta}(k) = 1$.

1. Sketch the functions $f(x)$ listed below, and show that their Fourier transforms are as given. Here a and σ are positive constants, while b and μ are real constants. $\theta(x)$ is the unit step function.

(a) $f(x) = e^{-a|x|}\sin\, bx \qquad \Longrightarrow \qquad \widetilde{f}(k) = \dfrac{-4iabk}{(a^2+b^2+k^2)^2 - 4b^2k^2}\,.$

(b) $f(x) = e^{-a|x|}\cos\, bx \qquad \Longrightarrow \qquad \widetilde{f}(k) = \dfrac{2a(a^2+b^2+k^2)}{(a^2+b^2+k^2)^2 - 4b^2k^2}\,.$

(c) $f(x) = \dfrac{[\theta(x+a) - \theta(x-a)]}{2a} \qquad \Longrightarrow \qquad \widetilde{f}(k) = \dfrac{\sin\, ka}{ka}\,.$

(d) $f(x) = \dfrac{\sin\, ax}{x} \qquad \Longrightarrow \qquad \widetilde{f}(k) = \pi\,[\theta(k+a) - \theta(k-a)].$

(e) $f(x) = \dfrac{1}{\sqrt{2\pi\,\sigma^2}}\, e^{-(x-\mu)^2/(2\sigma^2)} \qquad \Longrightarrow \qquad \widetilde{f}(k) = e^{-i\mu k - \frac{1}{2}k^2\sigma^2}.$

The functions in (c) and (e) above can be regarded as normalized probability density functions corresponding to a random variable x. The first of these corresponds to a **uniform distribution**, while the second corresponds, of course, to the normal or **Gaussian distribution**. The Fourier transform of the probability density of a random variable is called its **characteristic function**. Knowing the characteristic function of a random variable is equivalent to knowing all its moments and cumulants.

2. Sketch the functions $\widetilde{f}(k)$ listed below, and show that their inverse Fourier transforms $f(x)$ are as given. (σ, k_0 and λ are positive constants.)

(a) $\widetilde{f}(k) = e^{-\frac{1}{2}(k-k_0)^2\sigma^2} \;\Longrightarrow\; f(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\, e^{ik_0x}\, e^{-x^2/(2\sigma^2)}.$

(b) $\widetilde{f}(k) = \dfrac{k}{|k|}\,\theta(k_0 - |k|) \;\Longrightarrow\; f(x) = \dfrac{i(1-\cos\, k_0x)}{\pi x}\,.$

(c) $\widetilde{f}(k) = \theta(k_0 - |k|)\,(k_0 - |k|) \;\Longrightarrow\; f(x) = \dfrac{1 - \cos(k_0x) + k_0x\,\sin(k_0x)}{2\pi x^2}\,.$

(d) $\widetilde{f}(k) = e^{-\lambda|k|} \;\Longrightarrow\; f(x) = \dfrac{\lambda}{\pi\,(x^2+\lambda^2)}\,.$

The function $f(x)$ in case (d) is a 'Lorentzian'. It is is the probability distribution function corresponding to the **Cauchy distribution**.

Relative 'spreads' of a Fourier transform pair: A general feature of great importance emerges from the examples above. Roughly speaking, when the function $f(x)$ is of *compact support*, i.e., it is nonzero in only a finite interval in x and is zero outside it, or is mostly concentrated in a finite interval and rapidly decreases to zero outside it, its transform $\widetilde{f}(k)$ is spread out in k. This is why the rectangular pulse considered earlier, which is strictly zero for $|x| > a$, has a Fourier transform that decays to zero relatively slowly, like $|k|^{-1}$. The same statement is applicable with the roles of $\widetilde{f}(k)$ and $f(x)$ interchanged. The compact pulses represented by $\widetilde{f}(k)$ in the examples above have transforms that decay like $|x|^{-1}$ for large $|x|$. When one member of the transform pair decays exponentially, which means it falls off faster than any inverse power of the argument, the other member decays like the inverse square of the argument. This is demonstrated by case of the Lorentzian. Finally, when one of the pair is a Gaussian (with possible linear terms in the exponent), so is the other member of the pair. This result is so useful that I'll write it down again for ready reference:

$$f(x) = \frac{1}{\sqrt{2\pi\,\sigma^2}}\, e^{-(x-\mu)^2/(2\sigma^2)} \qquad \Longleftrightarrow \qquad \widetilde{f}(k) = e^{-i\mu k - \frac{1}{2}k^2\sigma^2}.$$

The feature under discussion has a far-reaching physical consequence. This sort of 'duality'—compactness in one variable, spreading in the 'conjugate' variable—is at the very heart of the Heisenberg Uncertainty Principle. The fact that $\hbar \not\equiv 0$ necessitates a quantum mechanical description of systems, in terms of state vectors (or wave functions). In turn, the 'fuzziness' implicit in such a description is inevitably subject to the duality mentioned above.

The convolution theorem for Fourier transforms is a basic and most useful relationship. Given two functions $f(x)$ and $g(x)$, the function

$$h(x) = \int_{-\infty}^{\infty} dx'\, f(x')\, g(x - x') = \int_{-\infty}^{\infty} dx'\, f(x - x')\, g(x')$$

is called the *convolution* of the functions $f(x)$ and $g(x)$. This relationship is sometimes written as $h = f * g$. (This should not be confused with ordinary multiplication of the functions concerned.)

3. Establish the **convolution theorem** for Fourier transforms: If $f(x)$ and $g(x)$ have Fourier transforms $\widetilde{f}(k)$ and $\widetilde{g}(k)$ respectively, then the Fourier transform of their convolution $h(x)$ is given by $\widetilde{h}(k) = \widetilde{f}(k)\ \widetilde{g}(k)$.

Thus, the Fourier transform of the convolution of two functions is just the product of the Fourier transforms of the two functions. The Fourier transform operation converts

convolution to an ordinary multiplication.

4. It should be clear from the foregoing that this works in reverse, too. Consider the Fourier transform of the convolution of $\widetilde{f}$ and $\widetilde{g}$, namely, of the function of k given by

$$\widetilde{H}(k) = \int_{-\infty}^{\infty} dk'\, \widetilde{f}(k')\, \widetilde{g}(k-k') \equiv \int_{-\infty}^{\infty} dk'\, \widetilde{f}(k-k')\, \widetilde{g}(k').$$

Check that the inverse Fourier transform of $\widetilde{H}(k)$ is simply the product $H(x) = f(x)\, g(x)$.

5. Generalization of Parseval's formula: Let $f(x)$ and $F(x)$ be two *good functions*[30] of x, i.e., they have derivatives of all orders for all x, and vanish (along with their derivatives of all orders) as $x \to \infty$. Show that

$$\int_{-\infty}^{\infty} dx\, f(x)\, F(-x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, \widetilde{f}(k)\, \widetilde{F}(k)\,.$$

Note that (i) neither side of this equation is in the form of a convolution of two functions; (ii) there is a minus sign present in the argument of one of the functions in the integrand; and (iii) each side of the equation is a *number*, and not a function. We'll use this result in the sequel, in the derivation of the Poisson summation formula.

Note that Parseval's formula itself follows as a special case of the result above. All you have to do is to identify $F(-x)$ with $f^*(x)$, and observe that

$$F(-x) = f^*(x) \quad \Longleftrightarrow \quad \widetilde{F}(k) = \widetilde{f}^{\,*}(k).$$

Iterates of the Fourier transform operator: We've seen that the Fourier transform $\widetilde{f}(k)$ of a function $f(x)$ may be regarded as the result of applying a certain integral operator, namely, the Fourier transform operator $\mathcal{F}$, to the function. The kernel of the operator is just e^{-ikx}. In order to make the notion precise, we need to specify a function space, such that both the function $f(x)$ and its transform $\widetilde{f}(k)$ belong to the same space. As we know, the space $\mathcal{L}_2(-\infty\,,\,\infty)$ satisfies this requirement. Let us therefore restrict ourselves to this function space in what follows.

Given a function $f \in \mathcal{L}_2(-\infty\,,\,\infty)$, we have

$$[\mathcal{F}f](x) \equiv \widetilde{f}(x) = \int_{-\infty}^{\infty} dy\, e^{-ixy}\, f(y)\,.$$

[30] I believe this simple but expressive terminology is due to M. J. Lighthill.

(Observe the notation in this equation! I have written x for the argument of the *output* function *after* the Fourier transform operation is performed.) Something very interesting happens when the operator $\mathcal{F}$ is *iterated* i.e., applied repeatedly to a function. We have

$$\begin{aligned}
[\mathcal{F}^2 f](x) &\equiv [\mathcal{F}\mathcal{F}f](x) = [\mathcal{F}\widetilde{f}\,](x) = \int_{-\infty}^{\infty} dz\, e^{-ixz}\, \widetilde{f}(z) \\
&= \int_{-\infty}^{\infty} dz\, e^{-ixz} \int_{-\infty}^{\infty} dy\, e^{-izy}\, f(y) = \int_{-\infty}^{\infty} dy\, f(y) \int_{-\infty}^{\infty} dz\, e^{-i(x+y)z} \\
&= \int_{-\infty}^{\infty} dy\, f(y)\, 2\pi\, \delta\,(x+y) = 2\pi\, f(-x)\,.
\end{aligned}$$

Now, we know that the **parity operator** $\mathcal{P}$ changes the sign of the argument when acting on any function of x. That is, $(\mathcal{P}f)(x) \equiv f(-x)$. The result above shows that the square of the Fourier transform operator is just 2π times the parity operator. In symbols,[31]

$$\mathcal{F}^2 = 2\pi\, \mathcal{P}.$$

Moreover, since the square of the parity operator is obviously just the unit operator I, we have the operator relationship

$$\mathcal{F}^4 = (2\pi)^2\, \mathcal{P}^2 = (2\pi)^2\, I.$$

The Fourier transform operator is thus proportional to a 'fourth root' of the unit operator! These results find applications in **Fourier optics**.

Eigenvalues and eigenfunctions of $\mathcal{F}$: The fact that $\mathcal{F}^4 = (2\pi)^2\, I$ seems to suggest that the eigenvalues of the Fourier transform operator in $\mathcal{L}_2$ are given by $(2\pi)^{1/2}$ times the fourth roots of unity, namely, the four numbers $\pm(2\pi)^{1/2}$ and $\pm i\,(2\pi)^{1/2}$. Can we find the eigenvalues and the corresponding eigenfunctions explicitly?

Recall that the Hamiltonian of the quantum mechanical linear harmonic oscillator has exactly the same form in the position basis and in the momentum basis: In units such that m, ω and $\hbar$ are all equal to unity, the Hamiltonian is represented by the operators

$$\frac{1}{2}\Big(-\frac{d^2}{dx^2} + x^2\Big) \quad \text{and} \quad \frac{1}{2}\Big(-\frac{d^2}{dp^2} + p^2\Big),$$

respectively, in the position basis and the momentum basis. As a result, the normalized position-space wave functions $\Phi_n(x)$ and the momentum-space wave functions $\widetilde{\Phi}_n(p)$ representing the eigenstates of this Hamiltonian must be identical in form. But we also know that $\Phi_n(x)$ and $\widetilde{\Phi}_n(p)$ are elements of $\mathcal{L}_2(-\infty\,,\,\infty)$, and that they are Fourier transform pairs. It follows that these are precisely the eigenfunctions of $\mathcal{F}$ that we're

[31] The factor of 2π is a consequence of the particular Fourier transform convention I have used. It should be obvious that the factor can be made equal to unity by a suitable choice of convention.

looking for. It remains to prove this assertion directly, and also to find the eigenvalues of $\mathcal{F}$ explicitly. Let's now do so.

We want to show that the functions

$$\Phi_n(x) = \frac{1}{\left(2^n n! \sqrt{\pi}\right)^{1/2}} \, e^{-x^2/2} \, H_n(x) \quad (n = 0, 1, \ldots),$$

where $H_n(x)$ is the Hermite polynomial of degree n, are eigenfunctions of $\mathcal{F}$. The n-dependent constant of proportionality is irrelevant in an eigenvalue equation, and so let's consider the function $e^{-x^2/2} \, H_n(x)$. Now, the generating function for the Hermite polynomials is given by

$$e^{2tx - t^2} = \sum_{n=0}^{\infty} H_n(x) \, \frac{t^n}{n!} \, .$$

Such an equation must be regarded as an equation between two *analytic* functions of the complex variable t. The left-hand side is an entire function of t. The series on the right-hand side must therefore converge for all finite values of $|t|$.

Multiply both sides of the last equation by $\exp\left(-ikx - \frac{1}{2}x^2\right)$, and integrate over x from $-\infty$ to ∞. The left-hand side becomes a Gaussian integral that can be evaluated. The result is

$$(2\pi)^{1/2} \, e^{-k^2/2} \, e^{-2ikt + t^2} = \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} dx \, e^{-ikx} \, e^{-x^2/2} \, H_n(x) \, \frac{t^n}{n!} = \sum_{n=0}^{\infty} \mathcal{F}\left[e^{-x^2/2} \, H_n(x)\right] \frac{t^n}{n!} \, .$$

But the second exponential factor on the left-hand side is again a generating function for Hermite polynomials. We have

$$e^{-2ikt + t^2} = e^{2k(-it) + (-it)^2} = \sum_{n=0}^{\infty} H_n(k) \, \frac{(-it)^n}{n!} \, .$$

Equating the coefficients of t^n of the two absolutely convergent power series in t, we finally get

$$\int_{-\infty}^{\infty} dx \, e^{-ikx} \, e^{-x^2/2} \, H_n(x) = \mathcal{F}\left[e^{-x^2/2} \, H_n(x)\right] = (2\pi)^{1/2} \, (-i)^n \left(e^{-k^2/2} \, H_n(k)\right).$$

Hence the function $e^{-x^2/2} \, H_n(x)$, or this function multiplied by the normalization constant $(2^n \, n! \, \sqrt{\pi} \,)^{-1/2}$, is an eigenfunction of the Fourier transform operator $\mathcal{F}$ in the space $\mathcal{L}_2(-\infty \, , \, \infty)$. Here n runs over the values 0, 1, 2, ... The corresponding eigenvalue is $(2\pi)^{1/2} \, (-i)^n$.

- The eigenvalues of the operator $(2\pi)^{-1/2} \, \mathcal{F}$ are $(-i)^n$, where $n = 0, 1, 2, \ldots$.

6. Work through the steps of this derivation.

Unitarity of the Fourier transformation: Since $(-i)^n$ is always equal to 1, $-i$, -1 or i, the operator $(2\pi)^{-1/2}\,\mathcal{F}$ has just these four distinct eigenvalues. Each of these is infinitely degenerate, i.e., there is a countable infinity of linearly independent eigenfunctions corresponding to each eigenvalue. These eigenfunctions are, respectively,

$$\Phi_{4n}(x),\ \Phi_{4n+1}(x),\ \Phi_{4n+2}(x) \text{ and } \Phi_{4n+3}(x) \quad (n = 0,\, 1,\, 2,\, \ldots),$$

where $\Phi_n(x)$ is as given above.

As I have mentioned already, the factors of $(2\pi)^{1/2}$ in the eigenvalues of $\mathcal{F}$ arises from the Fourier transform convention we have adopted. Let me repeat this for clarity: we have defined

$$[\mathcal{F}\,f](k) = \widetilde{f}(k) = \int_{-\infty}^{\infty} dx\, e^{-ikx}\, f(x), \quad [\mathcal{F}^{-1}\,\widetilde{f}\,](x) = f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{ikx}\,\widetilde{f}(k).$$

Had we used the alternative (and more symmetrical) Fourier transform convention

$$[\mathcal{F}\,f](k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx\, e^{-ikx}\, f(x), \quad [\mathcal{F}^{-1}\,\widetilde{f}](x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\, e^{ikx}\,\widetilde{f}(k),$$

we would have found that $\mathcal{F}^4 = I$, and the eigenvalues of $\mathcal{F}$ would have been simply ± 1 and $\pm i$. Indeed, the customary normalization of quantum mechanical wave functions uses precisely this convention. But the convention chosen here is used in many other applications of Fourier analysis, and so let's stay with it.

Taking the Fourier transform of a function involves applying an integral operator to the function. What is the adjoint of this operator? Let us digress for a moment to find the adjoint of a general integral operator.

The adjoint of an integral operator: Let $f(x)$ be the function representing the vector $|\,f\,\rangle$ in the given function space (here, $\mathcal{L}_2(-\infty\,,\,\infty)$). If K is the integral operator concerned, the vector $\mathsf{K}\,|\,f\,\rangle$ is represented by the function $\int_{-\infty}^{\infty} dy\, K(x,y)\, f(y)$, where $K(x,y)$ is the kernel of the operator. The adjoint of K, denoted by $\mathsf{K}^\dagger$, is of course to be identified by applying the condition $\langle g\,|\,\mathsf{K}f\rangle = \langle \mathsf{K}^\dagger g\,|\,f\,\rangle$ for every pair of elements

$|\, f\,\rangle$, $|\, g\,\rangle$ in the function space. But

$$\begin{aligned}\langle g \,|\, \mathsf{K} f\rangle &= \int_{-\infty}^{\infty} dx\, g^*(x) \int_{-\infty}^{\infty} dy\, K(x,y)\, f(y) \\ &= \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx\, g^*(x)\, K(x,y)\, f(y) \;\; \text{(changing the order of integration)} \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, K(y,x)\, g^*(y)\, f(x) \;\; \text{(re-labeling } x \leftrightarrow y\text{)} \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, [K^*(y,x)\, g(y)]^*\, f(x) = \langle \mathsf{K}^\dagger g \,|\, f\,\rangle.\end{aligned}$$

It follows that the kernel of the integral operator $\mathsf{K}^\dagger$ is $K^*(y,x)$.[32]

Returning to the case at hand, we see that $(2\pi)^{-1/2}\,\mathcal{F}$ has the kernel $K(k,x) = (2\pi)^{-1/2}\, e^{-ikx}$, while its inverse $[(2\pi)^{-1/2}\,\mathcal{F}]^{-1}$ has the kernel $(2\pi)^{-1/2}\, e^{ikx}$. But the latter quantity is precisely $K^*(x,k)$. We therefore have the operator identity

$$[(2\pi)^{-1/2}\,\mathcal{F}]^{-1} = [(2\pi)^{-1/2}\,\mathcal{F}]^\dagger.$$

In other words, the operator $(2\pi)^{-1/2}\,\mathcal{F}$ is a *unitary* operator. It is not surprising, then, that all its eigenvalues lie on the unit circle in the complex plane, just as those of a unitary matrix do!

Hence the Fourier transform in $\mathcal{L}_2(-\infty\,,\,\infty)$ is not only a change of basis in the space, but also a **unitary transformation**. This fact has implications in quantum mechanics. For instance, it guarantees that the description of a particle (or a system of particles) in position space and in smomentum space are *unitarily equivalent*—i.e., you can use either description without altering the underlying physics.

The Fourier transform in d dimensions: The Fourier transform and most of the results in the foregoing are generalized in a straightforward manner to functions of $\mathbf{r} \in \mathbb{R}^d$, where $d = 2, 3, \ldots$. We have

$$f(\mathbf{r}) = \frac{1}{(2\pi)^d} \int d^d k\, e^{i\mathbf{k}\cdot\mathbf{r}}\, \widetilde{f}(\mathbf{k}) \quad \Longleftrightarrow \quad \widetilde{f}(\mathbf{k}) = \int d^d r\, e^{-i\mathbf{k}\cdot\mathbf{r}}\, f(\mathbf{r}).$$

Here $d^d k$ and $d^d r$ denote the volume elements in d-dimensional $\mathbf{k}$-space and $\mathbf{r}$-space, respectively. The Fourier representation of the d-dimensional δ-function is

$$\delta^{(d)}(\mathbf{r}) = \frac{1}{(2\pi)^d} \int d^d k\, e^{i\,\mathbf{k}\cdot\mathbf{r}}\,.$$

[32] Observe how reminiscent this is of the situation in the case of matrices. If the $(ij)^{\text{th}}$ element of a matrix A is a_{ij}, the $(ij)^{\text{th}}$ element of its hermitian adjoint $A^\dagger$ is a^*_{ji}. The close analogy between integral operators and matrices (as operators in an LVS) plays a crucial role in the theory of integral equations.

Fourier expansions of vector-valued functions can also be written down in an analogous manner. Thus, if $\mathbf{u}(\mathbf{r})$ is a vector field in $\mathbb{R}^d$, we have

$$\mathbf{u}(\mathbf{r}) = \frac{1}{(2\pi)^d} \int d^d k \, e^{i\mathbf{k}\cdot\mathbf{r}} \, \widetilde{\mathbf{u}}(\mathbf{k}) \quad \Longleftrightarrow \quad \widetilde{\mathbf{u}}(\mathbf{k}) = \int d^d r \, e^{-i\mathbf{k}\cdot\mathbf{r}} \, \mathbf{u}(\mathbf{r}) \, .$$

One of the great advantages of the Fourier expansion of functions now becomes evident. Consider the usual case, $d = 3$. A vector field is generally specified by differential equations for its divergence and curl, respectively.[33] We have

$$\nabla \cdot \mathbf{u}(\mathbf{r}) = \frac{1}{(2\pi)^3} \int d^3 k \, e^{i\mathbf{k}\cdot\mathbf{r}} \, i\mathbf{k} \cdot \widetilde{\mathbf{u}}(\mathbf{k})$$

and

$$\nabla \times \mathbf{u}(\mathbf{r}) = \frac{1}{(2\pi)^3} \int d^3 k \, e^{i\mathbf{k}\cdot\mathbf{r}} \, i\mathbf{k} \times \widetilde{\mathbf{u}}(\mathbf{k}).$$

Further, the set of functions $\left\{ e^{i\mathbf{k}\cdot\mathbf{r}} \,|\, \mathbf{k} \in \mathbb{R}^3 \right\}$ forms an orthonormal basis for integrable functions of $\mathbf{r} \in \mathbb{R}^3$. Hence *partial differential equations* for the divergence and curl of a vector field $\mathbf{u}(\mathbf{r})$ reduce to *algebraic equations* for its Fourier transform $\widetilde{\mathbf{u}}(\mathbf{k})$. The utility of this result should now be obvious to you.

The Poisson summation formula is a very useful result that has many applications. For instance, it can be used to sum many infinite series. It is also related to much deeper mathematical results (which we do not go into here).

Consider the infinite periodic array of δ-functions (a Dirac comb) given by

$$F(x) = \sum_{n=-\infty}^{\infty} \delta \, (x - nL),$$

where L is a positive constant. It is trivial to verify that $F(x) = F(-x)$. If we regard $F(x)$ as a periodic function of x with fundamental interval $(-\frac{1}{2}L, \, , \frac{1}{2}L)$, we have $F(x) = \delta(x)$ in the fundamental interval. It can then be expanded in a Fourier series according to

$$F(x) = \frac{1}{L} \sum_{n=-\infty}^{\infty} F_n \, e^{2\pi n i x/L} \, , \quad \text{where} \quad F_n = \int_{-\frac{1}{2}L}^{\frac{1}{2}L} e^{-2\pi n i x/L} \, \delta \, (x) = 1.$$

Therefore $F(x)$ can be written in two different ways—either as an infinite sum of exponentials, or as an infinite sum of δ-functions:

$$F(x) = \sum_{n=-\infty}^{\infty} \delta \, (x - nL) = \frac{1}{L} \sum_{n=-\infty}^{\infty} e^{2\pi n i x/L}.$$

[33] Maxwell's equations for electromagnetic fields provide a prominent example of this statement.

But

$$\delta\left(x-nL\right)=\delta\left(nL-x\right)=\frac{1}{L}\,\delta\,\left(n-\frac{x}{L}\right)\,.$$

We thus have a useful identity relating a sum of equally-spaced δ-functions to a sum of exponentials, namely,

$$\sum_{n=-\infty}^{\infty}\delta\,\left(n-\frac{x}{L}\right)=\sum_{n=-\infty}^{\infty}\exp\,\left(\frac{2\pi nix}{L}\right).$$

It is worth pausing once again for a moment to think about this most remarkable formula. Note that neither side of this equation is anything like an absolutely convergent series!

Next, consider the Fourier *transform* of $F(x)$. This is

$$\begin{aligned}\widetilde{F}(k) &= \int_{-\infty}^{\infty}dx\,e^{-ikx}\,F(x)\\ &= \sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}dx\,e^{-ikx}\,\delta\left(x-nL\right)=\sum_{n=-\infty}^{\infty}e^{inkL}.\end{aligned}$$

But, as we have seen already, an infinite sum of exponentials of this kind can be written as an infinite sum of δ-functions. Using this relationship, we get

$$\widetilde{F}(k)=\sum_{n=-\infty}^{\infty}\delta\,\left(n-\frac{kL}{2\pi}\right).$$

The final step is to insert the expressions for $F(x)$ and $\widetilde{F}(k)$ in the identity[34]

$$\int_{-\infty}^{\infty}dx\,f(x)\,F(-x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}dk\,\widetilde{f}(k)\,\widetilde{F}(k).$$

Using the fact that the function $F(x)$ in the present instance is an even function of x, we get

$$\int_{-\infty}^{\infty}dx\,f(x)\,\sum_{n=-\infty}^{\infty}\delta\left(x-nL\right)=\frac{1}{2\pi}\int_{-\infty}^{\infty}dk\,\widetilde{f}(k)\,\sum_{n=-\infty}^{\infty}\delta\,\left(n-\frac{kL}{2\pi}\right),$$

where $f(x)$ is any *arbitrary* function with a Fourier transform $\widetilde{f}(k)$. Finally, therefore,

$$\sum_{n=-\infty}^{\infty}f(nL)=(1/L)\,\sum_{n=-\infty}^{\infty}\widetilde{f}\,(2\pi n/L)\,.$$

[34]The identity was derived for *good* functions, but we're using it here for singular functions like the δ-function. I'll merely state here without proof that this step can be justified.

Here L is an arbitrary positive parameter. This is the famous **Poisson summation formula** It is helpful in the summation of series in cases when the left-hand side is difficult to evaluate, but the right-hand side is more tractable; or *vice versa*. Note that the parameter L occurs in the *numerator* of the argument of the function on the left-hand side, but on the right-hand side it occurs in the *denominator*. Hence one of the two representations may be useful for *small* values of L, while the other is useful for *large* values of L, in physical applications.

Some illustrative examples: Here are some instances of the use of Poisson's summation formula to derive useful identities and to sum certain infinite series.

7. If $f(x)$ is chosen to be a Gaussian function, the formula immediately gives

$$\sum_{n=-\infty}^{\infty} \exp\,(-\pi\, n^2\, \lambda^2) = \frac{1}{\lambda} \sum_{n=-\infty}^{\infty} \exp\,(-\pi\, n^2/\lambda^2)\,, \quad (\lambda > 0).$$

Establish this result using the Poisson summation formula.

This identity, due to Jacobi, is so useful and important that it is sometimes called the Poisson summation formula itself![35] For example, in the context of the phenomenon of diffusion, the parameter λ^2 is proportional to Dt, where D is the diffusion coefficient and t is the time. The Poisson summation formula can then be used to obtain valuable insight into the behavior of the solution to the diffusion equation for both small t and large t.

Another useful case is that of the function

$$f(x) = \frac{\cos\, bx}{x^2 + a^2} \quad (a > 0,\ 0 \le b < 2\pi).$$

The Fourier transform of this function is easily found by contour integration. You can, however, write down $\widetilde{f}(k)$ using the information that is already available. Recall that the inverse Fourier transform of $e^{-\lambda|k|}$ is a Lorentzian. That is,

$$f(x) = \frac{\lambda}{\pi(x^2 + \lambda^2)} \quad \Longleftrightarrow \quad \widetilde{f}(k) = e^{-\lambda\,|k|}.$$

Hence the Fourier transform of the function $f(x)$ is given by

$$\begin{aligned} \widetilde{f}(k) &= \int_{-\infty}^{\infty} dx\, \frac{\cos\, bx}{x^2 + a^2}\, e^{-ikx} = \frac{1}{2} \int_{-\infty}^{\infty} dx\, \frac{\left[e^{-i(k-b)x} + e^{-i(k+b)x}\right]}{x^2 + a^2} \\ &= \frac{\pi}{2a} \left[e^{-a\,|k-b|} + e^{-a\,|k+b|}\right]. \end{aligned}$$

[35] It is an example of a **Gaussian sum**, i.e., a sum over the exponentials of the *squares* of the integers. A vast literature exists on this subject, that has ramifications in several parts of mathematics and mathematical physics.

Setting $L = 1$ in the Poisson summation formula, we therefore have

$$\sum_{n=-\infty}^{\infty} \frac{\cos(nb)}{n^2 + a^2} = \sum_{n=-\infty}^{\infty} \frac{\pi}{2a} \left[e^{-a\,|2\pi n - b|} + e^{-a\,|2\pi n + b|} \right].$$

Some simplification then yields the identity

$$\sum_{n=1}^{\infty} \frac{\cos(nb)}{n^2 + a^2} = \frac{\pi}{2a} \left\{ \frac{\cosh\, a(\pi - b)}{\sinh(\pi a)} - \frac{1}{\pi a} \right\}.$$

Setting $b = 0$ in this relation, we recover an important and useful result that we have already derived using contour integration, namely,

$$S(a) \equiv \sum_{n=1}^{\infty} \frac{1}{n^2 + a^2} = \frac{\pi}{2a} \left\{ \coth\, \pi a - \frac{1}{\pi a} \right\}.$$

As usual, a relation of this sort may be regarded as a relation between two analytic functions—in this case, of the variable a. Recall that the quantity in curly brackets on the right-hand side is the **Langevin function**. It occurs, for instance, in the elementary theory of paramagnetism. Passing to the limit $a \to 0$ once again gives us $\sum_{n=1}^{\infty} 1/n^2 \equiv \zeta(2) = \pi^2/6$. Repeatedly differentiating $S(a)$ with respect to a and passing to the limit $a \to 0$ enables us to find $\zeta(4) = \pi^4/90$, $\zeta(6) = \pi^6/945$, ...

Generalization to higher dimensions: The Poisson summation formula is readily generalized to higher dimensions. If $\widetilde{f}(\mathbf{k})$ is the Fourier transform of $f(\mathbf{r})$ where $\mathbf{r} \in \mathbb{R}^d$, then

$$\sum_{\mathbf{n} \in \mathbb{Z}^d} f(\mathbf{n}L) = (1/L^d) \sum_{\mathbf{n} \in \mathbb{Z}^d} \widetilde{f}\,(2\pi \mathbf{n}/L)\ .$$

Here $\mathbb{Z}^d$ is the set of d-tuples of integers, in an obvious notation. Further generalizations are also possible, such as the counterpart of the formula above for $f(\mathbf{n}L + \mathbf{r}_0)$, where $\mathbf{r}_0$ is any given vector in $\mathbb{R}^d$.

Finally, it turns out that the generalizations of the Poisson summation formula are related to deep results in advanced mathematics. I merely mention some of these results, to whet your curiosity: the asymptotic behavior of the heat kernel, the spectra of the Laplacian operator and general elliptic operators on manifolds, Eisenstein series, automorphic functions, the Selberg trace formula, and so on.

9 QUIZ 1

1. Are the statements in quotation marks true or false?

   (a) "Every derivative of an analytic function of a complex variable is also an analytic function."

   (b) Let u and v denote the real and imaginary parts of an analytic function of $z = x + iy$.
   "The curves $u(x, y) =$ constant and $v(x, y) =$ constant intersect each other at right angles."

   (c) "An entire function must necessarily be singular at $z = \infty$, unless it is just a constant."

   (d) "A meromorphic function cannot have an essential singularity at the point at infinity."

   (e) "The radius of convergence of the power series $\sum_{n=1}^{\infty} n^{1/n}\, z^n$ is zero."

   (f) "The function $\sin(\pi/z)$ has an accumulation point of poles at $z = 0$."

   (g) "The relation $\Gamma(z)\,\Gamma(1 - z) = \pi \operatorname{cosec} \pi z$ is only valid in the region $0 < \operatorname{Re} z < 1$."

   (h) Let a be a positive constant.
   "$f(z) = \int_a^{\infty} dt\, t^z\, e^{-t}$ is an entire function of z."

   (i) "The power series $\sum_{n=1}^{\infty} z^n/n^4$ is absolutely convergent at all points inside and on the unit circle $|z| = 1$."

   (j) "The series $\displaystyle\sum_{n=0}^{\infty} (n + 1)^{z-1}$ converges in the region $\operatorname{Re} z > 0$."

   (k) Consider the Möbius transfomation $z \mapsto w = (2z + \sqrt{3}\,)/(\sqrt{3}\,z + 2)$.
   "This is a hyperbolic Möbius transformation."

   (l) Consider $z \mapsto w = (2z + \sqrt{3}\,)/(\sqrt{3}\,z + 2)$ once again.
   "The transformation maps the circle $|z| = 1$ to the circle $|w| = 1$."

(m) Legendre's differential equation is $(1-z^2)\phi'' - 2z\phi' + \nu(\nu+1)\phi = 0$.
"Since this equation is invariant under the interchange $\nu \leftrightarrow -\nu - 1$, *all* its solutions must also be invariant under this interchange."

(n) "The function $f(t) = e^{t^{3/2}}$, where $t \geq 0$, has no Laplace transform."

(o) "If $[\mathcal{L}f](s) = \int_0^\infty dt\, e^{-st}\, f(t)$, then $[\mathcal{L}^2 f](s) = \int_0^\infty dt\, f(t)/(s+t)$."

(p) "The product $\Gamma(z)\,\zeta(z)$ tends to a finite, nonzero limit as $z \to -2n$, where $n = 1, 2, \ldots$."

(q) Let $\phi(t)$ be a linear, causal, retarded response function, and $\widetilde{\phi}(s)$ its Laplace transform.
"The corresponding dynamic susceptibility $\chi(\omega)$ is the analytic continuation of $\widetilde{\phi}(s)$ to $s = -i\omega$."

(r) "The logarithmic derivative of the Riemann zeta function, $\zeta'(z)/\zeta(z)$, has a simple pole at $z = 1$ with residue equal to -1."

(s) "The only pole of the logarithmic derivative of the Riemann zeta function is at $z = 1$."

(t) Consider the function space $\mathcal{L}_2(-\infty\,,\,\infty)$.
"Any eigenfunction of the Fourier transform operator is also an eigenfunction of the parity operator, but the converse is not necessarily true."

2. Fill in the blanks in the following.

(a) The real part of an entire function $f(z)$ is given by $u(x,y) = (\cosh\, x)\,(\cos\, y)$. Hence the function is $f(z) = \cdots$.

(b) $\coth\, z$ is a periodic function of z, with a period equal to $\cdots$.

(c) The singularity of the polynomial $p(z) = z^n + a_1\, z^{n-1} + \cdots + a_n$ at $z = \infty$ is $\cdots$. (Select one from the following: (i) a removable singularity (ii) a simple pole (iii) a pole of order n (iv) an essential singularity.)

(d) The residue at $z = \infty$ of the polynomial $p(z) = z^n + a_1\, z^{n-1} + \cdots + a_n$ is $\operatorname*{Res}_{z=\infty} p(z) = \cdots$.

(e) Let C denote the circle $|z| = 2$ traversed once in the positive sense. Then $\oint_C dz/(z^4 - 1) = \cdots$.

(f) Let a and b be two different complex numbers, each with nonzero real and imaginary parts. The radius of convergence of the power series

$$f(z) = \sum_{n=0}^{\infty} \frac{\Gamma(n+a)}{\Gamma(a)} \frac{\Gamma(b)}{\Gamma(n+b)} \frac{z^n}{n!}$$

is $R = \cdots$.

(g) Given that $\sum\limits_{n=1}^{\infty} 1/n^4 = \pi^4/90$, it follows that $\sum\limits_{n=0}^{\infty} 1/(2n+1)^4 = \cdots$.

(h) Given that $\int_0^\infty dx\, (\sin kx)/x = \frac{1}{2}\pi$ (where $k > 0$), the value of the integral $\int_0^\infty dx\, (1 - \cos x)/x^2 = \cdots$. (*Hint*: Integrate k over a suitable range.)

(i) The numerical value of the product

$$\Gamma\left(-\tfrac{5}{4}\right) \Gamma\left(-\tfrac{3}{4}\right) \Gamma\left(-\tfrac{1}{4}\right) \Gamma\left(\tfrac{1}{4}\right) \Gamma\left(\tfrac{3}{4}\right) \Gamma\left(\tfrac{5}{4}\right) \Gamma\left(\tfrac{7}{4}\right) \Gamma\left(\tfrac{9}{4}\right) = \cdots .$$

(j) The value of the integral $\int_0^1 dt\, t^{-1/2}\, (1-t)^{-1/2} = \cdots$.

(k) Let an arbitrary initial point $z^{(0)}$ in the complex plane be mapped to the point $z^{(n)}$ under n iterations of the Möbius transfomation

$$z \mapsto (2z + \sqrt{3}\,)/(\sqrt{3}\, z + 2).$$

As $n \to \infty$, $z^{(n)} \to \cdots$ for all $z^{(0)}$, with *one* exception.

(l) Previous question continued: The exceptional point that does *not* tend to the limit point above is $z = \cdots$.

(m) Under the Möbius transformation $z \mapsto w = (2z + 3)/(z + 2)$, the circle $|z + 2| = 1$ is mapped to the circle $\cdots$.

(n) The function $f(z) = \sqrt{z}\,\ln\left[(z-1)/(z+1)\right]$ has branch points at $z = \cdots$.

(o) Let α and β be arbitrary complex numbers. The function

$$f(z) = (z^2-1)^\alpha/(z^2+1)^\beta$$

has branch points at $z = \cdots$.

(p) The residue of $f(z) = \exp\left(z + z^{-1}\right)$ at $z = 0$ is $\cdots$. (Express your answer in terms of a modified Bessel function.)

(q) Given that the Laplace transform of $\sin t$ is $1/(s^2+1)$, it follows that the Laplace transform of $\sinh t$ is $\cdots$.

(r) The generating function for the Hermite polynomial $H_n(z)$ is

$$e^{2tz-t^2} = \sum_{n=0}^{\infty} H_n(z)\,\frac{t^n}{n!}\,.$$

It follows that the Rodrigues formula for $H_n(z)$ is $H_n(z) = \cdots$.

(s) Let

$$f(x) = \begin{cases} 1, & |x| \le 1 \\ 0 & |x| > 1. \end{cases}$$

If $\widetilde{f}(k)$ denotes the Fourier transform of $f(x)$, the value of the integral $\int_{-\infty}^{\infty} dk\ |\widetilde{f}(k)|^2 = \cdots$.

(t) Consider a random walk on an infinite linear lattice whose sites are labelled by the integers. The walker jumps from any site j to $j-1$ with a probability per unit time given by λq, and from j to $j+1$ with a probability per unit time given by λp; further, the walker stays at the site j with probability per unit time given by λr. Here p, q and r are positive constants satisfying $p+q+r = 1$, and λ is a positive constant with the physical dimensions of $(\text{time})^{-1}$. Let $P(j,t)$ be the probability that the walker is at the site j at time t. The differential equation satisfied by $P(j,t)$ is $dP(j,t)/dt = \cdots$.

Quiz 1: Solutions

1. (a) **T**
   (b) **T**
   (c) **T**
   (d) **F**
   (e) **F**
   (f) **F**
   (g) **F**
   (h) **T**
   (i) **T**
   (j) **F**
   (k) **T**
   (l) **T**
   (m) **F**
   (n) **T**
   (o) **T**
   (p) **T**
   (q) **T**
   (r) **T**
   (s) **F**
   (t) **T**

2. (a) $f(z) = \cosh z$
   (b) $i\pi$
   (c) a pole of order n
   (d) 0
   (e) 0
   (f) ∞
   (g) $\pi^4/96$
   (h) $\frac{1}{2}\pi$
   (i) $4\pi^4$
   (j) π

(k) 1

(l) -1

(m) $|w-2|=1$

(n) $-1, 0, 1$ and ∞

(o) $1, i, -1, -i$ and ∞

(p) $I_1(2)$

(q) $1/(s^2-1)$

(r) $H_n(z) = \left[\dfrac{d^n}{dt^n} e^{2tz-t^2}\right]_{t=0}$, which simplifies to $H_n(z) = (-1)^n\, e^{z^2}\, \dfrac{d^n}{dz^n} e^{-z^2}$.

(s) 4π

(t) $\dfrac{dP(j,t)}{dt} = \lambda\Big[p\,P(j-1,t) + q\,P(j+1,t) - (p+q)\;P(j,t)\Big]$

10 The fundamental Green function for ∇^2

The Laplacian operator ∇^2 appears in all the standard and important linear partial differential equations of elementary mathematical physics: Laplace's equation, Poisson's equation, Helmholtz's equation, the diffusion or heat equation, and the wave equation. We'll now consider Poisson's equation, which requires the determination of the Green function of the Laplacian operator. Let's start with a brief recapitulation of the essentials of the Green function method.

Green functions: Consider an inhomogeneous, linear, ordinary differential equation of the form

$$\mathcal{D}_x\, f(x) = g(x)$$

in some interval (a, b) in the real variable x. Here $\mathcal{D}_x$ is a differential operator involving derivatives of various orders with respect to x, and $g(x)$ is a given function. It is required to find $f(x)$. Let's write the equation in abstract form in terms of the elements $|\, f\,\rangle$ and $|\, g\,\rangle$ of a linear space that are represented by $f(x)$ and $g(x)$ in a suitable function space. Recall that $\langle x\,|\, f\,\rangle \equiv f(x)$ and $\langle x\,|\, g\,\rangle \equiv g(x)$. Let D denote the abstract operator that is represented by $\mathcal{D}_x$ in function space. Then

$$\mathsf{D}\,|\, f\,\rangle = |\, g\,\rangle.$$

The formal, general solution of this equation is given by

$$|\, f\,\rangle = \mathsf{D}^{-1}\,|\, g\,\rangle + \sum_i c_i\,|h_i\rangle,$$

where the c_i are constants, and the kets $|h_i\rangle$ are the linearly independent solutions of the *homogeneous* equation

$$\mathsf{D}\,|h_i\rangle = 0 \quad \text{or} \quad \mathcal{D}_x\, h_i(x) = 0.$$

In function space, the inverse of the operator D is represented by the inverse of the differential operator $\mathcal{D}_x$. In general, the latter is an integral operator. To see this, take the scalar product of both sides of the formal solution for $|\, f\,\rangle$ above with $\langle\, x\,|$. We have

$$\begin{aligned}
\langle x\,|\, f\,\rangle \equiv f(x) &= \langle\, x\,|\mathsf{D}^{-1}\,|\, g\,\rangle + \sum_i c_i\,\langle\, x\,|h_i\rangle \\
&= \int_a^b dx'\,\langle\, x\,|\mathsf{D}^{-1}|\, x'\,\rangle\langle x'|\, g\,\rangle + \sum_i c_i\, h_i(x) \\
&= \int_a^b dx'\, G(x, x')\, g(x') + \sum_i c_i\, h_i(x),
\end{aligned}$$

where

$$G(x, x') \equiv \langle\, x \,|\mathsf{D}^{-1}|x'\rangle.$$

$G(x, x')$ is the **Green function** of the differential operator $\mathcal{D}_x$. It is just the 'matrix element' of the operator D^{-1} between the states $\langle\, x \,|$ and $|\, x' \,\rangle$, and is called the **kernel** of the integral operator D^{-1}.

We learn in elementary treatments of differential equations that the general solution of an inhomogeneous differential equation is made up of two parts: a **particular integral** (PI) that depends on $g(x)$, and a **complementary function** (CF) that does not. The first term on the right-hand side in the solution above is the PI, while the second is the CF. The right combination of the two is determined by fixing the values of the constants c_i using the boundary conditions.

The fact that $G(x, x')$ represents (in function space) the inverse of the abstract operator D^{-1} means that it, too, satisfies the same differential equation as $f(x)$, but with a δ-function as the inhomogeneous term. To see this, we start with

$$\mathsf{D}\,\mathsf{D}^{-1} = I \quad\Longrightarrow\quad \langle\, x \,|\mathsf{D}\,\mathsf{D}^{-1}|\, x' \,\rangle = \langle\, x \,|I|\, x' \,\rangle = \delta(x - x').$$

But

$$\begin{aligned}
\langle\, x \,|\mathsf{D}\,\mathsf{D}^{-1}|\, x' \,\rangle &= \langle\, x \,|\mathsf{D} \int_a^b dy\, |\, y \,\rangle\langle\, y \,|\, \mathsf{D}^{-1}|\, x' \,\rangle \\
&= \int_a^b dy\, \langle\, x \,|\mathsf{D}|\, y \,\rangle G(y, x') = \int dy\, \mathcal{D}_x\, \langle x|\, y \,\rangle G(y, x') \\
&= \int dy\, \mathcal{D}_x\, \delta(x - y) G(y, x') = \mathcal{D}_x\, G(x, x').
\end{aligned}$$

Therefore the Green function satisfies the differential equation

$$\mathcal{D}_x\, G(x, x') = \delta(x - x').$$

In order to write down the solution for $f(x)$ explicitly, we need to find the Green function $G(x, x')$. This must be done by solving the differential equation for G, using appropriate boundary conditions and adjusting the values of the constants c_i. In this manner, we arrive at the unique solution that satisfies the given boundary conditions on $f(x)$.

1. Here's a simple example that shows how the Green function method works. Consider the ordinary differential equation

$$\frac{d^2}{dx^2} f(x) = g(x),$$

where $x \in [0,1]$, and $g(x)$ is a given function of x. It is required to find the solution $f(x)$ that satisfies the general linear boundary conditions

$$f(0) + a f'(0) = b \quad \text{and} \quad f(1) + c f'(1) = d,$$

where a, b, c and d are given constants. Show that

$$G(x,x') = \begin{cases} A_1\, x + A_2 & \text{for} \quad 0 \le x < x' \\ A_3\, x + A_4 & \text{for} \quad x' < x \le 1, \end{cases}$$

where

$$\left.\begin{aligned} A_1 &= \frac{x' - 1 - b - c + d}{1 - a + c}, \quad A_2 = \frac{-a x' + (a+b)(1+c) - ad}{1 - a + c} \\ A_3 &= \frac{x' - a - b + d}{1 - a + c}, \qquad A_4 = \frac{-(1+c)x' + (a+b)(1+c) - ad}{1 - a + c}. \end{aligned}\right\}$$

Hence write down the solution for $f(x)$.

In essence, this is the Green function method. All that has been said above can be generalized to linear ordinary differential equations in a complex variable z, and further, to the case of linear partial differential equations in several variables—e.g., when x is replaced by $\mathbf{r}$ in any number d of spatial dimensions, or by $(\mathbf{r}, t)$ in $(d+1)$ dimensions.

Poisson's equation; the fundamental Green function for ∇^2: Let's turn now to a problem of great physical importance: the determination of the Green function of the Laplacian operator, required for the solution of Poisson's equation. In effect, we seek the inverse of the ∇^2 operator. To start with, consider the standard three-dimensional case. Subsequently, we'll go on to the cases $d = 2$ and $d \ge 4$.

As you know, the electrostatic potential $\phi(\mathbf{r})$ in the presence of a given static charge density $\rho(\mathbf{r})$ satisfies Poisson's equation, $\nabla^2 \phi(\mathbf{r}) = -\rho(\mathbf{r})/\epsilon_0$. The solution satisfying natural boundary conditions, namely, $\phi(\mathbf{r}) = 0$ as $r \to \infty$ along any direction in space, is $\phi(\mathbf{r}) = (4\pi\epsilon_0)^{-1} \int d^3r'\, \rho(\mathbf{r}')/|\mathbf{r} - \mathbf{r}'|$. This is just Coulomb's Law together with the superposition principle. This solution can be derived systematically, as follows.

Consider the general form of Poisson's equation together with natural boundary conditions, namely,

$$\nabla^2 f(\mathbf{r}) = g(\mathbf{r}), \quad \text{with} \quad f(\mathbf{r}) \to 0 \quad \text{as} \quad r \to \infty.$$

$g(\mathbf{r})$ is a given function that acts as the 'source' term for the scalar field $f(\mathbf{r})$. We're interested here in the PI, given by

$$f(\mathbf{r}) = \int d^3r'\, G(\mathbf{r}, \mathbf{r}')\, g(\mathbf{r}'),$$

where the Green function G satisfies the differential equation

$$\nabla^2_{\mathbf{r}}\, G(\mathbf{r}, \mathbf{r}\,') = \delta^{(3)}(\mathbf{r} - \mathbf{r}\,').$$

I've used a subscript on the gradient operator in order to indicate the variable with respect to which the differentiation is to be performed. We require the solution that vanishes as $r \to \infty$. Note that G may be interpreted (apart from a constant of proportionality) as the Coulomb potential due to a point charge at the source point $\mathbf{r}\,'$. Observe that

(i) the operator ∇^2 is translationally invariant—shifting $\mathbf{r}$ by $\mathbf{r}\,'$ does not change the operator;

(ii) the δ-function on the right-hand side is a function of the difference $\mathbf{r} - \mathbf{r}\,'$; and finally,

(iii) the boundary condition imposed involves $r \to \infty$, which is the same as $|\mathbf{r} - \mathbf{r}\,'| \to \infty$.

Together, these facts ensure that the solution $G(\mathbf{r}, \mathbf{r}\,')$ is a function of the difference $\mathbf{r} - \mathbf{r}\,'$ alone. Let's therefore set

$$\mathbf{R} = \mathbf{r} - \mathbf{r}\,',$$

and write the Green function as $G(\mathbf{R})$. Since $\mathbf{R}$ is just a shift of the variable $\mathbf{r}$, we have $\nabla^2_{\mathbf{r}} = \nabla^2_{\mathbf{R}}$. Then

$$\nabla^2_{\mathbf{R}}\, G(\mathbf{R}) = \delta^{(3)}(\mathbf{R}).$$

Now define the Fourier transform pair

$$G(\mathbf{R}) = \frac{1}{(2\pi)^3} \int d^3k\, e^{i\mathbf{k}\cdot\mathbf{R}}\, \widetilde{G}(\mathbf{k}) \quad \Longleftrightarrow \quad \widetilde{G}(\mathbf{k}) = \int d^3R\, e^{-i\mathbf{k}\cdot\mathbf{R}}\, G(\mathbf{R}).$$

The Fourier representation of the δ-function is

$$\delta^{(3)}(\mathbf{R}) = \frac{1}{(2\pi)^3} \int d^3k\, e^{i\mathbf{k}\cdot\mathbf{R}}.$$

Using the fact that $\nabla^2_{\mathbf{R}}\, (e^{i\mathbf{k}\cdot\mathbf{R}}) = -k^2\, (e^{i\mathbf{k}\cdot\mathbf{R}})$, we get

$$\widetilde{G}(\mathbf{k}) = -1/k^2.$$

Inverting the Fourier transform,

$$G(\mathbf{R}) = -\frac{1}{(2\pi)^3} \int d^3k\, \frac{e^{i\mathbf{k}\cdot\mathbf{R}}}{k^2}\,.$$

2. Use spherical polar coordinates to evaluate this integral. Since the integral is a scalar, it is rotationally invariant. You may choose the polar axis in $\mathbf{k}$-space along the

vector $\mathbf{R}$. You will need the Dirichlet integral $\int_0^\infty dk\,(\sin\,kR)/k = \frac{1}{2}\pi\,\mathrm{sgn}\,R = \frac{1}{2}\pi$. The final result is

$$G(\mathbf{r}-\mathbf{r}\,') = -\frac{1}{4\pi R} = -\frac{1}{4\pi|\mathbf{r}-\mathbf{r}\,'|}\,.$$

Note that the Green function is actually a function of R alone. The PI in the solution of Poisson's equation is then

$$f(\mathbf{r}) = -\,\frac{1}{4\pi}\int d^3r\,'\,\frac{g(\mathbf{r}\,')}{|\mathbf{r}-\mathbf{r}\,'|}\,.$$

3. Solution for a spherically symmetric source: When the source function $g(\mathbf{r}) = g(r)$, i.e.,when it is spherically symmetric, the solution above can be simplified further. For this purpose you need the expansion of the Coulomb kernel $1/|\mathbf{r}-\mathbf{r}\,'|$ in spherical harmonics. Let $\mathbf{r} = (r,\theta,\varphi)$ and $\mathbf{r}\,' = (r\,',\theta\,',\varphi\,')$, in spherical polar coordinates. Then

$$\frac{1}{|\mathbf{r}-\mathbf{r}\,'|} = \frac{1}{r_{\rm g}}\sum_{l=0}^{\infty}\frac{1}{(2l+1)}\left(\frac{r_{\rm s}}{r_{\rm g}}\right)^l\sum_{m=-l}^{l}Y_{lm}(\theta,\varphi)\,Y^*_{lm}(\theta\,',\varphi\,'),$$

where $r_{\rm s} = \min\,(r,r\,')$ and $r_{\rm g} = \max\,(r,r\,')$. Insert this expansion in the solution for $f(\mathbf{r})$, and and interchange the order of summation and integration. The integrations over the angular variables $\theta\,'$ and $\varphi\,'$ can then be carried out immediately by using the fact that

$$\int d\Omega\,'\;Y^*_{lm}(\theta\,',\,\varphi\,') = \sqrt{4\pi}\;\delta_{l,0}\,\delta_{m,0}\,.$$

Because of the Kronecker deltas, the dependence of $f(\mathbf{r})$ on the angles θ and φ also disappears, so that $f(\mathbf{r}) = f(r)$. Show that the solution simplifies to

$$f(r) = -\int_0^\infty dr\,'\,\frac{r\,'^{\,2}\,g(r\,')}{r_{\rm g}} = -\frac{1}{r}\int_0^r dr\,'\,r\,'^{\,2}\,g(r\,') - \int_r^\infty dr\,'\,r\,'\,g(r\,').$$

4. Check that the solution above satisfies Poisson's equation, $\nabla^2 f = g$. You will need the Laplacian in spherical polar coordinates, and also the formula for differentiation under the integral sign.

The Coulomb potential in d dimensions: The fundamental Green function of the ∇^2 operator in 3-dimensional Euclidean space, $-1/(4\pi|\mathbf{r}-\mathbf{r}\,'|)$, is essentially the Coulomb potential due to a point charge. This connection lends a deeper significance to the Coulomb potential, and hence to the inverse-square central force. As you know, the inverse-square force law leads directly to Gauss' Law in electrostatics: The flux of the electrostatic field over a closed surface is equal to the total charge enclosed by

the surface, apart from a multiplicative constant ($= 1/\epsilon_0$, in SI units). A counterpart obviously exists in gravitation as well. The inverse-square central force is the only force law in three-dimensional space for which such a property holds good. Interestingly enough, in spaces of other dimensionalities ($d = 2, 4, \ldots$) too, there exist force laws with an analogous property. They arise from the counterparts of the Coulomb potential in those spaces. The case $d = 2$ is somewhat exceptional, as you will see, and will be dealt with after we discuss the case $d \geq 3$.

What is the analog of the Coulomb potential (or the inverse-square central force) in a space of an arbitrary number of dimensions? The connection that I have just pointed out leads to the consistent way to *define* such a potential, via Poisson's equation for the potential due to a point source (or charge):

- The Coulomb potential in d-dimensional space is (apart from a constant of proportionality) the fundamental Green function of the Laplacian operator in d dimensions.

We therefore look for the fundamental solution of the equation

$$\nabla^2_{\mathbf{R}}\, G^{(d)}(\mathbf{R}) = \delta^{(d)}(\mathbf{R}), \quad \text{with} \quad G^{(d)}(\mathbf{R}) \underset{R\to\infty}{\longrightarrow} 0.$$

The superscript in $G^{(d)}$ is to remind us that we are concerned with the Green function in d-dimensional space. We'll find the solution in two different ways: first, by using Fourier transforms as in the $d = 3$ case derived above, and evaluating the resulting integral the 'hard' way; second, by using a much simpler argument that will get us to the same answer. This exercise is worth it because it is instructive.

Fourier-transforming the equation for $G^{(d)}(\mathbf{R})$ gives $\widetilde{G}^{(d)}(\mathbf{k}) = -1/k^2$, as before. Remember that we are now working in d dimensions. Therefore

$$G^{(d)}(\mathbf{R}) = -\frac{1}{(2\pi)^d} \int \frac{d^d k}{k^2}\, e^{i\mathbf{k}\cdot\mathbf{R}}.$$

That is, the Coulomb potential in d dimensions, i.e., the fundamental Green function of the Laplacian, is essentially the inverse Fourier transform of $-1/k^2$.

5. The next task is to evaluate the d-dimensional above for $G^{(d)}(\mathbf{R})$. Once again, This is most conveniently done in terms of the analog of spherical polar coordinates in d dimensions, in $\mathbf{k}$-space. These **ultraspherical coordinates** consist of (i) the magnitude $|\mathbf{k}| = k$, (ii) a set of $(d-2)$ 'polar' angles $\theta_1\,, \theta_2\,, \ldots\,, \theta_{d-2}\,$, and (iii) an 'azimuthal' angle φ. The ranges of these variables are given by

$$0 \leq k < \infty,\ 0 \leq \theta_j \leq \pi\ \ (1 \leq j \leq d-2),\ \text{ and }\ 0 \leq \varphi < 2\pi.$$

The Cartesian components k_1 , k_2 , ... , k_d are related to the polar coordinates by

$$\begin{aligned}
k_1 &= k\cos\theta_1 \\
k_2 &= k\sin\theta_1\cos\theta_2 \\
k_3 &= k\sin\theta_1\sin\theta_2\cos\theta_3 \\
\cdots &= \cdots\cdots\cdots\cdots\cdots\cdots \\
k_{d-1} &= k\sin\theta_1\sin\theta_2\ldots\sin\theta_{d-2}\cos\varphi \\
k_d &= k\sin\theta_1\sin\theta_2\ldots\sin\theta_{d-2}\sin\varphi.
\end{aligned}$$

The volume element d^dk is then given by

$$\begin{aligned}
d^dk &= dk_1\,dk_2\cdots dk_d \\
&= k^{d-1}\,(\sin^{d-2}\theta_1)\,(\sin^{d-3}\theta_2)\,\ldots\,(\sin\theta_{d-2})\,dk\,d\theta_1\,\ldots\,d\theta_{d-2}\,d\varphi.
\end{aligned}$$

We can choose the orientation of the axes such that $\mathbf{R}$ is along the direction of the first coordinate k_1 , so that $\mathbf{k}\cdot\mathbf{R} = kR\cos\theta_1$. The integration over φ gives a factor of 2π. Carry out the integration over each of the 'polar' angles θ_2 , ... , θ_{d-2} using the integral

$$\int_0^\pi d\theta\,\sin^r\theta = \frac{\sqrt{\pi}\,\Gamma\left(\frac{1}{2}(r+1)\right)}{\Gamma\left(1+\frac{1}{2}r\right)}\,.$$

But the integral over the last angular coordinate, θ_1 , is more complicated: It turns out to be essentially a representation of the Bessel functionof the first kind. The formula you require is

$$J_\nu(z) = \frac{\left(\frac{1}{2}z\right)^\nu}{\sqrt{\pi}\,\Gamma\left(\nu+\frac{1}{2}\right)}\int_0^\pi d\theta\,e^{\pm iz\cos\theta}\,\sin^{2\nu}\theta, \quad \mathrm{Re}\,\nu > -\tfrac{1}{2}.$$

Applying this formula,

$$\int_0^\pi d\theta_1\,e^{ikR\cos\theta_1}\,\sin^{d-2}\theta_1 = \sqrt{\pi}\;\Gamma\left(\tfrac{1}{2}(d-1)\right)(2/kR)^{(d-2)/2}\,J_{\frac{d}{2}-1}(kR).$$

Insert these results and simplify the expression obtained, to arrive at the following expression:

$$G^{(d)}(\mathbf{R}) = -\,\frac{1}{(2\pi)^{d/2}\,R^{(d-2)/2}}\,\int_0^\infty dk\,k^{(d-4)/2}\,J_{\frac{d}{2}-1}(kR).$$

Thus $G^{(d)}(\mathbf{R})$ has been reduced to a single integral. It is obvious from the last equation that the Green function is only a function of the magnitude R of the vector $\mathbf{R}$. Let's therefore write it as $G^{(d)}(R)$ from now on.

A divergence problem: A difficulty crops up now. The integral in the last equation above does not converge for arbitrary values of the dimensionality d. As you have seen, the case $d = 3$ already involves the Dirichlet integral $\int_0^\infty dk\,(\sin kR)/k$, which

is convergent, but not *absolutely* convergent. It is finite only because the sine function oscillates in sign, and there is some cancellation between positive and negative contributions to the integral. If the integrand is replaced by its *magnitude*, we find that $\int_0^\infty dk\,|\sin kR|/k$ diverges logarithmically owing to the slow ($\sim k^{-1}$) decay of the integrand as $k \to \infty$. For $d > 3$, you may expect the divergence of the relevant integral to get worse, essentially because the volume element $d^d k$ in d dimensions involves a factor k^{d-1}. This factor increases rapidly with k for larger values of d.

This sort of divergence, arising from the behavior of the integrand as $k \to \infty$, is an example of what is known as an **ultraviolet divergence** in physics, especially in the context of quantum field theory. The way to deal with such divergences in physical problems, and to extract meaningful results for physical quantities, is called **regularization**. There are many possible regularization methods. Here, I choose one that relies on the idea of analytic continuation, because you are already familiar with analytic continuation. It is called **dimensional regularization**. This is a powerful method of regularization, and is used quite commonly in modern quantum field theory.

Dimensional regularization: Consider, first, the conditions under which the integral

$$\int_0^\infty dk\, k^{(d-4)/2}\, J_{\frac{d}{2}-1}(kR)$$

actually converges. The argument is called **power-counting** and is quite simple in this instance. We must look at the behavior of the integrand at both the end-points of integration, and make sure that it is not too singular to be integrable at these points. Consider the upper limit first. As you know,

$$\int^\infty dk\, k^r < \infty \;\text{ provided } r < -1, \text{ or, more generally, } \;\operatorname{Re} r < -1.$$

Now, as $kR \to \infty$, the leading asymptotic behavior of the Bessel function $J_\nu(kR)$ is given by

$$J_\nu(kR) \sim (kR)^{-1/2},$$

independent of the order ν of the Bessel function.[36] The integrand therefore behaves like $k^{(d-5)/2}$ for large k. Hence the integral converges at the upper limit of integration only if

$$\tfrac{1}{2}(\operatorname{Re} d - 4) - \tfrac{1}{2} < -1, \quad \text{that is, if} \quad \operatorname{Re} d < 3.$$

This is precisely as expected: I have already pointed out that $d = 3$ is a marginal case in which the integral barely manages to converge to a finite value.

[36] The actual asymptotic form is $[2/(\pi\, kR)]^{1/2}$ times a cosine function whose magnitude does not exceed unity.

At the lower limit of integration, the situation is different. We know that

$$\int_0 dk\, k^r < \infty \text{ provided } \operatorname{Re} r > -1.$$

The leading behavior of the Bessel function $J_\nu(kR)$ as $kR \to 0$ is given by

$$J_\nu(kR) \sim (kR)^\nu.$$

The integrand therefore behaves like k^{d-3} in the neighborhood of $k = 0$, and the integral converges if

$$\tfrac{1}{2}(\operatorname{Re} d - 4) + \tfrac{1}{2}\operatorname{Re} d - 1 > -1, \quad \text{that is, if} \quad \operatorname{Re} d > 2.$$

For $\operatorname{Re} d \leq 2$, there is an **infrared divergence**. The terms 'ultraviolet divergence' and 'infrared divergence' originate in quantum field theory. The variable k is associated with momentum (recall that it is the Fourier conjugate of a position variable). Large k implies a small de Broglie wavelength ('ultraviolet'), while small k corresponds to a large de Broglie wavelength ('infrared').

Thus, the integral representing $G^{(d)}(R)$ converges in the range $2 < \operatorname{Re} d < 3$ of the parameter d. For other values of d, the integral is infinite. Let's now see how dimensional regularization helps us extract sensible results from the formally divergent integral we are faced with. The underlying idea is as follows.

(i) The dimension d itself is treated as a *complex variable*. In the region of the complex d-plane in which the integral converges, it defines an analytic function of d.

(ii) Analytic continuation is then used to go outside this region in the d-plane.

(iii) There are likely to be singularities (specifically, poles) present at the boundaries of the original region of convergence. If such a singularity occurs at a physical value of d, the *regular part* of the function (i.e., the function *with the singular part subtracted out*) is supposed to represent the value of the function at the point concerned.

Let's now see how this works in the case at hand.

6. First of all, *keeping d in the region* $2 < \operatorname{Re} d < 3$, we must evaluate the integral in the formula for $G^{(d)}(R)$. It turns out to be expressible in closed form, on using the following formula:

$$\int_0^\infty dx\, x^\mu\, J_\nu(ax) = \frac{2^\mu}{a^{\mu+1}} \frac{\Gamma\left(\frac{1}{2}(\nu + 1 + \mu)\right)}{\Gamma\left(\frac{1}{2}(\nu + 1 - \mu)\right)}, \quad \left(\operatorname{Re}\mu < -\tfrac{1}{2},\ \operatorname{Re}(\mu + \nu) > -1\right)$$

where a is a positive constant. First check that, when applied to the integral at hand, the conditions on the parameters μ and ν amount to precisely the restriction $2 < \mathrm{Re}\, d < 3$. Next, use the formula above to show that

$$G^{(d)}(R) = -\frac{\Gamma\left(\frac{1}{2}d - 1\right)}{4\pi^{d/2}\, R^{d-2}}\,.$$

But this expression for $G^{(d)}(R)$ is now in an explicit closed form that is analytic even to the right of $\mathrm{Re}\, d = 3$. Setting $d = 3$, it trivially checked that $G^{(3)}(R) = -1/(4\pi R)$, as we have found already. You can now proceed to set $d = 4, 5, \ldots$ in the final result to write down the fundamental Green function of the Laplacian in these dimensions. We conclude that:

- the fundamental Green function of the Laplacian (i.e., the Coulomb potential) in $d \geq 3$ spatial dimensions is proportional to $1/R^{d-2}$.

7. In this simple example, there is actually no singular part in the region $\mathrm{Re}\, d \geq 3$ that has to be extracted and discarded. The *ultraviolet* divergence in this problem is not a serious one. The *infrared* divergence, however, does exist. Recall that the leading (right-most) pole of the gamma function $\Gamma(z)$ is at $z = 0$. Therefore the expression found above for $G^{(d)}(R)$ has a pole at $d = 2$. Make this explicit by using the identity $\Gamma(z-1) = \Gamma(z)/(z-1)$ to express $G^{(d)}(R)$ in the form

$$G^{(d)}(R) = -\frac{\Gamma\left(\frac{1}{2}d\right)}{2\pi^{d/2}\,(d-2)\,R^{d-2}}\,.$$

A direct derivation using Gauss' Theorem: As I've mentioned already, there is a simple way to arrive at the final answer found above for $G^{(d)}(R)$, without going through the Fourier transform, its inversion, and analytic continuation in d. Let's go back to the equation

$$\nabla^2_{\mathbf{R}}\, G^{(d)}(\mathbf{R}) = \delta^{(d)}(\mathbf{R}),$$

and regard it as Poisson's equation for the electrostatic potential due to a unit charge (in suitable units) at $R = 0$. Let $\mathbf{F} \equiv \nabla G^{(d)}$ denote the 'field' due to this charge. Integrate both sides of Poisson's equation over a hypersphere of radius R. Hence

$$\int dV\, \nabla^2_{\mathbf{R}}\, G^{(d)}(\mathbf{R}) = \int dV\, \nabla \cdot \mathbf{F} = \int dV\, \delta^{(d)}(\mathbf{R}) = 1.$$

Now apply Gauss' Theorem in vector calculus[37], to obtain

$$\int dV\, (\nabla \cdot \mathbf{F}) = \int \mathbf{F} \cdot d\mathbf{S} = 1.$$

[37] This theorem is not restricted to three-dimensional space! It is valid in any number of dimensions, as it is really a form of the fundamental theorem of the calculus.

But the field required is spherically symmetric, and has only a radial component F_R which, moreover, depends only on the magnitude R. Hence

$$\int \mathbf{F} \cdot d\mathbf{S} = F_R \int dS = S_d(R)\, F_R = 1,$$

where $S_d(R)$ is the surface 'area' of a hypersphere of radius R in d-dimensional space. Hence $F_R = 1/S_d(R)$.

8. Determine $S_d(R)$. You can do so, for instance, by integrating the 'solid angle' element

$$d\Omega^{(d)} \stackrel{\text{def.}}{=} (\sin^{d-2}\theta_1)\,(\sin^{d-3}\theta_2)\,\ldots\,(\sin\,\theta_{d-2})\,d\theta_1\,\ldots\,d\theta_{d-2}\,d\varphi,$$

over the full ranges of the angular variables, and multiplying the result by R^{d-1}. The answer is

$$S_d(R) = \frac{2\pi^{d/2}\,R^{d-1}}{\Gamma\left(\frac{1}{2}d\right)}\,.$$

Returning to our problem, we have

$$F_R = (\nabla G^{(d)})_R = \frac{dG^{(d)}}{dR} = \frac{1}{S_d(R)} = \frac{\Gamma\left(\frac{1}{2}d\right)}{2\pi^{d/2}\,R^{d-1}}\,.$$

Integrating and imposing the boundary condition $G^{(d)} \to 0$ as $R \to \infty$, we get

$$\int_R^\infty \frac{dG^{(d)}(R')}{dR'}\,dR' = -G^{(d)}(R) = \frac{\Gamma\left(\frac{1}{2}d\right)}{2\pi^{d/2}} \int_R^\infty \frac{dR'}{R'^{\,d-1}}.$$

The integral on the right-hand side converges provided $\mathrm{Re}\,d > 2$. *Keeping d in this region and evaluating the integral*, we get

$$G^{(d)}(R) = -\frac{\Gamma\left(\frac{1}{2}d\right)}{2\pi^{d/2}\,(d-2)\,R^{d-2}}\,.$$

This is precisely the expression found earlier. As already pointed out, it is singular at $d = 2$. The singularity is a simple pole.

The reason why an inverse-square central force (in three dimensions) leads to an integral theorem like Gauss' Law is also obvious now. The surface area of a sphere increases with its radius like R^2. If the field drops off like $1/R^2$, there is obviously an exact compensation in the flux of the field across a sphere centered at the origin, and the total flux becomes independent of the radius R. Precisely the same thing happens in d dimensions, provided the field drops off like $1/R^{d-1}$, because the 'surface' of a hypersphere increases like R^{d-1}. The potential must then decrease like $1/R^{d-2}$, exactly

as we have deduced.

9. The Coulomb potential in $d = 2$ dimensions: The case $d = 2$ requires a separate treatment. The reason can be traced, ultimately, to a very simple fact: namely, that $\int dR/R = \ln R$, rather than a power of R. As we've found, the analytic formula for $G^{(d)}(R)$ has a simple pole at $d = 2$. This is a reflection of the fact that the original integral representation for $G^{(d)}(R)$ does indeed have an infrared divergence when $\mathrm{Re}\, d \leq 2$. The prescription of dimensional regularization, as applicable to the problem at hand, is as follows:

(i) Write $G^{(d)}(R)$, which is an analytic function of d, in the form of a Laurent series about the point $d = 2$, i.e.,

$$G^{(d)}(R) = \underbrace{\frac{\text{residue}}{(d-2)}}_{\text{singular part}} + \underbrace{\sum_{n=0}^{\infty} c_n\,(d-2)^n}_{\text{regular part}}\,.$$

(ii) Subtract out the singular part. Setting $d = 2$ in the regular part leaves behind just the coefficient c_0, which is guaranteed to be the Green function $G^{(2)}(R)$ that we seek. To find it, expand each of the d-dependent factors in the expression for $G^{(d)}(R)$, except the pole factor $(d-2)^{-1}$, in a Taylor series about $d = 2$. Retain only terms up to the first order in $(d-2)$, to obtain

$$G^{(d)}(R) = -\frac{1}{2\pi(d-2)} + \frac{1}{2\pi}\,\ln\,R + \frac{1}{4\pi}(\ln\,\pi + \gamma) + \mathcal{O}(d-2),$$

where γ is the Euler-Mascheroni constant. Therefore, according to the prescription described above, the true fundamental Green function of the Laplacian in two dimensions is

$$G^{(2)}(R) = \frac{1}{2\pi}\,\ln\,R + \text{constant}.$$

The constant is actually arbitrary, and is fixed by specifying a boundary condition.

Note that the boundary condition $G^{(2)}(R) \to 0$ as $R \to \infty$ is not possible in $d = 2$, owing to the logarithmic R-dependence of the potential. You will recognize that this logarithmic potential is essentially the same as the electrostatic potential ϕ due to an uniformly charged, infinitely long straight line in three-dimensional space, with R replaced by ϱ, the *axial* distance from the line. Recall that, in this problem too, the potential does not vanish, but instead diverges, as $\varrho \to \infty$. What is done then is to specify the value of the potential at some axial distance a to be equal to a given value ϕ_a. The potential difference $\phi(\varrho) - \phi_a$ is then $(\lambda/2\pi\epsilon_0)\,\ln\,(\varrho/a)$, where λ is the line charge density.

A direct derivation, once again: Should we believe the result obtained above, as it seems to have been derived using a prescription that appears to be arbitrary? The answer is 'Yes'. Corroboration comes from the same direct physical argument as was given for the case $d \geq 3$. As in that case, we go back to the differential equation

$$\nabla^2_{\mathbf{R}}\, G^{(2)}(\mathbf{R}) = \delta^{(2)}(\mathbf{R}).$$

Regard this as Poisson's equation for the electrostatic potential due to a unit charge (in suitable units) at $R = 0$. Let $\mathbf{F} \equiv \nabla G^{(2)}$ denote the planar vector field due to this charge. Integrate both sides of Poisson's equation over a circle of radius R. The right-hand side of the equation becomes unity, of course. Gauss' Theorem, applied to the left-hand side, gives (using the circular symmetry of the field)

$$2\pi R\, F_R = 1, \;\text{ so that }\; F_R = 1/(2\pi R).$$

But $F_R = dG^{(2)}/dR$. Integrating with respect to R gives precisely the expression found earlier, namely, $G^{(2)}(R) = (1/2\pi)\,\ln R + \text{constant}$.

- The fundamental Green function of the Laplacian (i.e., the Coulomb potential) in 2-dimensional Euclidean space is proportional to the *logarithm* of R.

This seemingly simple fact has profound consequences in diverse areas of physics, such as condensed matter physics and quantum field theory, among others. It even seems to have a bearing on the phenomenon of quark confinement!

11 The diffusion equation

Fick's laws of diffusion: Diffusion is the process by which an uneven concentration of a substance gets gradually smoothed out spontaneously—e.g., a concentration of a chemical species (like a drop of ink or dye) in a beaker of water spreads out 'by itself', even in the absence of stirring. The microscopic mechanism of diffusion involves a very large number of collisions of the dye molecules with those of the fluid, which cause the dye molecules to move essentially randomly and disperse throughout the medium, even without any stirring of the fluid. A macroscopic description of the process is based on Fick's Laws, and leads to a fundamental partial differential equation, the diffusion equation. This equation serves as a basic model of phenomena that exhibit **dissipation**, a consequence of the **irreversibility** of macroscopic systems in time.

The local, instantaneous concentration $\rho(\mathbf{r}, t)$ of dye molecules satisfies the equation of continuity, which is called Fick's first law in this context:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0, \quad \text{(Fick's I Law)}$$

where $\mathbf{j}$ the 'diffusion current density'. The crucial physical input is the specification of this quantity. We assume that $\mathbf{j}$ is proportional to the local difference in concentrations, i.e., to the gradient of the concentration itself. Thus

$$\mathbf{j}(\mathbf{r}, t) = -D \nabla \rho(\mathbf{r}, t) \quad \text{(Fick's II Law)}$$

The positive constant D called the **diffusion coefficient**. It has the physical dimensions of $(\text{length})^2/\text{time}$. The minus sign on the right-hand side signifies the fact that the diffusion occurs from a region of *higher* concentration to a region of *lower* concentration: That is, the diffusion current tends to make the concentration uniform. Eliminating $\mathbf{j}$, we get the diffusion equation for the concentration $\rho(\mathbf{r}, t)$:

$$\frac{\partial}{\partial t} \rho(\mathbf{r}, t) = D \nabla^2 \rho(\mathbf{r}, t).$$

This equation is of first order in the time variable, and second order in the spatial variables. It is a **parabolic equation** in the standard classification of second-order partial differential equations. In order to find a unique solution to it, you need an initial condition that specifies the initial concentration profile $\rho(\mathbf{r}, 0)$, as well as boundary conditions that specify $\rho(\mathbf{r}, t)$, for all $t \geq 0$, at the boundaries of the region in which the diffusion is taking place. The presence of the first-order time derivative in the diffusion equation implies that the equation is not invariant under the time reversal transformation $t \mapsto -t$. Irreversibility is thus built into the description of the phenomenon.

Fick's II Law is an example of a very general feature of diverse physical systems called **linear response**, that we discussed when we derived dispersion relations for

the generalized susceptibility. In the present case, linear response implies that the diffusion current that is set up in the medium as a result of the unequal concentrations at different points is proportional to the gradient of ρ, rather than the gradient of some nonlinear function of ρ (such as a power of ρ other than unity). Another example of linear response, leading to an exact analog of the diffusion equation, is provided by the phenomenological description of **heat conduction**. Given the initial temperature distribution $T(\mathbf{r}, 0)$ of a body, the problem is to find the temperature distribution $T(\mathbf{r}, t)$ at any later time $t > 0$. Analogous to Fick's second law, it is assumed that the heat flux is proportional to the negative of the temperature gradient (a linear response). The constant of proportionality in this case is the **thermal conductivity** of the body, κ. The equation for $T(\mathbf{r}, t)$ reads

$$\frac{\partial}{\partial t} T(\mathbf{r}, t) = \kappa \nabla^2 T(\mathbf{r}, t).$$

This is why the diffusion equation is also known as the **heat equation**.

1. The fundamental solution in d dimensions: At the level of individual particles, it turns out that the positional **probability density function** (or PDF) $p(\mathbf{r}, t)$ of a particle satisfies exactly the same diffusion equation as the concentration $\rho(\mathbf{r}, t)$ does in the macroscopic description of the diffusion process. Consider this equation in a (Euclidean) space of an arbitrary number of spatial dimensions, d. We can subsequently set $d = 1, 2, 3, \ldots$ in the solution, as required. Let's therefore begin with

$$\frac{\partial}{\partial t} p(\mathbf{r}, t) = D \nabla^2 p(\mathbf{r}, t)$$

where $p(\mathbf{r}, t)$ satisfies natural boundary conditions, i.e., $p(\mathbf{r}, t) \to 0$ as $r \to \infty$ along any direction. We may start with the initial condition

$$p(\mathbf{r}, 0) = \delta^{(d)}(\mathbf{r}),$$

where $\delta^{(d)}(\mathbf{r})$ is the d-dimensional δ-function. This means that the diffusing particle starts at the origin at $t = 0$. In the context of the diffusion equation for the concentration $\rho(\mathbf{r}, t)$, such an initial condition represents a point source of unit concentration at the origin. In a space of infinite extent, we may take the starting point to be the origin of coordinates without any loss of generality. The solution thus obtained is the fundamental solution (or Green function) of the diffusion equation. As you'll see, it can be used to write down the solution corresponding to an arbitrary initial PDF $p(\mathbf{r}, 0) = p_{\text{init}}(\mathbf{r})$ (or an initial concentration profile $\rho_{\text{init}}(\mathbf{r})$; all the results for $p(\mathbf{r}, t)$ that follow are applicable, as they stand, to the case of $\rho(\mathbf{r}, t)$.)

The diffusion equation presents an initial value problem. Moreover, it is a linear equation in the unknown function $p(\mathbf{r}, t)$. It is therefore well-suited to the application of Laplace transforms (with respect to the time variable). As far as the spatial variable

r is concerned, it is natural to use Fourier transforms. To avoid confusion, let's adopt the following notation:

$$
\begin{aligned}
p(\mathbf{r},t) &= \text{the PDF of the position at time } t;\\
\widetilde{p}(\mathbf{r},s) &= \int_0^\infty dt\, e^{-st}\, p(\mathbf{r},t),\ \text{the Laplace transform of } p(\mathbf{r},t);\\
\phi(\mathbf{k},t) &= \int d^d\mathbf{r}\, e^{-i\mathbf{k}\cdot\mathbf{r}}\, p(\mathbf{r},t),\ \text{the Fourier transform of } p(\mathbf{r},t);\\
\widetilde{\phi}(\mathbf{k},s) &= \int_0^\infty dt\, e^{-st}\, \phi(\mathbf{k},t),\ \text{the Laplace transform of } \phi(\mathbf{k},t)\\
&= \int d^d\mathbf{r}\, e^{-i\mathbf{k}\cdot\mathbf{r}}\, \widetilde{p}(\mathbf{r},s),\ \text{the Fourier transform of } \widetilde{p}(\mathbf{r},s).
\end{aligned}
$$

(You can tell which function we're dealing with by looking at the arguments of the functions concerned.) Taking the Laplace transform of both sides of the diffusion equation, we get

$$s\,\widetilde{p}(\mathbf{r},s) - p\,(\mathbf{r},0) = D\nabla^2\,\widetilde{p}(\mathbf{r},s).$$

Therefore

$$(s - D\nabla^2)\,\widetilde{p}(\mathbf{r},s) = \delta^{(d)}(\mathbf{r}).$$

Now expand $\widetilde{p}(\mathbf{r},s)$ in a Fourier integral with respect to the spatial variable **r**, according to

$$\widetilde{p}(\mathbf{r},s) = \frac{1}{(2\pi)^d}\int d^dk\ e^{i\mathbf{k}\cdot\mathbf{r}}\,\widetilde{\phi}(\mathbf{k},s).$$

The δ-function has the familiar Fourier representation

$$\delta^{(d)}(\mathbf{r}) = \frac{1}{(2\pi)^d}\int d^dk\ e^{i\mathbf{k}\cdot\mathbf{r}}.$$

Use these expressions in the equation for $\widetilde{p}(\mathbf{r},s)$ and equate the coefficients of the basis vector $e^{i\mathbf{k}\cdot\mathbf{r}}$ in the space of functions of **r**. We must then have, for each **k**,

$$(s + Dk^2)\,\widetilde{\phi}(\mathbf{k},s) = 1, \quad \text{or} \quad \widetilde{\phi}(\mathbf{k},s) = \frac{1}{s+Dk^2}\,.$$

We thus obtain a very simple expression for the double transform $\widetilde{\phi}(\mathbf{k},s)$. The transforms must be inverted to find the PDF $p(\mathbf{r},t)$. It is easier to invert the Laplace transform first, and then the Fourier transform.[38] The Laplace transform is trivially inverted: recall that $\mathcal{L}^{-1}[1/(s+a)] = e^{-at}$. Therefore

$$\phi(\mathbf{k},t) = e^{-Dk^2t}, \quad \text{and hence} \quad p\,(\mathbf{r},t) = \frac{1}{(2\pi)^d}\int d^dk\, e^{i\mathbf{k}\cdot\mathbf{r}}\, e^{-Dk^2t}.$$

[38] You can also invert the Fourier transform first, and then the Laplace transform. This procedure leads to the same result, as it ought to, but it is needlessly complicated.

This d-dimensional integral factors into a product of d integrals upon writing $\mathbf{r}$ and $\mathbf{k}$ in Cartesian coordinates. Each of the factors is the familiar 'shifted' Gaussian integral. Show that the final result is

$$p\,(\mathbf{r},t) = \frac{1}{(4\pi Dt)^{d/2}}\, e^{-r^2/(4Dt)}.$$

This is the fundamental Gaussian solution to the diffusion equation in d spatial dimensions.

Observe that the solution is spherically symmetric: the coordinate dependence of $p\,(\mathbf{r},t)$ is restricted to the radial coordinate r, with no dependence on the direction of $\mathbf{r}$. It is important to understand why this comes about. The basic reason, of course, is that the diffusion equation involves the scalar operator ∇^2, which is rotationally invariant. But it is *also* necessary for the boundary conditions and the initial condition to be spherically symmetric. These requirements are satisfied in the present instance.

Solution for an arbitrary initial distribution: We can now write down the particular integral that solves the diffusion equation for any specified initial probability density function. As a trivial generalization, we could take the initial instant to be any $t\,'$. Then, given the PDF $p_{\rm init}(\mathbf{r},t\,')$ we have, for all $t > t\,'$,

$$p(\mathbf{r},t) = \frac{1}{[4\pi D(t-t\,')]^{d/2}} \int d^d r\,' \, \exp\left\{-\frac{(\mathbf{r}-\mathbf{r}\,')^2}{4D(t-t\,')}\right\} p_{\rm init}(\mathbf{r}\,',t\,').$$

Thus, the PDF at any time $t > t\,'$ is an integral transform of the initial PDF. You will recognize that the kernel of the transform is just the Green function of the operator $(\partial/\partial t - D\nabla^2)$, and is given by

$$G(\mathbf{r},t\,;\,\mathbf{r}\,',t\,') = \frac{1}{[4\pi D(t-t\,')]^{d/2}}\, \exp\left\{-\frac{(\mathbf{r}-\mathbf{r}\,')^2}{4D(t-t\,')}\right\}$$

As I have mentioned already, the heat conduction equation has exactly the same form as the diffusion equation. For this reason, the Gaussian kernel above is called the **heat kernel** in the mathematical literature.

2. Moments of the distance travelled in a time interval t: Let's revert to $t\,' = 0$ as the initial instant. Since $p(\mathbf{r},t)$ is a function of r alone, the mean *displacement* of the diffusing particle vanishes for all $t \geq 0$, as you would expect:

$$\langle \mathbf{r}(t) \rangle = \int d^d r\, \mathbf{r}\, p(\mathbf{r},t) = 0.$$

(Integration over the angular coordinates of the vector $\mathbf{r}$ yields zero.) The mean *distance* travelled by the particle does not vanish, of course, for any $t > 0$. In fact, *all*

the moments of the distance travelled in a given time interval t can be written down easily. We find

$$\langle r^l(t)\rangle = \int d^d r\, r^l\, p(\mathbf{r},t) = \int_0^\infty dr \int d\Omega_d\; r^{d+l-1}\, p(\mathbf{r},t),$$

on going over to polar coordinates in d dimensions. Since the PDF depends on r alone, the angular integration can be carried out at once. It gives $\int d\Omega_d = S_d(1) = 2\pi^{d/2}/\Gamma\left(\frac{1}{2}d\right)$, the surface 'area' of a sphere of unit radius in d-dimensional space. The integration over r involves a Gaussian integral. Verify that

$$\langle r^l(t)\rangle = \frac{\Gamma\left(\frac{1}{2}(d+l)\right)}{\Gamma\left(\frac{1}{2}d\right)}\,(4Dt)^{l/2}.$$

The noteworthy point is that the moments $\langle r^l(t)\rangle$ increase with time like $t^{l/2}$. Moreover, this is *independent of the dimensionality d of the space* in which the diffusion takes place! In particular, the variance of the displacement is obtained by setting $l = 2$, and is given by

$$\mathrm{Var}\,(\mathbf{r}) = \langle \mathbf{r}^2\rangle - \langle \mathbf{r}\rangle^2 = \langle r^2\rangle = 2dDt.$$

3. Diffusion in one dimension: continuum limit of a random walk: Let's turn now to the case $d = 1$, i.e., diffusion in one spatial dimension. The calculations are simpler in this instance, and at the same time we obtain a number of useful insights into the nature of the diffusion problem. In particular, the effects of boundary conditions can be examined in some detail.

In order to see how diffusion arises as the continuum limit of a random walk, consider a random walk on a linear lattice. A little more generality is achieved by considering a *biased* random walk, with different probabilities for jumps to the right and left, respectively. on the lattice. This leads to a diffusion equation that is appropriate an important physical situation: namely, *diffusion on a line in the presence of a constant external force field*. The force-free case is a special case of this equation. The random walker starts at the site $j = 0$ at time $n = 0$. The probability of a step to the right is α, while that of a step to the left is $\beta = 1 - \alpha$. The probability that the walker is at the site j (that is, at the point ja on the line) at time step n (that is, at time $n\tau$) is

$$P(ja,\, n\tau) = \alpha\, P(ja - a,\, n\tau - \tau) + \beta\, P(ja + a,\, n\tau - \tau).$$

I have explicitly introduced the lattice constant a and the time step τ, so that it becomes easy to see how the continuum limit (in both space and time) arises. This happens when the quantities a, τ and $\alpha - \beta$ tend to zero simultaneously, as follows. You will find it instructive to work through the steps outlined below.

Subtract $P(ja,\, n\tau - \tau)$ from each side of the equation for $P(ja,\, n\tau)$. On the right-hand side, re-write β as $\alpha - (\alpha - \beta)$, and the coefficient of $P(ja,\, n\tau - \tau)$ as $-1 = -2\alpha + (\alpha - \beta)$. Collect terms suitably to get

$$\begin{aligned} P(ja,\, n\tau) - P(ja,\, n\tau - \tau) \\ &= \alpha \Big\{ P(ja - a,\, n\tau - \tau) - 2P(ja,\, n\tau - \tau) + P(ja + a,\, n\tau - \tau) \Big\} \\ &\quad - (\alpha - \beta) \Big\{ P(ja + a,\, n\tau - \tau) - P(ja,\, n\tau - \tau) \Big\}. \end{aligned}$$

Divide both sides by τ. Multiply and divide the first term on the right-hand side by a^2, and the second term by a. Now let $a \to 0$, $\tau \to 0$ and $\alpha - \beta \to 0$ (that is, let $\alpha \to \frac{1}{2}$, $\beta \to \frac{1}{2}$), such that

$$\lim \frac{a^2 \alpha}{\tau} = D \quad \text{and} \quad \lim \frac{a(\alpha - \beta)}{\tau} = c,$$

where D and c are finite, nonzero constants. Since $\alpha \to \frac{1}{2}$, the constant D is essentially $\lim_{a,\tau \to 0} a^2/(2\tau)$. D has the physical dimensions of (length)^2/(time), while c has the physical dimensions of (length)/(time), i.e., a velocity. Further letting $j \to \infty$ and $n \to \infty$ such that ja and $n\tau$ tend to the continuous variables x and t respectively, the difference equation above for the probability $P(ja, n\tau)$ reduces to the following partial differential equation for the probability *density* function (PDF) $p(x,t)$ of the position:

$$\frac{\partial p(x,t)}{\partial t} = -c \, \frac{\partial p\,(x,t)}{\partial x} + D \, \frac{\partial^2 p\,(x,t)}{\partial x^2}.$$

This is called the **Smoluchowski equation**. It describes diffusion in the presence of a drift—for instance, the diffusion of colloidal particles in one dimension under the influence of a constant field of force. The first term on the right-hand side is the **drift term** and the second is the **diffusion term**. The parameter c represents the **mean drift velocity**, while D is the diffusion coefficient, as usual. I will refer to diffusion in the absence of a drift as **free diffusion**. Observe that the Smoluchowski equation above can be written in the form of a continuity equation, according to

$$\frac{\partial p}{\partial t} + \frac{\partial j}{\partial x} = 0,$$

where the current density $j(x,t)$ is given by

$$j(x,t) = c\, p(x,t) - D \, \frac{\partial p\,(x,t)}{\partial x}\,.$$

The ratio D/c is a *natural length scale* in the problem, with the following physical significance: it is a measure of the relative importance of the drift and diffusion contributions to the motion of the colloidal particles. When the ratio c/D is multiplied by

a characteristic length L (such as that used in the definition of the Reynold's number in fluid flow), we get a dimensionless number Lc/D, called the **Péclet number**: it quantifies the relative strengths of the **advective transport rate** and the **diffusive transport rate** in fluid flow.

An important physical application of the Smoluchowski equation is provided by the phenomenon of **sedimentation**, the constant field of force being provided by gravity. As this phenomenon involves a medium with a boundary, we'll turn our attention, next, to the solution of the diffusion equation in the presence of finite boundaries.

Absorbing and reflecting boundary conditions: In general, in physical applications the region in which diffusion occurs is finite rather than infinite. We then need to specify appropriate boundary conditions at the end-points of the region. It is important to recognize that the solutions of a given PDE with the same initial conditions but different boundary conditions may be very different functions. Let's consider, for simplicity and ease of illustration, diffusion in one spatial dimension (the x-axis, say), inside the 'box' given by $b_1 \leq x \leq b_2$. If the diffusing particle gets absorbed at the ends of the box (or can leak out through the ends and thus leave the region of physical interest), we must impose the so-called **absorbing boundary conditions**, for which I'll use the abbreviation ABC. I'll also write the corresponding PDF as $p_{\rm abs}(x,t)$, to avoid any confusion.

$$p_{\rm abs}(b_1\,,\,t) = 0 \text{ and } p_{\rm abs}(b_2\,,\,t) = 0. \quad \text{(ABC)}$$

On the other hand, if the substance cannot leak out of the ends and stays confined in the region $b_1 \leq x \leq b_2$ at all times, the diffusion *current* must vanish at the end-points. Hence we must impose the so-called **reflecting boundary conditions**, for which I'll use the abbreviation RBC. The corresponding PDF will be written as $p_{\rm ref}(x,t)$. These conditions imply that the *flux* or current, rather than the PDF itself, vanishes at the boundaries. In the case of free diffusion, this means that

$$\frac{\partial p_{\rm ref}}{\partial x} = 0 \text{ at } x = b_1 \text{ and } x = b_2\,, \text{ for all } t \geq 0. \quad \text{(RBC)}$$

RBCs are a little more involved when both diffusion and drift are present. Recall that the Smoluchowski equation for the PDF $p(x,t)$ in the case of diffusion with drift in one dimension is given by

$$\frac{\partial p(x,t)}{\partial t} = -c\,\frac{\partial p\,(x,t)}{\partial x} + D\,\frac{\partial^2 p\,(x,t)}{\partial x^2}.$$

As mentioned already, we can write this in the form of an equation of continuity,

$$\frac{\partial p}{\partial t} + \frac{\partial j}{\partial x} = 0, \quad \text{where} \quad j(x,t) = c\,p(x,t) - D\,\frac{\partial p\,(x,t)}{\partial x}\,.$$

Hence RBCs in this case are given by

$$cp_{\rm ref} - D\frac{\partial p_{\rm ref}}{\partial x} = 0 \text{ at } x = b_1 \text{ and } x = b_2\,. \quad \text{(RBC)}$$

Other boundary conditions, or combinations of these, may be imposed, depending on the physical problem.

There is a drastic difference in the long-time behavior of the respective solutions for reflecting and absorbing boundary conditions. In the case of reflecting boundaries, the particle never leaves the region between the boundaries. (Alternatively, the total amount of the diffusing substance remains conserved.) Therefore the normalization condition, $\int_{b_1}^{b_2} dx\, p_{\rm ref}(x,t) = 1$, remains valid for all $t \geq 0$. In the case of absorbing boundaries, however, the particle is absorbed when it happens to hit either boundary. The diffusion process then comes to an end (or the diffusing substance leaks out through the end points). In this case it is clear that $p_{\rm abs}(x,t) \to 0$ as $t \to \infty$, at any point x inside the box. The total probability $\int_{b_1}^{b_2} dx\, p_{\rm abs}(x,t) = S(t)$ then represents the **survival probability** inside the box. It starts with an initial value equal to unity, and decreases to zero with increasing time. The question of precisely how $S(t) \to 0$ as $t \to \infty$ is of interest in its own right. We'll consider it shortly, in the case of free diffusion.

4. Free diffusion on a semi-infinite line: Consider free diffusion in the semi-infinite region $-\infty < x < b$, where b is a positive constant. On the left, we have the natural boundary condition $p(x,t) \to 0$ as $x \to -\infty$. On the right, we have either (i) $p_{\rm abs}(b,t) = 0$ if the barrier at $x = b$ is an absorber, or (ii) $(\partial p_{\rm ref}/\partial x)_{x=b} = 0$ if the barrier is a reflector. Verify that, for the initial condition $p(x,0) = \delta(x)$, the solutions in the two cases are give by

$$p_{\rm ref}(x,t) = \frac{1}{(4\pi Dt)^{1/2}}\,\left[e^{-x^2/(4Dt)} + e^{-(2b-x)^2/(4Dt)}\right]$$

and

$$p_{\rm abs}(x,t) = \frac{1}{(4\pi Dt)^{1/2}}\,\left[e^{-x^2/(4Dt)} - e^{-(2b-x)^2/(4Dt)}\right],$$

respectively.

5. Finite boundaries: Solution by the method of images: The solutions above have a striking interpretation. They are superpositions (the sum and the difference, respectively) of (Gaussian) solutions of the diffusion equation with *natural* boundary conditions, but with x and $(2b - x)$, respectively, as the coordinate argument. Now, $(2b - x)$ is precisely the coordinate of the *image* of the point x reflected in a mirror located at the boundary b. This is no coincidence! The solutions above can indeed be obtained by a general technique called **the method of images**. You have come across

this method in the context of electrostatics, where it yields the solution to Poisson's equation in the presence of specific boundary conditions, provided the problem has a convenient symmetry. However, the method is quite general, and is applicable in other instances as well. Broadly speaking, the method of images is a technique to find the Green function of a linear differential operator with specific boundary conditions, provided the problem has some symmetry that can be exploited. It is based on the fact that the differential equation concerned, taken together with appropriate initial and boundary conditions, has a *unique* solution. My purpose in using the method of images in the present context of the diffusion equation is two-fold. The first is to demonstrate how powerful the method is, when the conditions are right. The second is to show that the method is also applicable to time-*dependent* problems, and not just time-independent problems as in electrostatics.

Consider free diffusion in the line segment region $-b \leq x \leq b$, where b is a positive constant. For simplicity, we assume once again that the diffusing particle starts at the origin at $t = 0$, so that $p(x, 0) = \delta(x)$. We want the solution to the diffusion equation subject to this initial condition and

$$\text{(i)} \ \left[\partial p_{\text{ref}}(x,t)/\partial x\right]_{x=\pm b} = 0 \ \ (\text{RBC}); \text{ or } \ \text{(ii)} \ p_{\text{abs}}(\pm b, t) = 0 \ \ (\text{ABC}).$$

Imagine placing mirrors at both boundaries, facing each other. A source point x therefore has an infinite number of images, at the points $-2b-x$, $2b-x$, $-4b+x$, $4b+x$, ..., as we move away from the source point on both sides of the x-axis. (Draw a figure and mark the successive image points.) The solution is a superposition of the fundamental Gaussian solution with each of these points as the coordinate argument. The boundary conditions above are incorporated by the following simple prescriptions:

(i) For RBC, the coefficient of the contribution from each image is just 1.

(ii) For ABC, the coefficient of the contribution from an image arising from n reflections is $(-1)^n$.

The complete solutions in the two cases may therefore be written down. Show that, after a bit of simplification, they are given by

$$p_{\text{ref}}(x,t) = \frac{1}{(4\pi Dt)^{1/2}} \sum_{n=-\infty}^{\infty} e^{-(x+2nb)^2/(4Dt)} \quad (\text{RBC at } x = \pm b)$$

and

$$p_{\text{abs}}(x,t) = \frac{1}{(4\pi Dt)^{1/2}} \sum_{n=-\infty}^{\infty} (-1)^n \, e^{-(x+2nb)^2/(4Dt)} \quad (\text{ABC at } x = \pm b).$$

6. You may now ask: can each of these solutions be written as the fundamental Gaussian solution on the infinite line, plus an 'extra' piece arising from the presence of the finite boundaries? Show that the solutions above can be re-written in the form

$$p_{\rm ref}(x,t) = \frac{e^{-x^2/(4Dt)}}{(4\pi Dt)^{1/2}} \left\{ 1 + 2\sum_{n=1}^{\infty} e^{-n^2b^2/(Dt)} \cosh\left(\frac{nbx}{Dt}\right) \right\} \quad \text{(RBC)}$$

and

$$p_{\rm abs}(x,t) = \frac{e^{-x^2/(4Dt)}}{(4\pi Dt)^{1/2}} \left\{ 1 + 2\sum_{n=1}^{\infty} (-1)^n\, e^{-n^2b^2/(Dt)} \cosh\left(\frac{nbx}{Dt}\right) \right\} \quad \text{(ABC)},$$

respectively. It is important to remember, however, that the physical region in which these solutions are valid is restricted to the line interval $[-b, b]$.

- It is remarkable that the *only* difference between the solutions for reflecting and absorbing boundary conditions is the extra factor $(-1)^n$ in the summand in the latter case. But this is sufficient to alter completely the long-time behavior of the PDF $p(x,t)$, as you will see below.

7. Finite boundaries: Solution by separation of variables: You are undoubtedly familiar with an elementary method for the solution of partial differential equations, namely, the method of separation of variables The diffusion equation, with an initial condition $p(x,t) = \delta(x)$ and either reflecting or absorbing boundary conditions at $x = \pm b$, is tailor-made for solution by this method. Set $p(x,t) = T(t)\,X(x)$, as usual, in the diffusion equation. It follows that $(1/DT)dT/dt = (1/X)dX/dx = -C^2$, a constant. The boundary conditions yield the allowed values of C. The general solution is a superposition of solutions for the various allowed values of C, as the diffusion equation is a linear equation. The calculations are simplified by noting that the diffusion equation, the set of boundary conditions, as well as the initial condition, are *all* unchanged under the transformation $x \mapsto -x$. As a consequence, the solution $p(x,t)$ remains an even function of x at all times. This fact will help you select the correct solution in each case. You will also need the following representation of the initial PDF,

$$\delta(x) = \frac{1}{2b} + \frac{1}{b}\sum_{n=1}^{\infty} \cos\,\frac{n\pi x}{b}\,,$$

which follows from the relation $\delta\big(x/(2b)\big) = \sum\limits_{n=-\infty}^{\infty} e^{\pi nix/b}$. Show that

$$p_{\rm ref}(x,t) = \frac{1}{2b} + \frac{1}{b}\sum_{n=1}^{\infty} e^{-\,n^2\pi^2 Dt/b^2} \cos\frac{n\pi x}{b} \quad \text{(RBC)}$$

and

$$p_{\rm abs}(x,t) = \frac{1}{b}\sum_{n=0}^{\infty} e^{-(2n+1)^2\pi^2 Dt/(4b^2)} \cos\frac{(2n+1)\pi x}{2b} \quad \text{(ABC)}.$$

8. Survival probability: The representations of $p(x,t)$ just found are useful in reading off the *long-time* behavior of the PDF, because they are superpositions of decaying exponential functions of time. It follows by inspection that $p_{\rm ref}(x,t) \to 1/(2b)$ as $t \to \infty$ in the case of reflecting boundaries. This uniform distribution in the interval $[-b\,,\,b]$ is precisely what we would expect on physical grounds as the asymptotic PDF. On the other hand, in the case of absorbing boundaries, $p_{\rm abs}(x,t) \to 0$ as $t \to \infty$ for every value of x. This is an indication of the fact that absorption at either one boundary or the other is a **sure event** for the random process concerned—that is, it will occur with probability 1. The **survival probability** $S(t, \pm b \,|\, 0)$ is the probability that, starting from the origin at $t = 0$, the diffusing particle survives in the open interval $(-b, b)$ till time t, without hitting either of the boundary points.[39] Its definition is obvious:

$$S(t, \pm b \,|\, 0) = \int_{-b}^{b} dx\, p_{\rm abs}(x,t).$$

(a) Using the solution for $p_{\rm abs}(x,t)$ with ABC, check that

$$S(t, \pm b \,|\, 0) = \frac{4}{\pi}\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)} e^{-\,(2n+1)^2\pi^2 Dt/(4b^2)}.$$

(b) Hence verify that $S(0, \pm b \,|\, 0) = 1$, as required. You will need the value[40] of the series $\sum_0^\infty (-1)^n/(2n+1)$.

$S(t, \pm b \,|\, 0)$ is a superposition of an infinite number of decaying exponentials that decays monotonically to zero, with a leading asymptotic behavior $\sim e^{-\pi^2 Dt/(4b^2)}$.

9. First-passage-time distribution & mean first-passage time: A diffusing particle starting from $x = 0$ will hit one of the end-points $\pm b$ for the *first* time at some random instant of time, which may be called the **first-passage time**.[41] What is the distribution of this time? This sort of question occurs in numerous applications—for instance, in reaction-diffusion problems in chemical physics.

[39] More generally, the survival probability in any region should be computed by averaging over all possible starting points as well, but I do not go into this detail here.

[40] This was first deduced by the father of mathematical analysis, Madhava of Sangamagrama (1350-1425), as a special case of the power series for $\tan^{-1} x$ discovered by him. The latter was rediscovered a few centuries later by Gregory (1638-1675), and is now known as the Madhava-Gregory series.

[41] Also called the *escape time*, or *hitting time*, or *exit time*, depending on the application.

Let $Q(t, \pm b \,|\, 0)\, dt$ be the probability for a particle starting from $x = 0$ at $t = 0$ to reach either b or $-b$ for the *first* time in the time interval $(t, t+dt)$, *without* ever having hit either of the end-points at any earlier time. It is obvious that the probability density $Q(t, \pm b \,|\, 0)$ is nothing but the rate of decrease of the survival probability $S(t, \pm b \,|\, 0)$. That is,

$$Q(t, \pm b \,|\, 0) \equiv -\frac{d}{dt}\, S(t, \pm b \,|\, 0) = \frac{\pi D}{b^2} \sum_{n=0}^{\infty} (-1)^n (2n+1)\, e^{-(2n+1)^2 \pi^2 D t/(4b^2)}.$$

(a) Verify that $Q(t, \pm b \,|\, 0)$ is normalized to unity, i.e., $\int_0^\infty dt\, Q(t, \pm b \,|\, 0) = 1$.

Hence a first passage to either one or the other of the end-points is a sure event, as already stated.

The **mean first-passage time** (MFPT) $\langle t(0 \to \pm b) \rangle$ is also easily determined. By dimensional considerations, it must be proportional to b^2/D, which is the only time scale in the problem.

(b) Show that the MFPT $\langle t(0 \to \pm b) \rangle = b^2/(2D)$, justifying the term 'diffusion time' for this mean value. You will need the value of the sum $\sum_0^\infty (-1)^n/(2n+1)^3$, which is $\pi^3/32$.

In many instances, first-passage-time problems are conveniently handled by considering the Laplace transforms of the corresponding first-passage-time densities. In the present instance, let $\widetilde{Q}(s, \pm b \,|\, 0)$ denote the Laplace transform of $Q(t, \pm b \,|\, 0)$. First passage (from the starting point 0 to either of the points $\pm b$) is a sure event if and only if

$$\widetilde{Q}(0, \pm b \,|\, 0) = 1.$$

Further, the mean first-passage time is given by

$$\langle t(0 \to \pm b) \rangle = -\Big[\frac{d\widetilde{Q}}{ds}\Big]_{s=0}.$$

Connection with the Schrödinger equation for a free particle: Let's return to the fundamental solution of the diffusion equation in d-dimensional space, for an arbitrary initial PDF. This solution enables us to write down a similar solution for another very important equation, namely, the Schrödinger equation for the position-space wave function $\psi(\mathbf{r}, t)$ of a nonrelativistic free particle of mass m moving in d-dimensional space:

$$\frac{\partial \psi(\mathbf{r}, t)}{\partial t} = \frac{i\hbar}{2m} \nabla^2 \psi(\mathbf{r}, t).$$

It is obvious that this equation has exactly the same form as the diffusion equation, with the physically and mathematically important difference that the *real* diffusion constant D in the latter is replaced by the *pure imaginary* constant $i\hbar/(2m)$. We can now write down the formal solution to the Schrödinger equation by analogy with the solution of the diffusion equation. Given that the wave function is $\psi(\mathbf{r}, t')$ at some instant of time t', the solution at any time $t > t'$ is given by

$$\psi(\mathbf{r}, t) = \left[\frac{m}{2\pi i\hbar(t-t')}\right]^{d/2} \int d^d r' \, \exp\left\{\frac{im(\mathbf{r}-\mathbf{r}')^2}{2\hbar(t-t')}\right\} \psi(\mathbf{r}', t').$$

But we know from quantum mechanics that the solution must have the form

$$\psi(\mathbf{r}, t) = \int d^d r' \, K(\mathbf{r}, t\,;\, \mathbf{r}', t')\, \psi(\mathbf{r}', t'), \quad (t > t')$$

where $K(\mathbf{r}, t\,;\, \mathbf{r}', t')$ is the free-particle **Feynman propagator**. This quantity is defined as

$$K(\mathbf{r}, t\,;\, \mathbf{r}', t') = \left\langle \mathbf{r} \middle| e^{-iH(t-t')/\hbar} \middle| \mathbf{r}' \right\rangle,$$

where H is the free-particle Hamiltonian $\mathbf{p}^2/(2m)$, and $|\mathbf{r}\rangle$ is the position eigenstate of the particle corresponding to the position eigenvalue $\mathbf{r}$. Hence the explicit form of the free-particle propagator for a nonrelativistic particle moving in d-dimensional Euclidean space is

$$K(\mathbf{r}, t\,;\, \mathbf{r}', t') = \left[\frac{m}{2\pi i\hbar(t-t')}\right]^{d/2} \exp\left\{\frac{im(\mathbf{r}-\mathbf{r}')^2}{2\hbar(t-t')}\right\} \quad (t > t').$$

The replacement of the real constant D by the pure imaginary constant $i\hbar/(2m)$ actually represents a drastic change. The kernel $\exp\left[-(\mathbf{r}-\mathbf{r}')^2/4D(t-t')\right]$ decays to zero as $|\mathbf{r}-\mathbf{r}'| \to \infty$. On the other hand, the kernel $\exp\left[im(\mathbf{r}-\mathbf{r}')^2/2\hbar(t-t')\right]$ is an oscillatory function with a modulus equal to unity. As a result, questions of convergence arise, and these require careful handling. I do not go into these aspects here, but merely mention that the formal similarity between the diffusion equation and the Schrödinger equation has important consequences. The solution for $\psi(\mathbf{r}, t)$ written down above serves as the starting point for the **path integral formulation** of quantum mechanics.

10. Spreading of a quantum mechanical wave packet: As you know from elementary quantum mechanics, the wave packet representing a free particle undergoes **dispersion**, i.e., it broadens or spreads in time, even though the particle moves in a vacuum. The physical reason for this fact is easy to see. Let ε and p denote the energy and the magnitude of the momentum of the particle, respectively. Wave-particle duality is expressed by the Einstein-de Broglie relations $\varepsilon = \hbar\omega$ and $p = \hbar k$, where ω is the angular frequency and k is the wave number. Hence the relation $\varepsilon = p^2/(2m)$ becomes $\omega = \hbar k^2/(2m)$. This nonlinear dispersion relation immediately implies that the phase velocity ω/k is not equal to the group velocity $d\omega/dk$, i.e., there is dispersion. In other

words, an initial wave packet will change shape and spread with time, *even though the particle is free*, and not under the influence of any force.

In order to see this quantitatively, let's consider the problem in one spatial dimension. Setting $t' = 0$ for simplicity, we have

$$\psi(x,t) = \Big(\frac{m}{2\pi i\hbar t}\Big)^{1/2} \int_{-\infty}^{\infty} dx\, e^{im(x-x')^2/(2\hbar t)}\, \psi(x',0).$$

We need an initial state that represents the quantum mechanical counterpart of a classical free particle moving with some constant momentum p_0. We could start with $|p_0\rangle$, which is an eigenstate of the momentum of the particle, corresponding to the eigenvalue p_0. The position-space wave function corresponding to this state is $\langle x \,|\, p_0\rangle$. This is a plane wave proportional to $\exp(ip_0\, x/\hbar)$. But such a wave function has a constant modulus, and is clearly not normalizable in $(-\infty, \infty)$, whereas we would like to work with square-integrable functions throughout. This is achieved by *modulating* the plane wave with a Gaussian, centered at the origin, say. The initial wave function is then given by

$$\psi(x,0) = (2\pi\sigma^2)^{-1/4}\; e^{ip_0 x/\hbar}\, e^{-x^2/4\sigma^2},$$

where σ is a positive constant with the physical dimensions of a length. The wave function is normalized according to $\int_{-\infty}^{\infty} dx\, |\psi(x,0)|^2 = 1$.

(a) Show that

$$\langle x(0)\rangle = \int_{-\infty}^{\infty} dx\, x\, |\psi(x,0)|^2 = 0, \;\; \langle x^2(0)\rangle = \int_{-\infty}^{\infty} dx\, x^2\, |\psi(x,0)|^2 = \sigma^2.$$

Therefore the initial uncertainty (i.e., the standard deviation) in the position of the particle is $\Delta x(0) = \sigma$.

(b) Show also that

$$\langle p(0)\rangle = \int_{-\infty}^{\infty} dx\, \psi^*(x,0)\, (-i\hbar)\, \frac{d}{dx}\, \psi(x,0) = p_0\,,$$

while

$$\langle p^2(0)\rangle = \int_{-\infty}^{\infty} dx\, \psi^*(x,0)\, (-i\hbar)^2\, \frac{d^2}{dx^2}\, \psi(x,0) = p_0^2 + \frac{\hbar^2}{4\sigma^2}\,.$$

Hence the initial uncertainty (or standard deviation) in the momentum is $\Delta p(0) = \hbar/(2\sigma)$. It follows that $\Delta x(0)\, \Delta p(0) = \frac{1}{2}\hbar$. In other words, the initial state is a **minimum uncertainty state**.

The Hamiltonian of a free particle is given by just the kinetic energy term, $p^2/(2m)$. As this is quadratic in the dynamical variable concerned, **Ehrenfest's Theorem** implies that the expectation value $\langle p(t)\rangle$ will remain equal to p_0 for all t, while $\langle x(t)\rangle$ will be given by $p_0 t/m$: these are the expressions that would obtain for a classical particle. These statements can be checked out directly, as follows.

(c) The integral representing $\psi(x,t)$ now becomes a shifted Gaussian integral, and can be evaluated, to find the wave function at any time t. Show that

$$\psi(x,t) = \frac{e^{\phi(x,t)}}{(2\pi\sigma^2)^{1/4}[1+(i\hbar t/2m\sigma^2)]^{1/2}}\,,$$

where the exponent $\phi(x,t)$ is given by

$$\phi(x.t) = \frac{imx^2}{2\hbar t} - \frac{im(x-p_0t/m)^2}{2\hbar t[1+(\hbar^2t^2/4m^2\sigma^4)]} - \frac{(x-p_0t/m)^2}{4\sigma^2[1+(\hbar^2t^2/4m^2\sigma^4)]}\,.$$

The first two terms on the right-hand side of the last equation are purely imaginary, and hence only contribute phase factors to $\psi(x,t)$. They do not, therefore, affect the probability density function $|\psi(x,t)|^2$. This quantity is given by

$$|\psi(x,t)|^2 = \left\{\frac{1}{2\pi[\sigma^2+(\hbar^2t^2/4m^2\sigma^2)]}\right\}^{1/2} \exp\left\{-\frac{(x-p_0t/m)^2}{2[\sigma^2+(\hbar^2t^2/4m^2\sigma^2)]}\right\}.$$

Thus, the PDF of the position of the particle at any time $t>0$ is also a normalized Gaussian, with its peak at $x=p_0t/m$, and a variance that increases with time.

(d) Show that

$$\left.\begin{aligned} \langle x(t)\rangle &= \frac{p_0t}{m}\,, \\ \langle x^2(t)\rangle - \langle x(t)\rangle^2 &\equiv \big(\Delta x(t)\big)^2 = \sigma^2 + \frac{\hbar^2t^2}{4m^2\sigma^2}\,. \end{aligned}\right\}$$

(e) Show also that

$$\left.\begin{aligned} \langle p(t)\rangle &= \int_{-\infty}^{\infty} dx\,\psi^*(x,t)\,(-i\hbar)\frac{d}{dx}\psi(x,t) = p_0\,, \\ \langle p^2(t)\rangle &= \int_{-\infty}^{\infty} dx\,\psi^*(x,t)\,(-i\hbar)^2\frac{d^2}{dx^2}\psi(x,t) = p_0^2 + \frac{\hbar^2}{4\sigma^2}\,. \end{aligned}\right\}$$

We thus arrive at the following conclusions:

(i) The initial Gaussian wave packet in position space remains a Gaussian wave packet, but broadens as a function of time.

(ii) The uncertainty in the position of the particle increases with time according to

$$\Delta x(t) = \sigma\left(1+\frac{\hbar^2t^2}{4m^2\sigma^4}\right)^{1/2}.$$

At very long times ($t \gg 2m\sigma^2/\hbar$), the uncertainty in the position increases linearly with time.

(iii) The uncertainty in the momentum at any time t remains equal to its initial value, $\hbar/(2\sigma)$. Thus the state of the particle is no longer a minimum uncertainty state for any $t > 0$.

Observe, incidentally, that the energy of the particle (i.e., the expectation value of its Hamiltonian) in the state under consideration is not $p_0^2/(2m)$, but rather

$$E \equiv \langle H \rangle = \frac{1}{2m} \langle p^2 \rangle = \frac{p_0^2}{2m} + \frac{\hbar^2}{8m\sigma^2} \, .$$

$\langle H \rangle$, of course, is constant in time.

(f) The particle is not in a momentum eigenstate either at $t = 0$ or at any time $t > 0$. Why, then, do the expectations values of p and p^2 remain unchanged from their initial values, while those of x and x^2 are functions of time?

12 The Green function for $(\nabla^2 + k^2)$; nonrelativistic scattering

The Helmholtz operator: After the Laplacian operator ∇^2 and the diffusion operator $(\partial/\partial t - D\nabla^2)$, we turn now to the Helmholtz operator $\nabla^2 + k^2$, where k is a constant. As you know, the normal modes of vibration of a region $\mathcal{R}$ are given by the solutions of the Helmholtz equation $(\nabla^2 + k^2)u(\mathbf{r}, t) = 0$, i.e., the eigenvalue equation $\nabla^2 u = -k^2 u$, with appropriate boundary conditions on the 'displacement' u at the boundary $\mathcal{S}$ of the region. The most common boundary condition is the **Dirichlet boundary condition** $u(\mathbf{r}, t) = 0$ for $\mathbf{r} \in \mathcal{S}$. A vast literature exists on this problem and its solutions.

Here, we shall not be concerned with the foregoing eigenvalue problem, but rather, with the inhomogeneous equation $(\nabla^2 + k^2)\, f(\mathbf{r}, t) = g(\mathbf{r}, t)$. More specifically, we are interested in finding the fundamental Green function for the operator $\nabla^2 + k^2$. As it is useful to work in the context of a specific physical application, I will consider quantum mechanical scattering theory. For simplicity, we'll focus on the elastic scattering of a nonrelativistic particle of mass m from a static central potential $V(r)$ that vanishes as $r \to \infty$. It is convenient (and conventional) to introduce a parameter λ as a measure of the 'strength' of the potential, so that various physical quantities can be expressed as power series in λ, and approximations to different orders in λ can be made.

The scattering amplitude; differential and total cross-sections: The time-independent Schrödinger equation for the position-space wave function of the particle is given by

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + \lambda V(r)\,\psi(\mathbf{r}) = E\,\psi(\mathbf{r}).$$

We are interested here in the **scattering states** of the particle, i.e., states belonging to the *continuous* part of the spectrum of the Hamiltonian, with energy eigenvalue $E > 0$. These states are not normalizable, i.e., the corresponding wave functions are not square-integrable.[42] The initial state of the particle is taken to be a momentum eigenstate, with eigenvalue $\mathbf{p} = \hbar\mathbf{k}$. It is therefore represented by the position-space wave function $\psi_{\text{inc}}(\mathbf{r}) \equiv \langle \mathbf{r}|\mathbf{p}\rangle = e^{i\mathbf{p}\cdot\mathbf{r}/\hbar} = e^{i\mathbf{k}\cdot\mathbf{r}}$. Thus $\mathbf{k}$ is the direction of the momentum of the incident particle. After scattering, the wave vector changes direction without changing its magnitude, because the scattering is elastic: the energy remains

$$E = \hbar^2 k^2/(2m)$$

throughout. The scattered wave vector can be directed along any direction in space. The quantity of interest is the *probability* of scattering in any particular direction. This

[42] This technical difficulty can be overcome by working with normalizable wave packets rather than plane waves as wave functions. But we shall not enter into this aspect here.

is measured by the **differential cross-section** $d\sigma/d\Omega$, defined as

$$\frac{d\sigma}{d\Omega} = \frac{\text{flux of particles per unit solid angle in the direction concerned}}{\text{incident flux}}\,.$$

Sufficiently far away from the scattering center (i.e., as $r \to \infty$, or more precisely, for $kr \gg 1$), the scattered wave has the form of an *outgoing spherical wave*. The radial coordinate dependence of this wave is (in three-dimensional space) e^{ikr}/r, and it is, in general, modulated by an amplitude f relative to the unit amplitude of the incident plane wave. The amplitude f is, in general, a function of the angular coordinates θ and φ, as well as the energy E (or the wave number k). Without loss of generality, we may choose spherical polar coordinates such that the polar axis is along the direction of the incident wave vector $\mathbf{k}$. Then, the spherical symmetry of a central potential clearly implies that there can be no φ-dependence in the amplitude f. There remains a nontrivial θ-dependence, because the presence of an incident wave vector $\mathbf{k}$ breaks the spherical symmetry, leaving an axial symmetry about the direction of $\mathbf{k}$. The total wave function is a superposition of the incident and scattered wave functions. Its asymptotic form given by

$$\psi(\mathbf{r}) = \psi_{\text{inc}}(\mathbf{r}) + \psi_{\text{sc}}(\mathbf{r}) \xrightarrow[kr\gg 1]{} e^{i\mathbf{k}\cdot\mathbf{r}} + f(k,\theta)\,\frac{e^{ikr}}{r}\,,$$

The energy and angle-dependent factor $f(k,\theta)$ is called the **scattering amplitude**.

To repeat: The scattering amplitude f is independent of the azimuthal angle φ because the scattering potential is spherically symmetrical. But it does have θ-dependence, because the incident wave vector $\mathbf{k}$ singles out a special direction. The potential is spherically symmetric, but the symmetry of the scattered wave function is reduced to an axial or cylindrical symmetry about the direction of the initial momentum of the particle.

1. There is a direct relation between the differential cross-section and the scattering amplitude. incident current density is $\hbar(\psi^*_{\text{inc}}\,\nabla\psi_{\text{inc}} - \psi_{\text{inc}}\,\nabla\psi^*_{\text{inc}})/(2mi)$, which is easily simplified to $\hbar\mathbf{k}/m$. Its magnitude $\hbar k/m$ is the incident flux. The scattered current density is given by $\hbar(\psi^*_{\text{sc}}\,\nabla\psi_{\text{sc}} - \psi_{\text{sc}}\,\nabla\psi^*_{\text{sc}})/(2mi)$. What we need here is the radially outward scattered flux through a cone of solid angle $d\Omega$ located at any angular position (θ,φ). This is given by

$$\frac{\hbar}{2mi}\left(\psi^*_{\text{sc}}\,\frac{\partial\psi_{\text{sc}}}{\partial r} - \psi_{\text{sc}}\,\frac{\partial\psi^*_{\text{sc}}}{\partial r}\right) r^2 d\Omega.$$

Using the asymptotic form above for ψ_{sc}, show that this quantity simplifies to

$$(\hbar k/m)\,|f(k,\theta)|^2\,d\Omega.$$

Hence the differential cross-section is given by

$$d\sigma/d\Omega = |f(k,\theta)|^2.$$

The **total cross-section** for scattering is then

$$\sigma \stackrel{\text{def.}}{=} \int d\sigma = \int d\Omega\, |f(k,\theta)|^2 = 2\pi \int_{-1}^{1} d\,(\cos\,\theta)\, |f(k,\theta)|^2.$$

σ is obviously a function of k alone, i.e., a function of the energy E of the incident particle.

Integral equation for scattering: In order to find the scattering amplitude, we need the asymptotic form of the solution for the wave function. For this purpose, it is convenient to convert the time-independent Schrödinger equation from a differential equation to an **integral equation**. First write the differential equation in the form

$$(\nabla^2 + k^2)\,\psi(\mathbf{r}) = \lambda\, U(r)\,\psi(\mathbf{r}), \quad \text{where} \quad U(r) = (2m/\hbar^2)\, V(r).$$

Suppose, for a moment, we treat the right-hand side $\lambda U(r)\psi(\mathbf{r})$ as a 'source' term in an inhomogeneous differential equation. The general solution will be the sum of the complementary function and the particular integral. The former is a solution of the homogeneous equation $(\nabla^2 + k^2)\,\psi(\mathbf{r}) = 0$, chosen so as to satisfy the boundary conditions. In the present case, the complementary function is just the incident wave $e^{i\mathbf{k}\cdot\mathbf{r}}$. The particular integral involves $G(\mathbf{r},\mathbf{r}\,')$, the Green function for the Helmholtz operator $(\nabla^2 + k^2)$. Thus

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} + \lambda \int d^3r\,'\, G(\mathbf{r},\mathbf{r}\,')\, U(r\,')\, \psi(\mathbf{r}\,').$$

The presence of the unknown function ψ in the integrand in the last term makes this an inhomogeneous integral equation for the wave function, rather than a solution for this quantity.

2. Green function for the Helmholtz operator: The Green function satisfies the equation

$$(\nabla^2_{\mathbf{r}} + k^2) G(\mathbf{r},\mathbf{r}\,') = \delta^{(3)}(\mathbf{r} - \mathbf{r}\,').$$

Now, the operator $\nabla^2_{\mathbf{r}} + k^2$, the the δ-function on the right-hand side, as well as the free boundary condition (the vanishing of G as $r \to \infty$) are all translation invariant (i.e., invariant under the shift $\mathbf{r} \to \mathbf{r} - \mathbf{r}\,'$). Hence G will once again turn out to be a function of $\mathbf{R} = \mathbf{r} - \mathbf{r}\,'$. Introducing the Fourier transform of $G(\mathbf{R})$ and following the same steps as in the case of ∇^2, we now get

$$G(\mathbf{R}) = -\frac{1}{(2\pi)^3} \int d^3q\, \frac{e^{i\mathbf{q}\cdot\mathbf{R}}}{q^2 - k^2}\,.$$

I have used **q** for the Fourier transform variable conjugate to **R**, since **k** has already been used for the incident wave vector. As usual, go over to spherical polar coordinates in **q**-space, and choose the polar axis along the direction of the vector **R**. Carry out the angular integrations to get

$$G(R) = -\frac{1}{2\pi^2 R}\int_0^\infty dq\,\frac{q\,\sin\,qR}{q^2-k^2} = -\frac{1}{8\pi^2 iR}\int_{-\infty}^\infty dq\,\frac{q\,(e^{iqR}-e^{-iqR})}{q^2-k^2}\,.$$

Note that the answer is a function of R alone, which is why $G(\mathbf{R})$ has been written as $G(R)$. We must now use an appropriate $i\epsilon$-prescription to avoid the singularities of the integrand at the points $q = \pm k$ on the path of integration.

- This prescription will also help us select the correct Green function, i.e., the one that corresponds to an outgoing spherical wave with an asymptotic behavior $\sim e^{ikr}/r$.

The next step is to close the contour of integration in the complex q-plane by adding a semicircle to the line integral. The contribution of this semicircle to the integral must vanish as its radius tends to infinity. It must therefore lie (i) in the *upper* half-plane for the term involving e^{iqR}, and (ii) in the *lower* half-plane for the term involving e^{-iqR}. Moreover, we want $G(R)$ to have an asymptotic behavior $\sim e^{ikr}/r$ as $r \to \infty$. Hence:

(i) Only the pole at $q = k$ must contribute to the term proportional to e^{iqR}. Therefore, the pole at $q = k$ must be displaced into the upper half-plane.

(ii) Only the pole at $q = -k$ must contribute to the term proportional to e^{-iqR}. Therefore, the pole at $q = -k$ must be displaced into the lower half-plane.

The correct $i\epsilon$-prescription is thus implemented by setting

$$G(R) = -\frac{1}{8\pi^2 iR}\lim_{\epsilon\to 0}\int_{-\infty}^\infty dq\,\frac{q\,(e^{iqR}-e^{-iqR})}{q^2-(k+i\epsilon)^2}\,,$$

where ϵ is an infinitesimal positive number. Close the contour as prescribed above (draw the corresponding figures) and use the residue theorem to show that

$$G(R) = -\frac{e^{ikR}}{4\pi R} = -\frac{e^{ik|\mathbf{r}-\mathbf{r}\,'|}}{4\pi|\mathbf{r}-\mathbf{r}\,'|}\,.$$

3. It is easy to work out the Green functions obtained by adopting the three other ways in which the poles at $q = \pm k$ can be given infinitesimal imaginary displacements. Let ϵ be a positive infinitesimal, as usual.

(a) Suppose we displace the pole at $+k$ into the lower half-plane in q, and the pole at $-k$ into the upper half-plane. (Draw a figure.) Verify that the Green function then vanishes identically, i.e.,

$$-\frac{1}{(2\pi)^3}\lim_{\epsilon\to 0}\int d^3q\,\frac{e^{i\mathbf{q}\cdot\mathbf{R}}}{q^2-(k-i\epsilon)^2}=0.$$

(b) Similarly, suppose both poles are displaced into either the upper or the lower half-plane. (Draw the corresponding figures.) Show that the corresponding Green function is

$$-\frac{1}{(2\pi)^3}\lim_{\epsilon\to 0}\int d^3q\,\frac{e^{i\mathbf{q}\cdot\mathbf{R}}}{(q\pm i\epsilon)^2-k^2}=-\frac{\cos\,kR}{4\pi R}\,.$$

Exact formula for the scattering amplitude: Using the correct Green function, the integral equation for scattering becomes

$$\psi(\mathbf{r})=e^{i\mathbf{k}\cdot\mathbf{r}}-\frac{\lambda}{4\pi}\int d^3r'\,\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}U(r')\,\psi(\mathbf{r}').$$

We need to extract the asymptotic behavior of the wave function from this integral equation, and identify the scattering amplitude. As $r\to\infty$ (more precisely, for $kr\gg 1$), we may replace the factor $1/|\mathbf{r}-\mathbf{r}'|$ in the integrand by just r^{-1}. The implicit assumption here is that the potential $V(r')$ decays rapidly for very large values of r', so that the contribution to the integral from the region in which *both* r and r' are large is quite negligible. In the pure phase factor $\exp(ik|\mathbf{r}-\mathbf{r}'|)$, however, we need to be more careful. The general field point $\mathbf{r}$ at which the wave function is being calculated is the point at which the particle detector is located, in order to measure the flux per unit solid angle in that direction. Thus the wave vector of the scattered particle, $\mathbf{k}'$, is in the same direction as $\mathbf{r}$: that is, $k\,\mathbf{e}_r=\mathbf{k}'$. We then have

$$\begin{aligned} ik\,|\mathbf{r}-\mathbf{r}'| &= ik\left(r^2-2\mathbf{r}\cdot\mathbf{r}'+r'^2\right)^{1/2}\simeq ikr\left(1-\frac{2\mathbf{r}\cdot\mathbf{r}'}{r^2}\right)^{1/2}\\ &\simeq ikr\Big(1-\frac{\mathbf{r}\cdot\mathbf{r}'}{r^2}\Big)=ikr-ik\,\mathbf{e}_r\cdot\mathbf{r}'=ikr-i\mathbf{k}'\cdot\mathbf{r}'.\end{aligned}$$

Inserting these approximations in the integral equation for $\psi(\mathbf{r})$, we get

$$\psi(\mathbf{r})\to e^{i\mathbf{k}\cdot\mathbf{r}}-\frac{e^{ikr}}{r}\,\frac{\lambda}{4\pi}\int d^3r'\,e^{-i\mathbf{k}'\cdot\mathbf{r}'}\,U(r')\,\psi(\mathbf{r}').$$

Comparing this result with the asymptotic form of the wave function, it follows at once that the scattering amplitude is given by the formula

$$f(k,\theta)=-\frac{\lambda}{4\pi}\int d^3r'\,e^{-i\mathbf{k}'\cdot\mathbf{r}'}\,U(r')\,\psi(\mathbf{r}').$$

While this formula is *exact*, it is not of much use unless we already know the full wave function $\psi(\mathbf{r})$ at all points in space.

Scattering geometry and the momentum transfer: For scattering in any given direction, the vector

$$\hbar\,\mathbf{Q} = \hbar\,(\mathbf{k}\,' - \mathbf{k})$$

is the final momentum of the particle minus its initial momentum. That is, $\mathbf{Q}$ is the wave vector corresponding to the **momentum transfer** associated with the scattering process. Recall that the energy E is related to $k = |\mathbf{k}| = |\mathbf{k}\,'|$ according to $E = \hbar^2 k^2/(2m)$. The relation between k, Q and θ is also easy to derive. (Draw a figure.) It is easily seen that

$$Q^2 = 2k^2\,(1 - \cos\,\theta) = 4k^2\,\sin^2\,\left(\tfrac{1}{2}\theta\right), \quad \text{so that} \quad Q = 2k\,\sin\,\left(\tfrac{1}{2}\theta\right).$$

$\theta = 0$ corresponds to **forward scattering**, while $\theta = \pi$ corresponds to **backward scattering**.

Born series and the Born approximation: Going back to the inhomogeneous integral equation for the wave function, we can solve it iteratively for sufficiently small values of the 'strength' $|\lambda|$ of the potential. We find

$$\begin{aligned}
\psi(\mathbf{r}) \;&=\; e^{i\mathbf{k}\cdot\mathbf{r}} - (\lambda/4\pi)\int d^3r\,'\;\frac{e^{ik|\mathbf{r}-\mathbf{r}\,'|}}{|\mathbf{r}-\mathbf{r}\,'|}\,U(r\,')\,e^{i\mathbf{k}\cdot\mathbf{r}\,'} \\
&+\; (\lambda/4\pi)^2\int d^3r\,'\int d^3r\,''\;\frac{e^{ik|\mathbf{r}-\mathbf{r}\,'|}}{|\mathbf{r}-\mathbf{r}\,'|}\,U(r\,')\frac{e^{ik|\mathbf{r}\,'-\mathbf{r}\,''|}}{|\mathbf{r}\,'-\mathbf{r}\,''|}\,U(r\,'')\,e^{i\mathbf{k}\cdot\mathbf{r}\,''} \\
&+\; \cdots
\end{aligned}$$

The precise conditions on the potential under which this is a convergent series are discussed in textbooks on quantum mechanics, and I shall not go into this aspect here. Very roughly speaking, the series solution is valid *when the effect of the potential is weak as compared to the kinetic energy of the incident particle.* Substitution of the solution above in the formula for $f(k,\theta)$ gives an expression for the scattering amplitude as a power series in λ, called the **Born series**.

It is trivial to write down the scattering amplitude to first order in λ. All we have to do is to approximate $\psi(\mathbf{r}\,')$ in the integrand by the incident wave $\exp\,(i\mathbf{k}\cdot\mathbf{r}\,')$ itself. This is called the (first) **Born approximation**. Denoting the scattering amplitude in this approximation by $f_{\rm B}(k,\theta)$, we have

$$f_{\rm B}(k,\theta) = -(\lambda/4\pi)\int d^3r\,'\;e^{-i\mathbf{Q}\cdot\mathbf{r}\,'}\,U(r\,'), \quad \text{where} \quad \mathbf{Q} = \mathbf{k}\,' - \mathbf{k}.$$

Hence the scattering amplitude in the Born approximation is, up to a constant factor, the (three-dimensional) Fourier transform of the potential, with the momentum

transfer wave vector $\mathbf{Q}$ playing the role of the variable conjugate to $\mathbf{r}$. Now, $e^{i\mathbf{k}\cdot\mathbf{r}} = e^{i\mathbf{p}\cdot\mathbf{r}/\hbar} = \langle \mathbf{r} \,|\, \mathbf{p} \rangle$ is the position space wave function corresponding to the initial momentum eigenstate $|\mathbf{p}\rangle$ of the particle. Similarly, $e^{-i\mathbf{k}'\cdot\mathbf{r}} = e^{-i\mathbf{p}'\cdot\mathbf{r}\hbar} = \langle \mathbf{r} \,|\, \mathbf{p}' \rangle^* = \langle \mathbf{p}' \,|\, \mathbf{r} \rangle$. Using the completeness relation $\int d^3r \, |\mathbf{r}\rangle \, \langle \mathbf{r}| = I$, we see that

$$f_{\rm B}(k,\theta) = -\frac{\lambda}{4\pi} \, \langle \mathbf{p}'|U|\mathbf{p}\rangle = -\frac{m\lambda}{2\pi\hbar^2} \, \langle \mathbf{p}'|V|\mathbf{p}\rangle.$$

In other words:

- The scattering amplitude in the Born approximation is essentially the matrix element of the potential energy operator between the initial and final free-particle momentum eigenstates.

4. The angular integration in the expression for $f_{\rm B}(k,\theta)$ is easily carried out. Show that

$$f_{\rm B}(k,\theta) = -\frac{2m\lambda}{\hbar^2 \, Q} \int_0^\infty dr \, r \, \sin{(Qr)} \, V(r),$$

where Q is the magnitude of the momentum transfer vector $\mathbf{Q}$.

5. For a finite potential barrier given by

$$V(r) = \begin{cases} V_0 & \text{for} \quad r \leq a \\ 0 & \text{for} \quad r > a \end{cases}$$

show that

$$f_{\rm B}(k,\theta) = \frac{2m\lambda V_0}{\hbar^3} \, \frac{(\sin \, Qa - Qa \, \cos \, Qa)}{Q^3}.$$

Hence show that the forward scattering amplitude in the Born approximation is $f_{\rm B}(k,0) = (2m\lambda V_0 \, a^3)/(3\hbar^2)$.

6. The Yukawa potential: According to quantum field theory, the forces between elementary particles arise from the exchange of other particles which are the quanta of **gauge fields**. In the nonrelativistic limit of scattering from a static potential, such a force generically reduces to a form called the **Yukawa potential**, given by

$$V(r) = \frac{e^{-r/\xi}}{r} \, ,$$

where ξ is a positive constant with the physical dimensions of length. It represents the 'range' of the potential, and is essentially the Compton wavelength of the exchanged particle. The latter is given by $h/(\mu c)$, where μ is the (rest) mass of the exchanged particle. The functional form of the Yukawa potential arises in other physical contexts as well. For instance, the **screened Coulomb potential** in a dielectric medium has

the Yukawa form.

Show that, in the Born approximation, the scattering amplitude for the Yukawa potential is given by

$$f_{\rm B}(k,\theta) = -\frac{2m\lambda\xi^2}{\hbar^2(Q^2\xi^2+1)},$$

where Q is the magnitude of the momentum transfer vector, as usual. Hence show that the total cross-section in the Born approximation is

$$\sigma_{\rm B}(E) = \frac{4\pi^2 m\lambda^2\xi^2}{E\hbar^2}\left(\frac{4mE\xi^2+\hbar^2}{8mE\xi^2+\hbar^2}\right).$$

For very large values of the incident energy E, the total cross-section falls off like $1/E$. This is a general feature of scattering from a wide class of potentials. (This class, however, does *not* include the Coulomb potential!)

The Coulomb potential corresponds to the $\xi \to \infty$ limit of the Yukawa potential.[43] It follows that the scattering amplitude in the Born approximation is now $f_{\rm B}(k,\theta) = -(2m\lambda)/(\hbar^2\,Q^2)$. The differential cross-section is therefore

$$\left[\frac{d\sigma}{d\Omega}\right]_{\rm B} = \frac{\lambda^2}{16E^2}\,{\rm cosec}^4\left(\tfrac{1}{2}\theta\right).$$

This is precisely the famous **Rutherford scattering formula**. There are several remarkable features about this result:

(i) It is also the *exact* expression for the differential cross-section for scattering in the Coulomb potential λ/r, because all the higher-order corrections to the first Born approximation happen to vanish in this instance.

(ii) Moreover, the expression is actually independent of Planck's constant, and is exactly the same as the differential scattering cross-section of *classical* particles in a Coulomb potential.

(iii) The differential cross-section for *forward* scattering ($\theta = 0$) diverges. So does the *total* cross-section, as the divergence at $\theta = 0$ is not integrable.

All these 'minor miracles' and drawbacks are related to the long-range nature of the Coulomb potential. It turns out that the assumption of an initial *plane* wave state ($\psi_{\rm inc} = e^{i\mathbf{k}\cdot\mathbf{r}}$) is incompatible with a potential that decays as slowly as $1/r$. The scattering problem has to be re-done in this case. The Schrödinger equation in the

[43] If we set $\xi = h/(\mu c)$, then $\xi \to \infty$ corresponds to $\mu = 0$. Recall that electromagnetic interactions are mediated by the exchange of photons, which are massless quanta.

presence of a Coulomb potential can be solved exactly, using parabolic cylindrical coordinates. The solution involves confluent hypergeometric functions. The initial state is found to be a plane wave that is amplitude-modulated by a factor of $1/r$, and phase-modulated by a term involving $\ln r$.

13 The wave equation

The last standard equation of mathematical physics to which we turn our attention in this course is the **wave equation**. To be specific, we'll consider the fundamental Green function of the wave operator, that corresponds to the **causal, retarded solution** of the wave equation. This is the solution that is of direct physical significance for the propagation of signals.

Formal solution for the causal Green function: The wave equation for a scalar function (or signal) $f(\mathbf{r},t)$ is given by

$$\Big(\frac{1}{c^2}\frac{\partial^2}{\partial t^2}-\nabla^2\Big)f(\mathbf{r},t)=g(\mathbf{r},t),$$

where $g(\mathbf{r},t)$ is a specified function of space and time that represents the *source* of the signal, and c is the speed of the signal. Although I have used c to denote this speed, what follows is not restricted to the propagation of light in free space. We'll first consider, formally, wave propagation in a general d-dimensional Euclidean space of infinite extent,[44] and then work out the explicit solutions for the cases $d=1,2$ and 3 that are of direct physical interest. Subsequently, I'll also comment on what happens when $d>3$.

The crucial difference between the wave equation and Poisson's equation, is, of course, the relative minus sign between the time derivatives and the Laplacian. This makes all the difference in the world (literally!), and is a reflection of the fact that the spacetime we consider has d space-like dimensions and 1 time-like dimension. From the mathematical point of view, the wave equation is a **hyperbolic** partial differential equation, while Poisson's equation is an **elliptic** equation, and the diffusion equation is a **parabolic** equation.

The particular integral of the inhomogeneous wave equation is given by the integral representation

$$f(\mathbf{r},t)=\int d^dr'\int_{-\infty}^{\infty}dt'\,G(\mathbf{r},t\,;\,\mathbf{r}',t')\,g(\mathbf{r}',t'),$$

where d^dr' is the volume element in d dimensions. The Green function G satisfies the equation

$$\Big(\frac{1}{c^2}\frac{\partial^2}{\partial t^2}-\nabla^2\Big)G(\mathbf{r},t\,;\,\mathbf{r}',t')=\delta^{(d)}(\mathbf{r}-\mathbf{r}')\,\delta(t-t'),$$

where $\delta^{(d)}$ denotes the d-dimensional δ-function. In physical terms, this Green function represents the signal at time t at the point $\mathbf{r}$, arising when a sharply-pulsed point source of unit strength is switched on at the point $\mathbf{r}'$ at the instant of time t'. We are interested in the Green function that satisfies natural boundary conditions, i.e., $G\to 0$

[44] So that ∇^2 stands for the Laplacian in d-dimensional Euclidean space.

as $|\mathbf{r}-\mathbf{r}\,'|\to\infty$. Further, we seek the *causal* Green function that vanishes identically for all $t<t\,'$, so that there is no signal anywhere before the source is switched on. This is the **principle of causality**. Hence G must satisfy the conditions

$$G=0 \;\text{ and }\; \partial G/\partial t=0 \;\text{ for all } t<t\,',\;\text{ at all points.}$$

As a consequence of the translational invariance in space and time of the differential equation as well as the boundary and initial conditions, we expect to find that the Green function has the form

$$G(\mathbf{r},t\,;\,\mathbf{r}\,',t\,')\equiv G(\mathbf{r}-\mathbf{r}\,',\,t-t\,')=\theta(t-t\,')\,K(\mathbf{r}-\mathbf{r}\,',\,t-t\,').$$

(In a region of *finite* extent, in the presence of boundary conditions at finite values of r, the dependence of G on $\mathbf{r}$ and $\mathbf{r}\,'$ cannot be reduced in general to a dependence on the difference $\mathbf{r}-\mathbf{r}\,'$ alone.) The quantity $K(\mathbf{r}-\mathbf{r}\,',\,t-t\,')$ is sometimes referred to as the **propagator**. As you will see, causality imposes an even stronger constraint. Since the propagation of the signal occurs with a finite speed c, a signal emanating from the point $\mathbf{r}\,'$ at time $t\,'$ cannot reach the point $\mathbf{r}$ till time $t\,'+|\mathbf{r}-\mathbf{r}\,'|/c$. This feature will also emerge automatically in the solution for G.

It is obviously convenient to shift variables from $\mathbf{r}$ and t to

$$\mathbf{R}=\mathbf{r}-\mathbf{r}\,' \quad\text{and}\quad \tau=t-t\,',$$

respectively. Then

$$\Big(\frac{1}{c^2}\frac{\partial^2}{\partial\tau^2}-\nabla^2\Big)G(\mathbf{R}\,,\,\tau)=\delta^{(d)}(\mathbf{R})\,\delta(\tau),$$

where ∇^2 now stands for the Laplacian operator with respect to $\mathbf{R}$. We must impose the conditions $G=0$ and $\partial G/\partial\tau=0$ for all $\tau<0$, and the boundary condition $G\to 0$ as $R\to\infty$. Now express G as a Fourier integral, according to

$$G(\mathbf{R}\,,\,\tau)=\int\frac{d^dk}{(2\pi)^d}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,e^{i(\mathbf{k}\cdot\mathbf{R}-\omega\tau)}\;\widetilde{G}(\mathbf{k},\omega).$$

Observe that we use a Fourier transform in both space *and* time, rather than a Fourier transform with respect to $\mathbf{r}$ and a Laplace transform with respect to t. This is helpful in imposing the initial conditions on G, as you will see. The inverse relation is of course

$$\widetilde{G}(\mathbf{k},\omega)=\int d^d\mathbf{R}\int_{-\infty}^{\infty}d\tau\,e^{-i(\mathbf{k}\cdot\mathbf{R}-\omega\tau)}\,G(\mathbf{R}\,,\,\tau).$$

Further,

$$\delta^{(d)}(\mathbf{R})\,\delta(\tau)=\int\frac{d^dk}{(2\pi)^d}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,e^{i(\mathbf{k}\cdot\mathbf{R}-\omega\tau)}.$$

Using these expressions in the partial differential equation satisfied by $G(\mathbf{R}\,,\,\tau)$, we get

$$\int \frac{d^dk}{(2\pi)^d} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, e^{i(\mathbf{k}\cdot\mathbf{R}-\omega\tau)} \left\{(\omega^2 - c^2k^2)\,\widetilde{G}(\mathbf{k},\omega) + c^2\right\} = 0.$$

But the set of functions $\{e^{i(\mathbf{k}\cdot\mathbf{R}-\omega\tau)}\}$, where ω and the Cartesian components of $\mathbf{k}$ run over all real values, forms a complete orthonormal basis in the space of integrable functions of τ and $\mathbf{R}$. Therefore the expression in curly brackets must vanish, for every value of ω and the components of $\mathbf{k}$. Hence

$$\widetilde{G}(\mathbf{k},\omega) = -\,\frac{c^2}{(\omega^2 - c^2k^2)}\,, \quad \text{where} \quad k^2 = |\mathbf{k}|^2 = k_1^2 + \ldots + k_d^2\,.$$

As expected, the introduction of the Fourier transform has converted the partial differential equation for G into a trivially-solved algebraic equation for its Fourier transform $\widetilde{G}(\mathbf{k},\omega)$. Inverting the Fourier transforms, we obtain the formal solution for $G(\mathbf{R}\,,\,\tau)$, namely,

$$G(\mathbf{R}\,,\,\tau) = -c^2 \int \frac{d^dk}{(2\pi)^d}\, e^{i\mathbf{k}\cdot\mathbf{R}} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, \frac{e^{-i\omega\tau}}{(\omega^2 - c^2k^2)}\,.$$

We now encounter a familiar difficulty: the formal solution above does not make sense as it stands! The integral over ω diverges because the integrand has poles at $\omega = -ck$ and $\omega = ck$ that lie *on* the path of integration.

The difficulty is overcome, as usual, by an appropriate $i\epsilon$-prescription that also takes care of the physical requirement of causality. This condition enables us to select the correct Green function unambiguously. Once the poles are displaced off the real axis in the ω-plane, the integration over ω can be carried out using contour integration. Let Ω be a large positive constant. Consider a closed contour C_+ [respectively, C_-] comprising a straight line from $-\Omega$ to $+\Omega$ along the real axis in the ω-plane, and a semicircle of radius Ω that takes us back from $+\Omega$ to $-\Omega$ in the upper [respectively, lower] half-plane. The limit $\Omega \to \infty$ is to be taken after the contour integral is evaluated. *Provided* the contribution from the semicircle vanishes in the limit $\Omega \to \infty$, the original line integral from $-\infty$ to $+\infty$ over ω is guaranteed to be precisely equal to the integral over the closed contour.[45]

Now, for $\tau < 0$, this semicircle *must* lie in the *upper* half-plane in ω, because it is only in this region that the factor $e^{-i\omega\tau}$ in the integrand vanishes exponentially as

[45] You will recall that we've already used this sort of contour completion more than once—for instance, in the derivation of dispersion relations for the generalized susceptibility in linear response theory, and of the Green function corresponding to outgoing spherical waves for the Helmholtz operator.

$\Omega \to \infty$. The addition of the semicircle to the contour would then simply add a vanishing contribution to the original line integral that we want to evaluate. Therefore, provided no singularities of the integrand lie on the real axis or in the *upper* half-plane in ω, the integral over the contour C_+ is guaranteed to vanish identically for $\tau < 0$. But this is precisely what is required by causality: namely, that $G(\mathbf{R}\,,\,\tau)$ be equal to 0 for all $\tau < 0$.

On the other hand, for $\tau > 0$, we do expect to have a signal that does not vanish identically. But now the semicircle closing the contour *must* lie in the *lower* half-plane, because it is only then that the factor $e^{-i\omega\tau}$ in the integrand vanishes exponentially as $\Omega \to \infty$; and hence so does the contribution from the semicircle to the integral over the contour C_-. Therefore, provided all the singularities of the integrand are in the *lower* half-plane, all our requirements are satisfied. This is ensured by displacing each of the poles of the integrand at $\omega = -ck$ and $\omega = +ck$ by an infinitesimal *negative* imaginary quantity $-i\epsilon$ where $\epsilon > 0$, and then passing to the limit $\epsilon \to 0$ after the integral is evaluated. Equivalently, we may replace ω by $\omega + i\epsilon$ in the denominator of the integrand, and take the limit $\epsilon \to 0$ after carrying out the integration.

The causal Green function we seek is therefore given by

$$G(\mathbf{R}\,,\,\tau) = -c^2 \lim_{\epsilon\to 0} \int \frac{d^dk}{(2\pi)^d}\, e^{i\mathbf{k}\cdot\mathbf{R}} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, \frac{e^{-i\omega\tau}}{(\omega+i\epsilon)^2 - c^2k^2},$$

where ϵ is a positive infinitesimal. But, as discussed above,

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{e^{-i\omega\tau}}{(\omega+i\epsilon)^2 - c^2k^2} &= \lim_{\Omega\to\infty} \int_{-\Omega}^{\Omega} \frac{d\omega}{2\pi} \frac{e^{-i\omega\tau}}{(\omega+i\epsilon)^2 - c^2k^2} \\ &= \lim_{\Omega\to\infty} \int_{C_\pm} \frac{d\omega}{2\pi} \frac{e^{-i\omega\tau}}{(\omega+i\epsilon)^2 - c^2k^2}\,.\end{aligned}$$

As explained in the foregoing, we must use C_+ for $\tau < 0$ and C_- for $\tau > 0$.

1. Show that

$$G(\mathbf{R}\,,\,\tau) = c\,\theta(\tau) \int \frac{d^dk}{(2\pi)^d}\, \frac{\sin\, c\tau k}{k}\, e^{i\mathbf{k}\cdot\mathbf{R}}.$$

Note how the step function $\theta(\tau)$ required by causality emerges automatically in the solution for $G(\mathbf{R}\,,\,\tau)$.

2. Each of the two poles of the integrand in the formal solution for $G(\mathbf{R}\,,\,\tau)$ the we started with can be displaced so as to lie either in the upper or lower half-plane. This leads to four possible ways of making the divergent integral finite. The particular $i\epsilon$-prescription we have used above is tailored to ensure that the correct causal solution is picked up from among the set of possible solutions. Find the other solutions.

We are now in a position to consider the solutions for different values of d. In order to make this explicit, I'll denote the Green function by $G^{(d)}(\mathbf{R}\,,\,\tau)$ instead of $G(\mathbf{R}\,,\,\tau)$, from now on.

3. The solution in $(1+1)$ dimensions: The case of one spatial dimension ($d = 1$) is somewhat distinct from the others, and simpler too. Note that the symbol k in the factor $(\sin c\tau k)/k$ above stands for $|\mathbf{k}|$. When $d = 1$, therefore, we should remember to write $|k|$ instead of just k in this factor. Further, $\mathbf{k}\cdot\mathbf{R}$ is just kX in this case, where $X = x - x'$. Therefore

$$G^{(1)}(X,\tau) = c\;\theta(\tau)\int_{-\infty}^{\infty}\frac{dk}{2\pi}\,\frac{\sin\,c\tau|k|}{|k|}\,e^{ikX} = c\;\theta(\tau)\int_{-\infty}^{\infty}\frac{dk}{2\pi}\,\frac{\sin\,c\tau k}{k}\,e^{ikX}.$$

It is obvious from this expression that $G^{(1)}(X,\tau)$ is a symmetric function of X, i.e.,

$$G^{(1)}(-X,\tau) = G^{(1)}(X,\tau).$$

(Change the variable of integration from k to $-k$, and the result follows.) As you will see shortly, $G^{(1)}(X,\tau)$ is actually a function of $|X|$. Show that

$$G^{(1)}(X,\tau) = \tfrac{1}{4}c\,\theta(\tau)\;\left[\varepsilon(c\tau+X)+\varepsilon(c\tau-X)\right],$$

where $\varepsilon(x) = 1$ for $x > 0$, and -1 for $x < 0$. Simplify this result to get

$$G^{(1)}(X,\tau) = \tfrac{1}{2}c\,\theta(\tau)\,\theta(c\tau-|X|).$$

The factor $\theta(c\tau-|X|)$ ensures that the signal arising from the source point x' at time t' does not reach any point x until time $t'+|x-x'|/c$, as required by causality and the finite speed of signal propagation. The presence of this step function makes the other step function, $\theta(\tau)$, redundant from a physical point of view. However, it is present in the formal mathematical solution for the quantity $G^{(1)}(X,\tau)$.

Another aspect of the solution is also noteworthy. An observer at x does not receive a *pulsed* signal, even though the sender sent out such a signal. (Recall that the initial disturbance was a Dirac δ-function.) In fact, the signal received at any x *persists* thereafter for all time. And it does so without diminishing in strength, a feature that is unique to the case $d = 1$.

Let's return to the general formula for $G(\mathbf{R}\,,\,\tau)$ in more than one spatial dimension, i.e., $d \geq 2$. The integrand on the right-hand side is a scalar, and so is the volume element d^dk. Further, the region of integration (namely, all of $\mathbf{k}$-space) is invariant under rotations of the coordinate axes. Hence $G^{(d)}(\mathbf{R}\,,\,\tau)$ is a scalar, i.e., it is unchanged under rotations of the spatial coordinate axes about the origin. This remains true for all integer values of $d \geq 2$. As a result of this rotational invariance, $G^{(d)}(\mathbf{R}\,,\,\tau)$ is actually a function of R and τ (where $R = |\mathbf{R}|$, as usual). This means that we can choose the

orientation of the axes in **k**-space according to our convenience, without affecting the result. This fact is of help in evaluating the integral. We'll write the Green function as $G^{(d)}(R,\tau)$, henceforth.

4. The solution in $(2+1)$ dimensions: Let's now consider the case $d=2$. It is evidently most convenient to work in plane polar coordinates in **k**-space, choosing one of the axes (say, the k_1-axis) along the vector **R**. Then

$$\begin{aligned} G^{(2)}(R,\tau) &= c\ \theta(\tau) \int_0^\infty \frac{k\,dk}{(2\pi)^2}\,\frac{\sin\,c\tau k}{k} \int_0^{2\pi} d\varphi\ e^{ikR\,\cos\,\varphi} \\ &= \frac{c\,\theta(\tau)}{2\pi} \int_0^\infty dk\ \sin\,(c\tau k)\,J_0(kR), \end{aligned}$$

on using the representation $\int_0^{2\pi} d\varphi\, e^{iz\,\cos\,\varphi} = 2\pi J_0(z)$ for the Bessel function (of the first kind) of order zero. The integral on the right-hand side is a little tricky to evaluate. As you know, $J_0(kR)$ decays quite slowly, like $k^{-1/2}$, as $k\to\infty$. However, it is an oscillatory function that changes sign. So is the factor $\sin\,(c\tau k)$. As a result of the partial cancellation of positive and negative contributions, it turn out that the definite integral above has a finite value. (You should be reminded of the case of the Dirichlet integral.) Here is a quick way to find it. Since $\sin\,(c\tau k)$ is the imaginary part of $e^{ic\tau k}$, and $J_0(kR)$ is real for real values of the argument kR, we have

$$G^{(2)}(R,\tau) = \frac{c\,\theta(\tau)}{2\pi}\mathrm{Im}\,\Big\{\int_0^\infty dk\,e^{ic\tau k}\,J_0(kR)\Big\}.$$

But we may regard this integral as the *analytic continuation*, to $s=-ic\tau$, of the Laplace transform of $J_0(kR)$, which is given by

$$\int_0^\infty dk\,e^{-sk}\,J_0(kR) = \frac{1}{\sqrt{s^2+R^2}}\,.$$

This analytic continuation can be justified properly, but I'll not digress to do so here. It is equivalent to putting in a decaying exponential factor (i.e., a regulator) like $e^{-\lambda k}$ in the integrand, and then passing to the limit $\lambda\to 0$ after the integral is evaluated. Show from the result above that

$$\mathrm{Im}\,\Big\{\int_0^\infty dk\,e^{ic\tau k}\,J_0(kR)\Big\} = \begin{cases} 0 & \text{for}\quad c^2\tau^2 < R^2 \\ 1/\sqrt{c^2\tau^2-R^2} & \text{for}\quad c^2\tau^2 > R^2. \end{cases}$$

A little care is needed in getting the overall sign right in the result. Note that the Laplace transform, regarded as an analytic function of s, has branch points at $s=\pm iR$, and a branch cut running between these points. Identify the phases of the analytic function $(s^2+R^2)^{-1/2}$ at different points in the complex s-plane, and use this information to arrive at the result quoted.

Since we are only interested in non-negative values of τ and R, the condition $c^2\tau^2 > R^2$ yields a factor $\theta(c\tau - R)$. We get, finally,

$$G^{(2)}(R,\tau) = \frac{c\ \theta(\tau)}{2\pi}\frac{\theta(c\tau - R)}{\sqrt{c^2\tau^2 - R^2}}\,.$$

The sharply-pulsed signal emanating from $\mathbf{r}\,'$ at time $t\,'$ thus reaches any point $\mathbf{r}$ only at time $t\,' + |\mathbf{r} - \mathbf{r}\,'|/c$, in accordance with causality and the finite velocity of propagation of the disturbance. But once again, the signal received at any field point $\mathbf{r}$ is no longer a sharply pulsed one: it persists for all $t > t\,' + |\mathbf{r} - \mathbf{r}\,'|/c$, although its strength slowly decays as t increases, like t^{-1} at very long times. Thus, both in $d = 1$ and in $d = 2$, there is an *after-effect* even for a sharply-pulsed initial signal emanating from the source.

5. The solution in $(3+1)$ **dimensions**: Something entirely different happens in the most important case of three-dimensional space. We have

$$G^{(3)}(R,\tau) = c\,\theta(\tau)\int \frac{d^3\mathbf{k}}{(2\pi)^3}\,\frac{\sin\,c\tau k}{k}\,e^{i\mathbf{k}\cdot\mathbf{R}}\,.$$

It is immediately clear that we should use spherical polar coordinates $(k\,,\,\theta\,,\,\varphi)$, and exploit rotational invariance to choose the polar axis in $\mathbf{k}$-space along the vector $\mathbf{R}$. Show that this leads to

$$G^{(3)}(R,\tau) = \frac{c\,\theta(\tau)}{2(2\pi)^2 R}\int_{-\infty}^{\infty} dk\,\big\{\cos\,(c\tau - R)k - \cos\,(c\tau + R)k\big\}.$$

Each cosine in the integrand can be replaced by the corresponding exponential, since the contribution from the sine function vanishes by symmetry. Hence show that

$$G^{(3)}(R,\tau) = \frac{\theta(\tau)\,\delta(\tau - R/c)}{4\pi R}.$$

Note that the other δ-function, namely, $\delta(c\tau + R)$, can be dropped, because we are only concerned with non-negative values of τ and R.

The crucial point about the solution in three-dimensional space is this: If the source pulse is a δ-function impulse emanating from $\mathbf{r}\,'$ at time $t\,'$, the signal at any field point $\mathbf{r}$ is *also* a δ-function pulse that reaches (and passes) this point at precisely the instant $t\,' + |\mathbf{r} - \mathbf{r}\,'|/c$. Hence there is no after-effect that lingers on at $\mathbf{r}$, in stark contrast to the situation in $d = 1$ and $d = 2$. Moreover, the amplitude of the pulse at $\mathbf{r}$ drops with the distance from the source like $1/R$, in exactly the way the Coulomb potential falls off. *These features are unique to three-dimensional space.*

Another interesting reduction takes place in the case $d = 3$. Formally, if the limit $c \to \infty$ is taken in the wave equation, the wave operator reduces to the negative of the

Laplacian operator. We might therefore expect the solution for $G^{(3)}(R,\tau)$ to reduce to the corresponding Green function for $-\nabla^2$. And indeed it does so, because the latter Green function is precisely $1/(4\pi R)$ in three spatial dimensions.

Retarded solution of the wave equation: Substituting the expression obtained above for $G^{(3)}(R,\tau)$ in the formal solution of the inhomogeneous wave equation in three dimensions yields the particular integral in the absence of boundaries (or natural boundary conditions in infinite space):

$$f(\mathbf{r},t) = \frac{1}{4\pi}\int \frac{d^3r\,'}{|\mathbf{r}-\mathbf{r}\,'|}\,\Big[g(\mathbf{r}\,',t\,')\Big]_{\rm ret},$$

where

$$[g(\mathbf{r}\,',t\,')]_{\rm ret} \stackrel{\rm def.}{=} g\Big(\mathbf{r}\,',\, t-\frac{|\mathbf{r}-\mathbf{r}\,'|}{c}\Big).$$

This is called the *causal* or **retarded solution** of the wave equation.

6. There is an immediate application to electromagnetism in the Lorenz gauge. In this gauge, the electromagnetic scalar and vector potentials in free space satisfy the inhomogeneous wave equations

$$\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} - \nabla^2\phi = \frac{\rho(\mathbf{r},t)}{\epsilon_0} \quad \text{and} \quad \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} - \nabla^2\mathbf{A} = \mu_0\,\mathbf{j}(\mathbf{r},t).$$

The retarded solutions to these equations with natural boundary conditions are therefore given by

$$\phi(\mathbf{r},t) = \frac{1}{4\pi\epsilon_0}\int \frac{d^3r\,'}{|\mathbf{r}-\mathbf{r}\,'|}\,\Big[\rho(\mathbf{r}\,',t\,')\Big]_{\rm ret}$$

and

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{4\pi}\int \frac{d^3r\,'}{|\mathbf{r}-\mathbf{r}\,'|}\,\Big[\mathbf{j}(\mathbf{r}\,',t\,')\Big]_{\rm ret}\,.$$

Hence write down the formal solutions for the electric field $\mathbf{E}(\mathbf{r},t)$ and magnetic field $\mathbf{B}(\mathbf{r},t)$ arising from arbitrary sources $\rho(\mathbf{r},t)$ and $\mathbf{j}(\mathbf{r},t)$, using the relations $\mathbf{E} = -\partial\mathbf{A}/\partial t - \nabla\phi$ and $\mathbf{B} = \nabla\times\mathbf{A}$. You must bear in mind that the solutions given above for ϕ and $\mathbf{A}$ are, of course, only valid in the Lorenz gauge. But the expressions obtained for the physical fields $\mathbf{E}$ and $\mathbf{B}$ are gauge-invariant, and hold good in any gauge.

Remarks on propagation in spatial dimensions $d > 3$: The results derived above essentially imply that the basic, linear wave equation permits the propagation of sharp pulses in three-dimensional space, but not in one- or two-dimensional space. Here are some comments on what happens in a general number of dimensions $d > 3$.

Interestingly enough, it turns out that the propagation of sharp signals is possible in all *odd*-dimensional spaces with $d \geq 3$, while it fails for all *even* values of d. This is yet another manifestation of the fundamental differences that exist between Euclidean spaces of even and odd dimensionalities, respectively. If a sharp δ-function pulse is emitted at the origin at $t = 0$, then:

(i) The signal received at any point $\mathbf{r}$ is sharply pulsed, arriving and passing on at time $t_0 + |\mathbf{r} - \mathbf{r}\,'|/c$ with no after-effect, in spaces of dimension $d = 3, 5, \ldots$.

(ii) In contrast, the signal lingers on for all $t > t_0 + |\mathbf{r} - \mathbf{r}\,'|/c$ in spaces of dimension $d = 2, 4, \ldots$. We have already seen an example of such an after-effect in the case $d = 2$.

(iii) There is, however, one feature that is quite unique to $d = 3$: this is the only case in which the original δ-function pulse is transmitted without any distortion, namely, as a pure δ-function pulse.

There is an elegant and powerful way to solve the general problem. It is based on the **relativistic invariance** of both the wave operator and the solution sought, i.e., on the fact that these remain invariant under rotations of the coordinate axes as well as boosts to other inertial frames. (Here we have identified c with the speed of light in a vacuum). I shall merely write down the final answer here. The causal Green function turns out to be proportional to a derivative of a δ-function, according to

$$G^{(d)}(R, \tau) \propto \frac{d^{(d-3)/2}}{d\xi^{(d-3)/2}}\, \delta(\xi^2), \quad \text{where} \quad \xi = (c^2\tau^2 - R^2)^{1/2}.$$

The order of the derivative, $\frac{1}{2}(d-3)$, is an integer when d is an odd number. The Green function remains a sharply-pulsed quantity in this case, although it is only for $d = 3$ that you get just a δ-function. For larger odd values of $d\,(= 5, 7, \ldots)$, the fundamental solution is given by higher and higher derivatives of the δ-function. These are increasingly singular quantities. When d is an even integer, on the other hand, the solution is a **fractional derivative** of a δ-function. Fractional derivatives are *nonlocal* objects, defined in terms of suitable transforms (such as the Fourier transform). This is how the extended nature of the Green function arises in the case of wave propagation in even-dimensional spaces, leading to the after-effects mentioned earlier.

There is another interesting connection between the Green functions in spaces of different dimensions. The solution in $(d + 2)$ spatial dimensions is related to that in d space dimensions by

$$G^{(d+2)} = -\frac{1}{2\pi d}\, \frac{\partial^2 G^{(d)}}{\partial \xi^2}\,.$$

This relationship shows how the solutions in $d = 5, 7, \ldots$ can be generated from the solution in $d = 3$, while those in $d = 4, 6, \ldots$ can be generated from that in $d = 2$.

Dispersion and nonlinearity: Finally, I mention that there are two important additional aspects of wave or signal propagation that can be adjusted so as to modify the fundamental solution considered here. The first is **dispersion**, which occurs because waves of different wavelengths propagate with different speeds in a medium. The corresponding dispersion relation (or frequency-wave number relationship) can be quite complicated. The second aspect is **nonlinearity**. While the simple wave equation we have considered here is linear in the signal $f(\mathbf{r}, t)$, physical situations often call for nonlinear equations. The *interplay between dispersion and nonlinearity* can be extremely intricate and interesting, and a vast variety of new phenomena can arise as a result. Among these are the so-called **solitary waves** and **propagating solitons** which represent very robust pulsed disturbances. These entities comprise a whole subject in their own right.

14 The rotation group and all that

Rotations of the coordinate axes: Consider n-dimensional Euclidean space, where $n \geq 2$. A **rotation** of the coordinate axes about the origin of coordinates takes a general point $\mathbf{r}$ to another point $\mathbf{r}\,'$, such that (i) the origin remains unchanged, and (ii) the distance between any two points remains unchanged. Therefore each rotation of the coordinate axes about the origin is a linear, homogeneous transformation of the Cartesian coordinates. Such a transformation is specified by a $(n \times n)$ **orthogonal matrix** R, as you'll see shortly. That is, R satisfies the condition $R\,R^{\mathrm{T}} = I$, where the superscript T denotes the transpose, and I is the unit matrix. Here's one of the simplest examples of a rotation matrix in 3-dimensional Euclidean space. The matrix corresponding to a rotation of the coordinate axes about the origin, in the xy-plane, and through an angle ψ, is given by

$$R(\psi) = \begin{pmatrix} \cos\,\psi & \sin\,\psi & 0 \\ -\sin\,\psi & \cos\,\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is easy to check that $R(\psi)$ is an orthogonal matrix. We could also have said, "a rotation about the z-axis through an angle ψ", because it is a rotation in 3-dimensional space. In general, however, the correct way to specify a rotation in any number of dimensions is to specify the *plane* in which the rotation takes place, rather than the *axis* about which it occurs. This is because no such axis may exist in general, although in $n = 3$ it so happens that a unique axis of rotation always exists for every rotation. Basically, this is because the number of independent, mutually orthogonal axes $(= n)$ becomes equal to the number of independent, mutually orthogonal planes $(= \frac{1}{2}n(n-1))$ only for $n = 3$.

Under a rotation, the components x_i of a point $\mathbf{r}$ change to the components $x\,'_i$ of the vector $\mathbf{r}\,'$, given by $x\,'_i = R_{ij}\,x_j$, where the indices run from 1 to n. (Summation over repeated indices is to be understood.) The definition of *any* vector $\mathbf{a}$ now follows: The ordered n-tuple $\mathbf{a} = (a_1\,,\,a_2\,,\,\ldots\,,\,a_n)$ is a vector if, under a rotation R of the coordinate axes, the new components are given by $a\,'_i = R_{ij}\,a_j$. Tensors of rank $2, 3, \ldots$ are sets of quantities that have transformation properties generalizing that for a vector. For example, tensors of rank 2 and 3 transform like $T\,'_{ij} = R_{ik}\,R_{jl}\,T_{kl}$ and $S\,'_{ijk} = R_{il}\,R_{jm}\,R_{kn}\,S_{lmn}$, respectively. The transformation rule for a tensor of rank 2 is of special interest. It can be written as

$$T\,'_{ij} = R_{ik}\,T_{kl}\,(R^{\mathrm{T}})_{lj} = R_{ik}\,T_{kl}\,(R^{-1})_{lj} = (RTR^{-1})_{ij}\,,$$

because the orthogonality condition on R implies that $R^{\mathrm{T}} = R^{-1}$. But we may also regard components T_{ij} of a tensor of rank 2 in n dimensions as the elements of a $(n \times n)$ matrix T. Thus the transformation rule can be written in the compact form $T\,' = R\,T\,R^{-1}$. In other words, $T\,'$ is obtained from T by a similarity transformation

involving R.

Rotations of the coordinated axes (equivalently, the matrices representing them) in n dimensions form the **rotation group**: Two rotations in succession are equivalent to a single 'resultant' rotation; no rotation at all corresponds to the identity element of the group; and for every rotation there is an 'inverse' rotation that takes us back to the original orientation of the axes.

Orthogonality of rotation matrices: It is now very easy to see why an arbitrary rotation of the coordinate axes about the origin is specified by an orthogonal matrix. Under such a rotation, the distance from the origin to any point remains unchanged. Therefore $r'^2 = r^2$, or

$$x'_i\, x'_i = x_j\, x_j = \delta_{jk}\, x_j\, x_k\,.$$

But we also have $x'_i = R_{ij}\, x_j$. Therefore

$$x'_i\, x'_i = R_{ij}\, x_j\, R_{ik}\, x_k = (R^{\mathrm{T}})_{ji}\, R_{ik}\, x_j\, x_k = (R^{\mathrm{T}} R)_{jk}\; x_j\, x_k\,.$$

The two expressions for r'^2 must be equal to each other for every point in space. Therefore we must have

$$(R^{\mathrm{T}} R)_{jk} = \delta_{jk}\,, \quad \text{or} \quad R^{\mathrm{T}}\, R = I.$$

For finite-dimensional square matrices, the left and right inverses are the same. Hence $R^{\mathrm{T}}\, R = I \Rightarrow R\, R^{\mathrm{T}} = I$.

Proper and improper rotations: As we have just seen, a rotation of the coordinate axes about the origin is specified (in n-dimensional Euclidean space) by an $(n \times n)$ orthogonal matrix. These matrices form a group, denoted by $O(n)$. Let R be a rotation matrix, i.e., a matrix whose elements tell you what linear combinations of the old coordinates yield the new coordinates. The orthogonality condition $R\, R^{\mathrm{T}} = I$ on the matrix R implies that $(\det\, R)^2 = 1$. Therefore $\det\, R = \pm 1$.

Rotations for which $\det\, R = +1$ are called continuous or **proper rotations**. They are obtainable "continuously from the identity transformation"—that is, *they can be built up by a succession of infinitesimal rotations, starting from the identity transformation* (or no rotation at all). For this reason, the set of proper rotations is called the **connected component** of the rotation group. Proper rotations constitute a group of their own, denoted by $SO(n)$. (The 'S' stands for *special*, which means 'unimodular' or 'with unit determinant', in this context.) This is a subgroup of $O(n)$. Proper rotations preserve the orientation or *handedness* of the coordinate system. That is, after a proper rotation, a right-handed coordinate system remains right-handed, and a left-handed coordinate system remains left-handed.

In contrast, transformations with det $R = -1$ are called discontinuous or **improper rotations**. They cannot be built up continuously from the identity transformation: in general, they involve proper rotations *together with* reflections, such that a right-handed coordinate system transforms to a left-handed one or vice versa. Examples of such orientation-reversing transformations in n dimensions are:

(i) **Reflection** about any plane in space. (The plane need not be one of the planes normal to the Cartesian axes.) In three dimensions, for example, a reflection about the yz-plane corresponds to a transformation under which $x \mapsto -x$, $y \mapsto y$, $z \mapsto z$.

(ii) The **parity transformation** $\mathbf{r} \mapsto -\mathbf{r}$, i.e., $x_i \mapsto -x_i$, , *provided* n *is odd.*[46] In three dimensions, for example, the transformation $x \mapsto -x$, $y \mapsto -y$, $z \mapsto -z$ is an improper rotation, while reversing the signs of any *two* of the three coordinates is actually a proper transformation: The determinant of the corresponding matrix remains equal to $+1$.

Note also that *improper rotations cannot form a subgroup of their own*, because they do not include the identity transformation. (Moreover, the product of two matrices, each with determinant equal to -1, is a matrix with determinant equal to $+1$.) The group $O(n)$ is thus made up of 2 *disjoint* sets: $(n \times n)$ orthogonal matrices with determinant equal to $+1$, and that constitute the group $SO(n)$; and $(n \times n)$ orthogonal matrices with determinant equal to -1. These are obtained by multiplying the elements of $SO(n)$ by the matrix corresponding to a parity or reflection transformation with determinant $= -1$.

The most important aspect of rotations in $n \geq 3$ dimensions is the following: Successive rotations do not *commute* with each other, unless they are rotations in the same plane. In other words, the net result of two successive rotations depends on the order in which the two are carried out. This **non-commutativity** is crucial to the understanding of rotations. It has truly profound consequences for the way the physical universe is. The rotation group in every dimension $n \geq 3$ is a noncommutative or **non-abelian** group.

Generators of infinitesimal rotations in 3 **dimensions** : Let's turn to the important special case of rotations in 3-dimensional space. There are many ways of parametrizing rotations in three dimensions. A very useful way in applications is via three **Euler angles**. These are used, for instance, in studying the rotational motion of a rigid body. Any given orientation of the triad of coordinate axes may be reached from an initial reference orientation by a succession of three rotations about a prescribed set of three different axes.[47] Our present objective is somewhat different: we

[46] When n is even, it is obvious that $\mathbf{r} \mapsto -\mathbf{r}$ is a proper rotation, with det $R = +1$.

[47] There are as many as twelve different conventions for defining Euler angles. I will not digress into these here.

are interested in the rotation matrices *per se* and in their algebraic properties. We are also interested in finding the *explicit* transformation formula of a vector under an arbitrary rotation of the coordinate axes. As you'll see below, there's quite an easy way to arrive at the exact answer without any tedious algebra.

A convenient way of parametrizing any given rotation is to specify the axis of rotation, i.e., the *direction* in space about which the triad of Cartesian coordinate axes is rotated, and the *amount* or angle of rotation about this axis. We may therefore denote a general rotation matrix by $R(\mathbf{n}, \psi)$, where $\mathbf{n}$ is the unit vector along the axis of rotation, and ψ is the angle of rotation about this axis. All you need in order to write down the rotation matrices corresponding to rotations about the three Cartesian axes is to recall a result from elementary coordinate geometry. A rotation by an angle ψ of the coordinate axes about the origin in the xy-plane gives the new coordinates $x' = x\,\cos\,\psi + y\,\sin\,\psi$ and $y' = -x\,\sin\,\psi + y\,\cos\,\psi$. The z coordinate is left unchanged. Hence, as we have already written down,

$$R(\mathbf{e}_z, \psi) = \begin{pmatrix} \cos\,\psi & \sin\psi & 0 \\ -\sin\,\psi & \cos\,\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

By cyclic permutation of xyz, we may write down the other two matrices

$$R(\mathbf{e}_x, \psi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\,\psi & \sin\,\psi \\ 0 & -\sin\,\psi & \cos\,\psi \end{pmatrix} \quad \text{and} \quad R(\mathbf{e}_y, \psi) = \begin{pmatrix} \cos\,\psi & 0 & -\sin\,\psi \\ 0 & 1 & 0 \\ \sin\,\psi & 0 & \cos\,\psi \end{pmatrix}.$$

It is easily checked that each of these matrices is orthogonal, and has a determinant equal to $+1$. Hence each of them can be built up from the identity matrix *by a succession of infinitesimal rotations* about the axis concerned. We can work backwards from the matrices written down above to see how this is done.

Consider, for definiteness, $R(\mathbf{e}_z, \psi)$. We could implement such a rotation by n successive rotations about the z-axis, each through an infinitesimal angle $\delta\psi$, such that $n\,\delta\psi = \psi$. The matrix $R(\mathbf{e}_z, \delta\psi)$ is easily written down: use the fact that $\sin\,\delta\psi \simeq \delta\psi$ and $\cos\,\delta\psi \simeq 1$ to first order in $\delta\psi$. Separating out the (3×3) unit matrix, which corresponds to the identity transformation (or zero rotation), we get

$$R(\mathbf{e}_z, \delta\psi) = I + i\,(\delta\psi)\,J_3\,, \quad \text{where} \quad J_3 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The parameter $\delta\psi$ has been factored out in the expression above. This makes the elements of the matrix J_3 pure numbers that are independent of the angle of rotation. The reason for separating out the factor i in the definition of J_3 is to ensure that J_3 is

a *hermitian* matrix.[48] The finite-angle rotation matrix $R(\mathbf{e}_z, \psi)$ is then given by

$$R(\mathbf{e}_z, \psi) = \underbrace{R(\mathbf{e}_z, \delta\psi) \cdots R(\mathbf{e}_z, \delta\psi)}_{n \text{ factors}} = [R(\mathbf{e}_z, \delta\psi)]^n = [I + i\,(\delta\psi)\,J_3]^n.$$

Setting $\delta\psi = \psi/n$ and passing to the limit $n \to \infty$,

$$R(\mathbf{e}_z, \psi) = \lim_{n\to\infty} \left(I + \frac{i\,\psi\,J_3}{n}\right)^n = e^{i\,\psi\,J_3}.$$

Repeat the procedure above for the matrices $R(\mathbf{e}_x, \psi)$ and $R(\mathbf{e}_y, \psi)$, to get

$$R(\mathbf{e}_x, \psi) = e^{i\,\psi\,J_1} \quad \text{and} \quad R(\mathbf{e}_y, \psi) = e^{i\,\psi\,J_2},$$

where

$$J_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \text{and} \quad J_2 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}.$$

The form $R(\mathbf{e}_z, \delta\psi) = I + i\,(\delta\psi)\,J_3$ makes it quite clear why the matrix J_3 is called the **generator** of an infinitesimal rotation about the z-axis (i.e., a rotation about the origin, in the xy-plane). Similarly, J_1 and J_2 may be identified as the generators of infinitesimal rotations about the x and y axes (i.e., about the origin, in the yz-plane and zx-plane), respectively.

The matrix corresponding to rotation by a finite angle is obtained *by exponentiating the corresponding generator*. This is a general feature for groups of transformations (more generally, **Lie groups**). The hermitian matrices J_1 , J_2 and J_3 satisfy the **commutation relations**

$$[J_k\,,\, J_l] = i\epsilon_{klm}\,J_m\,,$$

where ϵ_{klm} is the Levi-Civita totally antisymmetric symbol in three dimensions. These commutation relations comprise the **Lie algebra** $\mathfrak{so}(3)$ of the rotation group $SO(3)$. In general, the Lie algebra of the infinitesimal generators of a Lie group determine the 'structure' of the group in the neighborhood of the identity element, and, by continuity, elsewhere too (at least in the part of the group that is continuously connected to the identity element).

- In this sense, the Lie algebra corresponding to a Lie group contains all the information about the group except the *global* topological properties of its parameter space.

1. It is instructive to check out the statements made in the foregoing.

[48] This is the convention generally used in physics, because quantities like J_3 will be associated with operators that, in turn, represent physical observables which must have real values. Recall that the eigenvalues of hermitian matrices are guaranteed to be real.

(a) Work through the steps outlined above.

(b) Using the expressions given above for the matrices J_k , directly calculate the exponentials $e^{i\psi\, J_k}$ for $k = 1, 2$ and 3. (That is, sum the corresponding exponential series.) Verify that you recover the finite-angle rotation matrices $R(\mathbf{e}_x, \psi)$, $R(\mathbf{e}_y, \psi)$ and $R(\mathbf{e}_z, \psi)$.

(c) Verify that the generators J_k satisfy the commutation relations $[J_k\ ,\ J_l] = i\epsilon_{klm}\, J_m$.

2. The general rotation matrix in 3 dimensions: We've seen that the matrices corresponding to rotations about the three Cartesian axes can be written as exponentials of the corresponding infinitesimal generators. What about a *general* rotation $R(\mathbf{n}, \psi)$ about an axis $\mathbf{n}$ pointing in an arbitrary direction in space?

It turns out[49] that the three generators of infinitesimal rotations in 3-dimensional space, $(J_1\ ,\ J_2\ ,\ J_3)$, *themselves* transform under rotations like the components of a vector. It is therefore natural to denote the triplet by the vector symbol $\mathbf{J}$. Then, if the components of the direction vector $\mathbf{n}$ are given by $(n_1\ ,\ n_2\ ,\ n_3)$, we are guaranteed that

$$R(\mathbf{n}, \psi) = e^{i\,(J_1\, n_1 + J_2\, n_2 + J_3\, n_3)\,\psi} \equiv e^{i\,(\mathbf{J}\cdot\mathbf{n})\,\psi}.$$

Since the different matrices J_k do not commute with each other, however, the right-hand side of this equation is *not* equal to the product of exponentials, i.e.,

$$e^{i\,(J_1\, n_1 + J_2\, n_2 + J_3\, n_3)\,\psi} \neq e^{iJ_1\, n_1\, \psi}\, e^{iJ_2\, n_2\, \psi}\, e^{iJ_3\, n_3\, \psi}.$$

In spite of this problem, it turns out to be possible to compute the exponential of the (3×3) matrix $i\,(\mathbf{J}\cdot\mathbf{n})\,\psi$ exactly, and in closed form. Here's how this is done.

We want to find $e^{M\psi}$, where $M = i\,(\mathbf{J}\cdot\mathbf{n})$. Using the definitions of the matrices J_k above, we find

$$M = i\,(\mathbf{J}\cdot\mathbf{n}) = \begin{pmatrix} 0 & n_3 & -n_2 \\ -n_3 & 0 & n_1 \\ n_2 & -n_1 & 0 \end{pmatrix}.$$

In order to find the powers of M explicitly, it is helpful to note that the elements of M are given by

$$M_{ij} = \epsilon_{ijk}\, n_k\,.$$

It follows that

$$(M^2)_{ij} = n_i\, n_j - \delta_{ij}\,, \quad \text{and hence} \quad (M^3)_{ij} = -\epsilon_{ijk}\, n_k = -M_{ij}\,.$$

The fact that $M^3 = -M$ immediately enables us to simplify the exponential $e^{M\psi}$. The result is just a linear combination of the three matrices I, M and M^2. The final answer

[49] As you'll see subsequently.

for the matrix elements of the rotation matrix $R(\mathbf{n}, \psi)$ is both simple and elegant. It reads

$$R_{ij}(\mathbf{n}, \psi) = \delta_{ij} \cos \psi + n_i \, n_j \, (1 - \cos \psi) + \epsilon_{ijk} \, n_k \sin \psi.$$

Even more explicitly, if the spherical polar angles of the unit vector $\mathbf{n}$ are given by θ and φ, we have

$$n_1 = \sin \theta \cos \varphi, \;\; n_2 = \sin \theta \sin \varphi, \;\; n_3 = \cos \theta.$$

Using these expressions, you can write down the complete rotation matrix for an arbitrary rotation $R(\mathbf{n}, \psi)$. Work out the steps outlined above, and find the explicit expression for the rotation matrix $R(\mathbf{n}, \psi)$.

3. Find the eigenvalues and eigenvectors of the general rotation matrix $R(\mathbf{n}, \psi)$.

The finite rotation formula for a vector: Once we have $R(\mathbf{n}, \psi)$ explicitly, it is straightforward to apply it to an arbitrary position vector $\mathbf{r}$, to obtain

$$x'_i = R_{ij} \, x_j = x_i \cos \psi + n_i \, x_j \, n_j \, (1 - \cos \psi) + \epsilon_{ijk} \, x_j \, n_k \sin \psi.$$

Expressing this formula back in terms of the vectors $\mathbf{r}$ and $\mathbf{n}$ helps us understand it in physical terms:

$$\mathbf{r}' = (\cos \psi) \, \mathbf{r} + (1 - \cos \psi) \, (\mathbf{r} \cdot \mathbf{n}) \, \mathbf{n} + (\sin \psi) \, (\mathbf{r} \times \mathbf{n}).$$

It follows at once that, by its very definition, *any* vector $\mathbf{a}$ transforms under a general rotation $R(\mathbf{n}, \psi)$ according to

$$\mathbf{a} \xrightarrow{R(\mathbf{n}, \psi)} \mathbf{a}' = (\cos \psi) \, \mathbf{a} + (1 - \cos \psi) \, (\mathbf{a} \cdot \mathbf{n}) \, \mathbf{n} + (\sin \psi) \, (\mathbf{a} \times \mathbf{n}).$$

This is sometimes called the **finite rotation formula** for a vector. $\mathbf{a}'$ has a component along the original vector $\mathbf{a}$, a component along the axis of rotation $\mathbf{n}$, and a component in the direction normal to the plane formed by $\mathbf{a}$ and $\mathbf{n}$.

Relation between the groups $SO(3)$ and $SU(2)$: Orthogonal matrices and unitary matrices are foremost among the sets of special kinds of matrices that form groups under matrix multiplication, and that are of great importance in physics. All $(n \times n)$ orthogonal matrices with real elements form the orthogonal group $O(n)$, while all $(n \times n)$ unitary matrices with complex elements form the unitary group $U(n)$. When the determinants of the corresponding matrices are further restricted to the value $+1$, we have the special groups $SO(n)$ and $SU(n)$. It turns out that there is a very important connection between the rotation group $SO(3)$ and the group of unimodular, unitary (2×2) matrices, denoted by $SU(2)$. This relationship has profound physical consequences.

Let's begin by observing that the three Pauli matrices σ_1 , σ_2 , σ_3 (abbreviated as $\boldsymbol{\sigma}$) satisfy the same Lie algebra as the generators of rotations: if we set

$$J_k = \tfrac{1}{2}\sigma_k \;\; (\text{where} \;\; k = 1, 2, 3), \;\; \text{then} \;\; [J_k \, , \, J_l] = i\epsilon_{klm} \, J_m \, .$$

Therefore the (2×2) matrix

$$U(\mathbf{n}, \psi) = e^{i(\boldsymbol{\sigma}\cdot\mathbf{n})\psi/2}$$

also represents a rotation of the coordinate axes in 3-dimensional space about the direction $\mathbf{n}$, and through an angle ψ, just as the (3×3) orthogonal matrix $R(\mathbf{n}, \psi)$ does. But what does the matrix U matrix *act* on? Clearly, we need a representation of a point $\mathbf{r}$ in 3-dimensional space that is compatible with the (2×2) matrix representation of a rotation of the coordinate axes. This is obtained by noting that the point $\mathbf{r} = (x_1 \, , \, x_2 \, , \, x_3)$ in 3-dimensional space can *also* be represented as a (2×2) matrix given by

$$\mathbf{r} \cdot \boldsymbol{\sigma} \equiv x_1 \, \sigma_1 + x_2 \, \sigma_2 + x_3 \, \sigma_3 = \begin{pmatrix} x_3 & x_1 - ix_2 \\ x_1 + ix_2 & -x_3 \end{pmatrix} .$$

Similarly, *any* vector $\mathbf{a} = (a_1 \, , \, a_2 \, , \, a_3)$ in 3-dimensional space can *also* be represented as a (2×2) matrix given by

$$\mathbf{a} \cdot \boldsymbol{\sigma} \equiv a_1 \, \sigma_1 + a_2 \, \sigma_2 + a_3 \, \sigma_3 = \begin{pmatrix} a_3 & a_1 - ia_2 \\ a_1 + ia_2 & -a_3 \end{pmatrix} .$$

Note that the matix $(\mathbf{a} \cdot \boldsymbol{\sigma})$ *uniquely* determines the components $(a_1 \, , \, a_2 \, , \, a_3)$, and *vice versa*. The corresponding transformation rule under a rotation, when we choose to represent vectors by (2×2) matrices as above, is given by

$$(\mathbf{r}\,' \cdot \boldsymbol{\sigma}) = U(\mathbf{n}, \psi) \, (\mathbf{r} \cdot \boldsymbol{\sigma}) \, U^{-1}(\mathbf{n}, \psi),$$

and more generally,

$$(\mathbf{a}\,' \cdot \boldsymbol{\sigma}) = U(\mathbf{n}, \psi) \, (\mathbf{a} \cdot \boldsymbol{\sigma}) \, U^{-1}(\mathbf{n}, \psi).$$

4. The matrix $U(\mathbf{n}, \psi)$ can be determined explicitly.

(a) The square of any Pauli matrix is the unit matrix, i.e., $\sigma_j^2 = I$. Moreover, any two different Pauli matrices anticommute with each other, i.e., $\sigma_j \, \sigma_k + \sigma_k \, \sigma_j = 0, \; j \neq k$. Using these properties, show that

$$\begin{aligned} U(\mathbf{n}, \psi) \;\; &= \;\; \exp\left(\tfrac{1}{2} i \, \boldsymbol{\sigma} \cdot \mathbf{n} \, \psi\right) = I \, \cos\tfrac{1}{2}\psi + i \, (\mathbf{n} \cdot \boldsymbol{\sigma}) \, \sin\tfrac{1}{2}\psi \\ &= \;\; \begin{pmatrix} \cos\frac{1}{2}\psi + i \sin\frac{1}{2}\psi \, \cos\theta & i \sin\frac{1}{2}\psi \, \sin\theta \, e^{-i\varphi} \\ i \sin\frac{1}{2}\psi \, \sin\theta \, e^{i\varphi} & \cos\frac{1}{2}\psi - i \sin\frac{1}{2}\psi \, \cos\theta \end{pmatrix} , \end{aligned}$$

where θ and φ, respectively, denote the polar and azimuthal angles specifying the unit vector $\mathbf{n}$.

(b) Hence verify that U is unitary and unimodular, i.e., $UU^\dagger = I = U^\dagger U$, and $\det U = +1$.

In other words, $U(\mathbf{n}, \psi)$ is an element of the special unitary group $SU(2)$. Now, it is obvious that replacing the matrix $U(\mathbf{n}, \psi)$ by $-U(\mathbf{n}, \psi)$ does not alter these properties. Moreover, given any vector $\mathbf{a}$, a rotation transformation made using $-U$ in the place of U leads to exactly the same $\mathbf{a}'$.

- Hence there are *two* distinct elements $U(\mathbf{n}, \psi)$ and $-U(\mathbf{n}, \psi)$ of $SU(2)$, differing only by an overall sign, corresponding to every rotation matrix $R(\mathbf{n}, \psi) \in SO(3)$.
- This is the famous 2-to-1 **homomorphism** from $SU(2)$ to $SO(3)$.

Both the (2×2) unit matrix I and its negative $-I$ are elements of $SU(2)$. By themselves, these two matrices form a group, namely, the cyclic group of order 2, denoted by $\mathbb{Z}_2$.[50] This group ($\mathbb{Z}_2$) is said to be the **center** of $SU(2)$: its elements (I and $-I$) are the only elements of $SU(2)$ with which *all* the elements of the group commute. I and $-I$ constitute the **kernel** of the homomorphism from $SU(2)$ to $SO(3)$: that is, they are the only elements of $SU(2)$ that map into the unit element of $SO(3)$, the (3×3) unit matrix. Thus there is an **isomorphism** between the **quotient group** $SU(2)/\mathbb{Z}_2$ and $SO(3)$, written symbolically as

$$SU(2)/\mathbb{Z}_2 \simeq SO(3).$$

We'll resume our discussion of other group theoretical properties after completing an important bit of unfinished business: namely, the proof that the three generators of rotations in 3-dimensional space themselves transform like a vector.

Rotation generators in 3 dimensions transform like a vector: We have used the symbol $\mathbf{J}$ to denote the triplet $(J_1\,,\, J_2\,,\, J_3)$ of the generators $SO(3)$, implying that these quantities themselves transform like the components of a vector under rotations of the coordinate axes. We must now establish that this is indeed so.

Recall that the actual *representation* of each J_k depends on what it is required to act on. You've already come across two such representations: the defining representation in terms of (3×3) matrices, and a (2×2) matrix representation in which $J_k = \frac{1}{2}\,\sigma_k$. But the generators J_k operators can, and do, have an infinite number of other representations, including infinite-dimensional ones—for example, when they act on state vectors in infinite-dimensional Hilbert spaces in quantum mechanics. A rotation R of the coordinate axes *induces* a transformation in the Hilbert space. This is a **unitary**

[50] $\mathbb{Z}_2$ is also the group of integers under addition *modulo* 2: all the even integers are represented by the identity element of $\mathbb{Z}_2$, while all the odd integers are represented by the only other element of the group.

transformation given by $U(R) = e^{i\,(\mathbf{J}\cdot\mathbf{n})\,\psi}$, where the generators $J_k\,(k = 1, 2, 3)$ have the representation appropriate to the Hilbert space. What we need to do is to establish that the triplet $(J_1\,,\ J_2\,,\ J_3)$ transforms like a vector under rotations of the coordinate axes, *independent* of any particular representation for the generators. In effect, this means that the only input we can use is the *algebra* satisfied by the three operators, namely, $[J_k\,,\ J_l] = i\,\epsilon_{klm}\,J_m\,.$

Let's consider the general case of a rotation about an arbitrary axis $\mathbf{n}$ through an angle ψ. (All special cases can then be read off from it.) The task is to show that $\mathbf{J}$ satisfies the finite rotation formula, which is

$$e^{-i\,(\mathbf{J}\cdot\mathbf{n})\,\psi}\,J_i\,e^{i\,(\mathbf{J}\cdot\mathbf{n})\,\psi} = (\cos\,\psi)\,J_i + (1 - \cos\,\psi)\,(J_j\,n_j)\,n_i + (\sin\,\psi)\,\epsilon_{ijk}\,J_j\,n_k\,.$$

Observe that the transformation rule for the operator $\mathbf{J}$, given by the left-hand side of this equation, is *not* just $\mathbf{J}\,' = R(\mathbf{n}, \psi)\,\mathbf{J}$. This is because $\mathbf{J}$ is not an ordinary vector, but an *operator*-valued vector. I reiterate that the components of $\mathbf{J}$ are not necessarily (3×3) matrices, nor do we care what the actual representation is. In vector form, we need to show that

$$e^{-i\,(\mathbf{J}\cdot\mathbf{n})\,\psi}\,\mathbf{J}\,e^{i\,(\mathbf{J}\cdot\mathbf{n})\,\psi} = (\cos\,\psi)\,\mathbf{J} + (1 - \cos\,\psi)\,(\mathbf{J}\cdot\mathbf{n})\,\mathbf{n} + (\sin\,\psi)\,(\mathbf{J}\times\mathbf{n}).$$

The right-hand side of this equation has precisely the form of the finite rotation formula for the coordinate $\mathbf{r}$ of a point in three-dimensional space, with $\mathbf{J}$ replacing $\mathbf{r}$. Therefore, once the equation above is proved, we may assert that $\mathbf{J}$ itself transforms like a vector under rotations.

What you will need for the purpose is a remarkable and extremely useful operator identity called **Hadamard's Lemma**. This identity expresses the operator $e^{\lambda A}\,B\,e^{-\lambda A}$, where A and B are (in general, non-commuting) operators acting on some linear vector space and λ is a scalar constant (real or complex), as an infinite sum involving multiple nested commutators of A with $[A\,,\ B]$. The formula is

$$e^{\lambda A}\,B\,e^{-\lambda A} = B + \lambda\,[A\,,\ B] + \frac{\lambda^2}{2!}\,\big[A\,,\ [A\,,\ B]\,\big] + \frac{\lambda^3}{3!}\,\Big[A\,,\ \big[A\,,\ [A\,,\ B]\,\big]\,\Big] + \cdots$$

If any of the multiple commutators on the right-hand side should happen to be zero, the series terminates. The derivation of this identity is quite straightforward, and will be dealt with shortly.

5. Establish the transformation rule for $\mathbf{J}$ given above, using Hadamard's Lemma. Here's an outline of the steps involved. Start with the identifications

$$\lambda = -i\psi,\ \ A = \mathbf{J}\cdot\mathbf{n} = J_k\,n_k\,,\ \ B = J_i\,.$$

For ease of notation, write the r-fold multiple commutator in Hadamard's Lemma as $C_r\,.$ Then

$$[A\,,\ C_r] = C_{r+1}\,,\ \ \text{where}\ \ C_1 = [A\,,\ B]\ \ \text{and}\ \ r = 1, 2, \ldots\,.$$

Using the commutator algebra $[J_k\,,\,J_l] = i\,\epsilon_{klm}\,J_m$, it is straightforward to show that

$$C_1 = i\,\epsilon_{ilk}\,J_l\,n_k \quad \text{and} \quad C_2 = [A\,,\,C_1] = J_i - n_i\,n_k\,J_k\,.$$

Further,

$$C_3 = [A\,,\,C_2] = i\,\epsilon_{ilk}\,J_l\,n_k = C_1\,, \quad \text{and hence} \quad C_4 = C_2\,,\ \ C_5 = C_1\,,\,\ldots\,.$$

Thus, $C_{2r+1} = C_1$ and $C_{2r} = C_2$. The infinite series in Hadamard's lemma can now be summed easily. The transformation formula for $\mathbf{J}$ follows upon simplification.

There's a small subtlety involved in the left-hand side of the transformation formula. The factor $e^{i\,(\mathbf{J}\cdot\mathbf{n})\,\psi}$, which we have identified with $R(\mathbf{n},\psi)$, appears on the *right* of J_i, while its adjoint $e^{-i\,(\mathbf{J}\cdot\mathbf{n})\,\psi}$ appears on its left. You might have expected the reverse, based perhaps on the transformation rule for a second-rank tensor. But the order given here is correct. It is related to the fact that $\mathbf{J}$ itself is an operator (e.g., in a Hilbert space), and this is the transformation rule for an operator, as opposed to that for a state vector.

6. Derivation of Hadamard's Lemma: Show that, if A and B are operators acting on some linear vector space, and λ is a scalar parameter,

$$e^{\lambda A}\,B\,e^{-\lambda A} = B + \lambda\,[A\,,\,B] + \frac{\lambda^2}{2!}\,\big[A\,,\,[A\,,\,B]\,\big] + \frac{\lambda^3}{3!}\,\Big[A\,,\,\big[A\,,\,[A\,,\,B]\,\big]\,\Big] + \cdots\,.$$

Here's how the formula is derived. Define the operator-valued function the parameter λ according to

$$F(\lambda) = e^{\lambda A}\,B\,e^{-\lambda A}.$$

Differentiate both sides with respect to λ, to get the differential equation satisfied by $F(\lambda)$:

$$F\,'(\lambda) = A\,F(\lambda) - F(\lambda)\,A.$$

Note the ordering of the operators. A and F do not commute with each other, in general. Now, $F(\lambda)$ is a regular function of λ that can be expanded in a Taylor series in powers of λ, with coefficients that are operators, of course. Thus

$$F(\lambda) = F(0) + \lambda\,F\,'(0) + \frac{\lambda^2}{2!}\,F\,''(0) + \cdots\,.$$

The existence of such a power series expansion for $F(\lambda)$ implies the existence of an analogous series for its derivative $F\,'(\lambda)$ as well. This assertion follows from the fact that $F(\lambda)$ is an *analytic* function of the complex variable λ—in this instance, for all finite values of $|\lambda|$. Hence

$$F\,'(\lambda) = F\,'(0) + \lambda\,F\,''(0) + \frac{\lambda^2}{2!}\,F\,'''(0) + \cdots\,.$$

Insert these series for $F(\lambda)$ and $F'(\lambda)$ in the differential equation, and equate the coefficients of like powers of λ. Note that $F(0) = B$. Determine the successive derivatives $F^{(n)}(0)$ recursively, to arrive at Hadamard's formula. This formula is the starting point for the derivation of a number of extremely useful and important operator identities.

7. The general form of the elements of $U(2)$ and $SU(2)$: Let's return to our discussion of group theoretical aspects. You have seen that there is a homomorphism between the rotation group $SO(3)$ and the special unitary group $SU(2)$. This is a convenient place to determine the general forms of the elements of the unitary group $U(2)$ and the special unitary group $SU(2)$.

Consider an arbitrary (2×2) matrix

$$M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \quad \text{so that} \quad M^\dagger = \begin{pmatrix} \alpha^* & \gamma^* \\ \beta^* & \delta^* \end{pmatrix},$$

where α, β, γ and δ are complex numbers. Hence 8 real parameters are required to specify such a matrix. Show that, if the condition that M be unitary (i.e., $MM^\dagger = I$) is imposed, the matrix must be of the form

$$M = \begin{pmatrix} \alpha & \beta \\ -e^{i\theta}\,\beta^* & e^{i\theta}\,\alpha^* \end{pmatrix}, \quad \text{where} \quad |\alpha|^2 + |\beta|^2 = 1,$$

and θ is any real number. This is the form of a general element of the unitary group $U(2)$. It is specified by 4 independent real parameters, namely: θ, and 3 out of the 4 numbers α_1 , α_2 , β_1 and β_2 , where $\alpha_1 + i\alpha_2 = \alpha$, $\beta_1 + i\beta_2 = \beta$, and

$$\alpha_1^2 + \alpha_2^2 + \beta_1^2 + \beta_2^2 = 1.$$

If the further condition $\det M = +1$ is imposed, then θ must be 0 or an integer multiple of 2π; and M must then be of the form

$$M = \begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix}, \quad \text{where} \quad |\alpha|^2 + |\beta|^2 = 1.$$

This is the form of a general element of $SU(2)$. The four real parameters occurring in it satisfy a single constraint, reducing the number of independent parameters to three.

More generally, an $(n \times n)$ matrix with complex elements has $2n^2$ independent real parameters. Requiring that the matrix be unitary reduces this number to n^2. Thus $U(n)$ is an n^2-parameter group. The determinant of any element of $U(n)$ is a complex number of unit modulus, i.e., a number like $e^{i\theta}$. If, further, we require that the matrices be unimodular, the number of independent real parameters is reduced further to $n^2 - 1$. Such matrices comprise the special or unimodular unitary group $SU(n)$, which is a subgroup of $U(n)$.

The parameter spaces of $SU(2)$ and $SO(3)$: The parameter space of $SU(2)$ follows immediately: it is just the 'surface' of the unit (hyper)sphere

$$\alpha_1^2 + \alpha_2^2 + \beta_1^2 + \beta_2^2 = 1.$$

This is the 3-sphere $\mathbb{S}^3$, which is **simply connected**: that is, any closed path in it can be continuously deformed and shrunk to a point without leaving the space.

In contrast, the parameter space of $SO(3)$ is more complicated. As we've seen, a rotation in 3-space can be parametrized by a unit vector $\mathbf{n}$ and an angle of rotation ψ. The unit vector $\mathbf{n}$ is specified a polar angle $0 \leq \theta \leq \pi$ and an azimuthal angle $0 \leq \varphi < 2\pi$. You might expect, at first sight, that the angle ψ can run from 0 to 2π. But it is sufficient to let it take values in the range $0 \leq \psi \leq \pi$, because of a simple but profound fact: rotation about the axis $\mathbf{n}$ through an angle π leads to the *same* end result as a rotation about the oppositely-directed axis $-\mathbf{n}$ through an angle π, as is easily verified. This is 'a fact of life' in three-dimensional space! As a result, the parameter space spanned by θ, φ and ψ can be represented by a *solid* (or 3-dimensional) sphere of radius π (rather than 2π). The *direction* vector of any point in this sphere specifies the *direction* $\mathbf{n}$ of the corresponding rotation, while the *distance* of that point from the origin O represents the *amount* of rotation, ψ. The origin itself corresponds to the identity transformation, i.e., to the case of no rotation at all. The noteworthy point is that every point on the *surface* of this sphere is mathematically *identical* to its antipodal point on the surface. (You may imagine that every pair of antipodal points is connected by an invisible string of zero length!) This identification ensures that a rotation through π about any axis $\mathbf{n}$ is the same as a rotation through π about the oppositely directed axis, $-\mathbf{n}$. (Unfortunately, 3-dimensional Euclidean space is insufficient to embed this parameter space, precisely because it cannot exhibit the mathematical fact that every pair of antipodal points on the surface of the solid sphere is really a single point. That's why we can't display an actual real-life model of this space. The space is termed the real projective space $\mathbb{RP}^3$.)

As a consequence of this identification of antipodal points, the parameter space becomes **doubly-connected**. This means that there are two distinct *classes* of oriented closed paths in such a space:

(i) Closed paths that can be shrunk continuously to a point (without leaving the space).

(ii) Closed paths that can be shrunk to a point continuously only when traversed *twice*.

The existence of these two classes of closed paths leads, respectively, to the so-called single-valued or **tensor representations** of the rotation group, and the double-valued or **spinor representations** of the group. These correspond, in the language of quantum mechanics, to integral and half-odd-integral values of the angular momentum

quantum number j. Tensors return to their original values under a 2π rotation of the coordinate axes about any direction, while spinors change sign under such a rotation. A rotation of 4π is needed to return a spinor to its original value.

The universal covering group of a Lie group: It turns out that every multiply-connected Lie group G (that is, a Lie group with a parameter space that is not simply-connected) has a unique **universal covering group** $\widetilde{G}$ whose parameter space *is* simply connected. There is a homomorphism or many-to-one mapping from $\widetilde{G}$ to G.

- The Lie algebras of a Lie group and its covering group are, however, identical.

This means that the two groups are essentially identical in *local* neighborhoods (in parameter space), although they may be very different from each other *globally*. For one thing, the *topologies* of the parameter spaces are quite different from each other (their connectivities are different). In the case of the rotation group $SO(3)$, the parameter space (or **group manifold**) is doubly connected. The universal covering group of $SO(3)$ is the special unitary group $SU(2)$. The latter is simply connected: its parameter space is the 3-sphere $\mathbb{S}^3$. And, as I have already pointed out, there is a 2-to-1 homomorphism from $SU(2)$ to $SO(3)$: there are two matrices in $SU(2)$, differing by a sign, that correspond to each matrix in $SO(3)$.

The group $SO(2)$ and its covering group: These ideas are even more easily illustrated in the case of the group $SO(2)$ of rotations 2-dimensional space. The parameter space of $SO(2)$ is a single angle *modulo* 2π. In other words, it is the 1-sphere $\mathbb{S}^1$ (the circumference of a circle, in plain language). This space has an *infinite* number of distinct classes of closed paths, labelled by the number of times the (directed) path winds completely around the circle. When $\mathbb{S}^1$ is 'rolled out' on a line, the latter covers the circle an infinite number of times, the copies being labeled by the elements of the additive group of integers, $\mathbb{Z}$. The universal covering group of $SO(2)$ is therefore the real number line $\mathbb{R}$ (which is also a Lie group!). We have the isomorphism

$$\mathbb{R}/\mathbb{Z} \simeq SO(2),$$

so that $SO(2)$ is also a quotient group. Note also that $SO(2)$ is equivalent to $U(1)$, the multiplicative group of complex numbers of unit modulus (i.e., complex numbers of the form $e^{i\alpha}$, where α is real).

The groups $SO(n)$ and Spin (n): What about the rotation group $SO(n)$ for $n = 4, 5, \ldots$? It turns out that every one of these groups is also doubly connected, just as $SO(3)$ is. The universal covering group of $SO(n)$ for $n \geq 3$ is called Spin (n). There is again a 2-to-1 homomorphism from Spin (n) to $SO(n)$, and $SO(n)$ is a quotient group according to

$$\text{Spin}\,(n)/\mathbb{Z}_2 \simeq SO(n), \quad n \geq 3.$$

We've seen that Spin $(3) \simeq SU(2)$. For higher values of n the covering group does not reduce to anything so readily identifiable, except for the case of $SO(6)$. In that case it turns out that Spin $(6) \simeq SU(4)$. These are the only two cases when Spin (n) is a special unitary group.

Parameter spaces of $U(n)$ and $SU(n)$: For completeness, let me touch upon the parameter spaces of the unitary group $U(n)$ and the special unitary group $SU(n)$, where $n \geq 2$.[51] The parameter space of $SU(n)$ is simply connected. $U(n)$ is connected, but not simply connected. The determinant of each member of $U(n)$ is a complex number of unit modulus. $U(1)$ is a subgroup of $U(n)$: matrices of the form $e^{i\alpha}I$ are included among the elements of $U(n)$, and these matrices form a subgroup that is isomorphic to $U(1)$. The presence of this $U(1)$ subgroup, whose parameter space is the infinitely connected 1-sphere or circle $\mathbb{S}^1$, makes the parameter space of $U(n)$ also infinitely connected. In technical terms, $U(n)$ is the **semidirect product** of its **normal subgroup**[52] $U(1)$ and its subgroup $SU(n)$, and is written as

$$U(n) \simeq U(1) \rtimes SU(n).$$

Conversely, the special unitary group $SU(n)$ is just the unitary group $U(n)$ 'modulo' $U(1)$, i.e., it is a **quotient group** according to

$$U(n)/U(1) \simeq SU(n).$$

The universal covering group of $U(n)$ is $\mathbb{R} \otimes SU(n)$, which you can understand heuristically because $\mathbb{R}$ is the universal covering group of $U(1)$, and $SU(n)$ is already simply-connected.

A bit about the first homotopy group of a space: The connectivity of a space $\mathbb{V}$ is formally determined by finding the different **equivalence classes** of directed closed paths (or loops) in $\mathbb{V}$. (Two such paths are equivalent if either of them can be continuously and smoothly deformed into the other without leaving the space and without cutting open the closed path.) More formally, each closed path is a *map* $\mathbb{S}^1 \to \mathbb{V}$. The equivalence classes of such maps form a group, called the **fundamental group** or the **first homotopy group** of $\mathbb{V}$, denoted by $\pi_1(\mathbb{V})$. If *every* possible closed path in $\mathbb{V}$ can be shrunk continuously to a point, then $\pi_1(\mathbb{V})$ reduces to the trivial group with one element (this is often written as $\pi_1(\mathbb{V}) = 0$), and $\mathbb{V}$ is a simply-connected space. The converse is also true. Thus

$$\mathbb{V} \text{ is a simply-connected space } \iff \pi_1(\mathbb{V}) = 0.$$

[51] Recall that $U(1)$ is isomorphic to the group of complex numbers $\{e^{i\alpha}\}$, where α is real, under multiplication. It should be obvious that the group $SU(1)$ is the trivial group with just one element (the unit element).

[52] H is a *normal* subgroup of a group G if the following condition is satisfied: for any element $h \in H$ and any $g \in G$, the element $ghg^{-1} \in H$.

The fundamental group of the 1-sphere $\mathbb{S}^1$ itself is the set of equivalence classes of maps of $\mathbb{S}^1$ to $\mathbb{S}^1$. This is $\pi_1(\mathbb{S}^1) \simeq \mathbb{Z}$, the group of integers under addition. This follows from the fact that the class of a directed closed path on $\mathbb{S}^1$ is labeled by the winding number of the path (the number of times it goes around $\mathbb{S}^1$ completely). It is intuitively clear that $\pi_1(\mathbb{S}^2) = 0$.[53] It can be shown rigorously that $\pi_1(\mathbb{S}^n) = 0$ for all $n \geq 2$. As you might expect, since the 2-torus $\mathbb{T}^2$ is equivalent to the direct product space $\mathbb{S}^1 \otimes \mathbb{S}^1$, we have $\pi_1(\mathbb{T}^2) = \mathbb{Z} \otimes \mathbb{Z}$. This generalizes in an obvious fashion to the n-torus $\mathbb{T}^n = \underbrace{\mathbb{S}^1 \otimes \cdots \otimes \mathbb{S}^1}_{n \text{ factors}}$.

The fundamental groups of various group manifolds (and of other spaces as well, of course) are of great interest. Based on what has been mentioned in the foregoing, here are some of the results concerned.[54]

$$\begin{aligned}
\pi_1(\mathbb{R}^n) &= 0, \quad n \geq 1. \\
\pi_1(\mathbb{S}^1) &= \pi_1(\mathbb{T}^1) = \mathbb{Z}. \\
\pi_1(\mathbb{S}^n) &= 0, \quad n \geq 2. \\
\pi_1(\mathbb{T}^n) &= \pi_1(\mathbb{S}^1 \otimes \cdots \otimes \mathbb{S}^1) = \mathbb{Z} \otimes \cdots \otimes \mathbb{Z}. \\
\pi_1(SO(2)) &= \pi_1(U(1)) = \pi_1(\mathbb{S}^1) = \mathbb{Z}. \\
\pi_1(\text{Spin}\,(n)) &= 0, \quad n \geq 3. \\
\pi_1(SO(n)) &= \pi_1(\text{Spin}\,(n)/\mathbb{Z}_2) = \mathbb{Z}_2\,, \quad n \geq 3. \\
\pi_1(SU(2)) &= \pi_1(\mathbb{S}^3) = 0. \\
\pi_1(SU(n)) &= 0. \\
\pi_1(U(n)) &= \pi_1(U(1) \rtimes SU(n)) = \mathbb{Z}.
\end{aligned}$$

[53] A fact that is picturesquely expressed by the statement, "You can't lasso a basketball!"

[54] When we write $\pi_1(G)$ where G is a Lie group, we mean the fundamental group of the parameter space of G.

15 QUIZ 2

I. Are the statements in quotation marks true or false?

1. "The function $\sin(1/z)$ does not have a Taylor series expansion in the neighborhood of $z = 0$."

2. A function $f(z)$ is defined by the power series $\sum_{n=0}^{\infty} z^{2n+1}/[n!\,(n+1)!]$ about the origin.

   "$f(z)$ is an entire function of z."

3. "The only singularity of $1/\Gamma(z)$ is a simple pole at $z = 0$."

4. "The Mittag-Leffler expansion of the gamma function is given by $\Gamma(z) = \sum_{n=0}^{\infty}(-1)^n/[(z+n)\,n!]$."

5. Let $f(z) = z + z^3 + z^9 + z^{27} + \cdots$ *ad infinitum*, for $|z| < 1$.

   "$f(z)$ cannot be analytically continued outside the unit circle."

6. "The power series $\sum_{n=1}^{\infty}(\ln n)\,z^n/n$ converges at all points on its circle of convergence."

7. "The function $f(z) = 1/(e^z - 1)$ is a meromorphic function of z."

8. "Dispersion relations for the real and imaginary parts of a generalized susceptibility $\chi(\omega)$ can be derived only if the corresponding response function $\phi(t)$ decays to zero faster than any negative power of t, as $t \to \infty$."

9. "The derivative of the gamma function, $\Gamma'(z)$, has zero residue at each of its poles."

10. "The Legendre function of the second kind, $Q_\nu(z)$, has branch points in the z-plane even when ν is a positive integer."

11. "The Laplace transform of the function $f(t) = \cosh \pi t$ has no singularities in the region $\operatorname{Re} s > \pi$."

12. Bessel's differential equation is $\left[z^2 \dfrac{d^2}{dz^2} + z \dfrac{d}{dz} + (z^2 - \nu^2)\right] f(z) = 0$, where ν is a parameter.

    "If $\phi_\nu(z)$ is any solution of this equation, then $\phi_{-\nu}(z)$ must be equal to $\phi_\nu(z)$, apart from a possible multiplicative constant."

13. "The group $\mathrm{M\ddot{o}b}\,(2, \mathbb{C})$ of Möbius transformations of the complex plane has continuous subgroups, but no discrete subgroups."

14. "The group $\mathrm{M\ddot{o}b}\,(2, \mathbb{C})$ of Möbius transformations of the complex plane is isomorphic to the group $SO(3, 1)$ of homogeneous proper Lorentz transformstions in $(3 + 1)$-dimensional spacetime."

15. "The Riemann surface of the function $f(z) = z^{1/2}\,(z - 1)^{-1/3}$ has 6 sheets."

16. "It is possible to find a contour integral representation of the beta function $B(z, w)$ that is valid for all complex values of both z and w."

17. "The Riemann zeta function $\zeta(z)$ cannot be continued analytically to the left of the line $\operatorname{Re} z = \frac{1}{2}$, because it has an infinite number of zeroes on that line."

18. "The Fourier transform operator in $\mathcal{L}_2(-\infty\,,\,\infty)$ has a finite number of eigenvalues, each of which is infinitely degenerate."

19. Let $G(x, x')$ denote the Green function of the differential operator d^2/dx^2 where $x \in [-1, 1]$.

    "As a function of x, G is continuous at $x = x'$, but its derivative $\partial G/\partial x$ has a finite discontinuity at $x = x'$."

20. "The fundamental Green function of the Laplacian operator ∇^2 in four-dimensional Euclidean space is $G(\mathbf{r}, \mathbf{r}\,') = -1/[4\pi^2(\mathbf{r} - \mathbf{r}\,')^2]$."

21. Consider the diffusion equation in d-dimensional space, $\partial f/\partial t = D\nabla^2 f$ with boundary condition $f(\mathbf{r}, t) \to 0$ as $r \to \infty$ and initial condition $f(\mathbf{r}, 0) = \delta^{(d)}(\mathbf{r})$.

    "The fundamental solution to this equation is a Gaussian in each Cartesian component of $\mathbf{r}$, for all positive integer values of the dimension d."

22. The scattering amplitude for the scattering of a nonrelativistic particle of mass m in a central potential $\lambda V(r)$ is given by

$$f(k, \theta) = -\frac{m\lambda}{2\pi\hbar^2} \int d^3r\, e^{-i\mathbf{k}\,'\cdot\,\mathbf{r}}\, V(r)\, \psi(\mathbf{r}),$$

    where $\mathbf{k}\,'$ is the scattered wave vector.

    "This formula is valid only if the potential $V(r)$ decays to zero as $r \to \infty$ more rapidly than any inverse power of r."

23. Continuation: "In the Born approximation, the scattering amplitude in the forward direction ($\theta = 0$) vanishes identically."

24. Continuation: "In the Born approximation, the imaginary part of the scattering amplitude vanishes identically."

25. Consider the Helmholtz operator $\nabla^2 + \mathbf{k}^2$ in three-dimensional space.

    "The fundamental Green function of this operator, corresponding to outgoing spherical waves, is $G(\mathbf{r} - \mathbf{r}\,') = -e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}\,')}/\big(4\pi|\mathbf{r} - \mathbf{r}\,'|\big)$."

26. Consider the wave operator $(1/c^2)\,\partial^2/\partial t^2 - \nabla^2$ in $(d+1)$-dimensional spacetime, where c is the speed of light in a vacuum. Let $G^{(d)}(R, \tau)$ denote the causal retarded Green function of the operator.

    "$G^{(d)}(R, \tau)$ vanishes identically when $(c\tau, \mathbf{R})$ is a time-like four-vector."

27. Continuation: "$G^{(d)}(R, \tau)$ is singular when $(c\tau, \mathbf{R})$ is a light-like four-vector."

28. Let $\mathbf{J} = (J_i\,,\, J_2\,,\, J_3)$ be the generators of rotations in three-dimensional space, satisfying the Lie algebra $[J_j\,,\, J_k] = i\epsilon_{jkl}\, J_l\,.$

    "The lowest-dimensional, non-trivial, unitary representation of the generators is in terms of (2×2) matrices with complex elements."

29. "The parameter space of the group $SU(n)$ is doubly connected."

30. "The first homotopy group of the parameter space of the special orthogonal group $SO(n)$, for every $n \geq 3$, is $\mathbb{Z}_2$."

II. Fill in the blanks in the following.

1. Given that the imaginary part of an entire function $f(z)$ is

$$v(x, y) = e^{(x^2 - y^2)}\, \sin(2xy),$$

   the function is $f(z) = \cdots$.

2. The meridian of longitude φ on the Riemann sphere is mapped into a straight line in the complex plane. The equation of this straight line is $y = mx + c$, where $m = \cdots$ and $c = \cdots$.

3. The region of absolute convergence in the complex z-plane of the power series $\sum\limits_{n=0}^{\infty} \big[(n+1)/(n^2+1)\big](\frac{1}{2}z)^n$ is $\cdots$.

4. The residue at infinity of the function $f(z) = (z - z^{-1})^3$ is $\operatorname{Res}\limits_{z=\infty} f(z) = \cdots$.

5. Let $[z_1, z_2\,;\, z_3, z_4]$ denote the cross-ratio of the four points z_1, z_2, z_3 and z_4 in the complex plane. Then $[z_1, z_2\,;\, z_3, z_4] + [z_1, z_3\,;\, z_2, z_4] = \cdots$.

6. The Möbius transformation $z \mapsto w$ such that three given points z_1, z_2, z_3 are mapped respectively into three other given points w_1, w_2, w_3 is expressed by a relation between w and z that reads $\cdots$.

7. Under the Möbius transformation $z \mapsto w = (z+1)/(z+2)$, an infinitesimal area element δA centered at the point $z = -3/2$ is mapped to an element of area $\lambda\, \delta A$, where the value of λ is $\cdots$.

8. The Bernoulli numbers B_n are defined via the expansion $z/(e^z-1) = \sum_{n=0}^{\infty} B_n\, z^n/n!$. Therefore B_n is given by the contour integral $B_n = \cdots$. (You must specify both the integrand and the contour.)

9. The Chebyshev polynomial of the second kind, $U_n(\cos\theta)$, has the generating function
$$\frac{1}{1 - 2t\,\cos\theta + t^2} = \sum_{n=0}^{\infty} U_n(\cos\theta)\, t^n,$$
where $\theta \in [0, \pi]$. Therefore $U_n(\cos\theta)$ can be expressed as a contour integral in the t-plane given by $U_n(\cos\theta) = \cdots$. (You must specify both the integrand and the contour.)

10. Continuation: Evaluating the contour integral and simplifying the result, the final expression for $U_n(\cos\theta)$ is $U_n(\cos\theta) = \cdots$. (You must express your answer in terms of trigonometric functions of θ.)

11. Continuation: Hence the polynomial $U_1(\cos\theta)$ reduces to $U_1(\cos\theta) = \cdots$.

12. The function $f(z) = (z^2+2)^{1/3}$ has branch points at $z = \cdots$.

13. Express your answer in terms of a Bessel function:
The residue of $f(z) = \exp(z - z^{-1})$ at $z = 0$ is $\operatorname{Res}_{z=0} f(z) = \cdots$.

14. The inverse Laplace transform of $\widetilde{f}(s) = 1/(s^2 - 2s + 1)$ is $f(t) = \cdots$.

15. Let λ be a positive constant. The Laplace transform of the function
$$f(t) = \int_0^t dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} dt_1\, e^{-\lambda(t-t_1)}$$
is $\widetilde{f}(s) = \cdots$.

16. Let
$$f(x) = \begin{cases} 1 - |x|, & |x| \le 1 \\ 0, & |x| > 1. \end{cases}$$
If $\widetilde{f}(k)$ denotes the Fourier transform of $f(x)$, the value of the integral
$$\int_{-\infty}^{\infty} dk\; |\widetilde{f}(k)|^2 = \cdots .$$

17. The positional probability distribution at any time $t \ge 0$ of a random walker on a square lattice with sites labelled by the integers (ℓ, m) is given by
$$P(\ell, m, t) = e^{-\lambda t}\, (p_1/q_1)^{\ell/2}\, (p_2/q_2)^{m/2}\, I_\ell\big(2\lambda t\sqrt{p_1\, q_1}\,\big)\, I_m\big(2\lambda t\sqrt{p_2\, q_2}\,\big),$$

where λ is the mean jump rate and p_i, q_i are directional probabilities such that $p_1 + q_1 + p_2 + q_2 = 1$. The leading asymptotic behavior of $P(\ell, m, t)$ at very long times ($\lambda t \gg 1$) is given by $P(\ell, m, t) \sim \cdots$.

18. The diffusion equation for the positional probability density of a particle diffusing on the x-axis in the region $-\infty < x \leq a$, in the presence of a constant force, is given by

$$\frac{\partial p(x,t)}{\partial t} = -c\,\frac{\partial p\,(x,t)}{\partial x} + D\,\frac{\partial^2 p\,(x,t)}{\partial x^2}.$$

Here c and D are positive constants denoting the drift velocity and diffusion constant, respectively. $p(x,t)$ is normalized to unity for all $t \geq 0$. There is a reflecting boundary at the point $x = a$. The boundary condition satisfied by $p(x,t)$ at $x = a$ is then given by $\cdots$.

19. Continuation: As $t \to \infty$, $p(x,t)$ tends to the stationary probability density $p_{\rm st}(x)$. This quantity satisfies the ordinary differential equation $\cdots$.

20. Continuation: The normalized solution for $p_{\rm st}(x)$ is $p_{\rm st}(x) = \cdots$.

21. A quantum mechanical particle of mass m moving in one dimension has the Hamiltonian $H = p^2/(2m)$, where p is the momentum operator of the particle. Its momentum-space wave function at $t = 0$ is given to be $\phi(p)$. Therefore its momentum-space wave function at any time $t \geq 0$ is given by $\psi(p,t) = \cdots$.

22. The scattering amplitude for a nonrelativistic particle of mass m in a central potential $\lambda V(r)$ is given, in the Born approximation, by

$$f_{\rm B}(k,\theta) = -\,\frac{2m\lambda}{\hbar^2\,Q}\int_0^\infty dr\,r\,\sin{(Qr)}\,V(r),$$

where Q is the magnitude of the momentum transfer vector $\mathbf{Q}$. The forward scattering amplitude in the Born approximation is therefore given by the expression $f_{\rm B}(k,0) = \cdots$.

23. Continuation: The backward scattering amplitude in the Born approximation is therefore given by the expression $f_{\rm B}(k,\pi) = \cdots$.

24. Let $\mathbf{R} = \mathbf{r} - \mathbf{r}\,'$ and $\tau = t - t'$, as usual. Let $G^{(d)}(\mathbf{R},\tau)$ denote the fundamental Green function of the Klein-Gordon operator $\Box + \mu^2$, where μ is a positive constant and $\Box = (1/c^2)(\partial^2/\partial t^2) - \nabla^2$, in $(d+1)$-dimensional spacetime. Then $G^{(d)}(\mathbf{R},\tau)$ can be expressed in the form

$$G^{(d)}(\mathbf{R},\tau) = \frac{1}{(2\pi)^d}\int d^d\mathbf{k}\;\phi(\mathbf{k},\mathbf{R},\tau),$$

where $\phi(\mathbf{k},\mathbf{R},\tau) = \cdots$.

25. Let $\mathbf{J} = (J_1\,,\, J_2\,,\, J_3)$ denote the generators of rotations in three-dimensional space, and let ψ be an arbitrary angle. The quantity $e^{-iJ_1\,\psi}\, J_2\, e^{iJ_1\,\psi}$, expressed as a linear combination of the generators, is $\cdots$.

26. Continuation: Let $\mathbf{n} = (n_1\,,\, n_2\,,\, n_3)$ be an arbitrary unit vector. Then the commutator $[\mathbf{J}\cdot\mathbf{n}\,,\, \mathbf{J}^2] = $.

27. The number of generators of the orthogonal group $O(n)$ and the special orthogonal group $SO(n)$ are, respectively, $\cdots$ and $\cdots$.

28. The number of generators of the unitary group $U(n)$ and the special unitary group $SU(n)$ are, respectively, $\cdots$ and $\cdots$.

29. Let x and p denote the position and momentum operators of a quantum mechanical particle moving in one dimension, so that their commutator $[x\,,\, p] = i\hbar I$, where I is the unit operator. Let a be a real constant with the physical dimensions of length. Using Hadamard's Lemma, the operator $e^{iap/\hbar}\, x\, e^{-iap/\hbar}$ simplifies to
$$e^{iap/\hbar}\, x\, e^{-iap/\hbar} = \cdots\,.$$
This result tells us why the momentum operator p is the generator of translations in position space.

30. Continuation: Let b be a real constant with the physical dimensions of linear momentum. Once again, using Hadamard's Lemma, the operator $e^{-ibx/\hbar}\, p\, e^{ibx/\hbar}$ simplifies to
$$e^{-ibx/\hbar}\, p\, e^{ibx/\hbar} = \cdots\,.$$
This result tells us why the position operator x is the generator of translations in momentum space.

Quiz 2: Solutions

I. True or false:

1. T
2. T
3. F
4. F
5. T
6. F
7. T
8. F
9. T
10. T
11. T
12. F
13. F
14. T
15. T
16. T
17. F
18. T
19. T
20. T
21. T
22. F
23. F
24. T

25. F
26. F
27. T
28. T
29. F
30. T

II. Fill in the blanks:

1. $f(z) = e^{z^2}$.
2. $m = \tan\,\varphi$ and $c = 0$.
3. $|z| < 2$.
4. $\operatorname{Res}\limits_{z=\infty} f(z) = -3$.
5. 1.
6. $\dfrac{(w-w_2)(w_1-w_3)}{(w-w_3)(w_1-w_2)} = \dfrac{(z-z_2)(z_1-z_3)}{(z-z_3)(z_1-z_2)}$.
7. $\lambda = 16$.
8. $B_n = \dfrac{n!}{2\pi i}\displaystyle\oint_C \frac{dz}{z^n\,(e^z-1)}$, where C encloses the origin once in the positive sense.
9. $U_n(\cos\,\theta) = \dfrac{1}{2\pi i}\displaystyle\oint_C \frac{dt}{t^{n+1}\,(1-2t\,\cos\,\theta+t^2)}$, where C encloses the origin once in the positive sense.
10. $U_n(\cos\,\theta) = \dfrac{\sin\,(n+1)\theta}{\sin\,\theta}$.
11. $U_1(\cos\,\theta) = 2\,\cos\,\theta$.
12. $i\sqrt{2}$, $-i\sqrt{2}$, ∞.
13. $-J_1(2)$.
14. $t\,e^t$.
15. $1/[s(s+\lambda)^n]$.
16. $4\pi/3$.

17. $P(\ell, m, t) \sim \dfrac{(p_1/q_1)^{\ell/2}\,(p_2/q_2)^{m/2}\,\exp\,\big[-\lambda t\,\big(1-2\sqrt{p_1\,q_1}-2\sqrt{p_2\,q_2}\big)\big]}{4\pi\lambda t\,(p_1\,q_1\,p_2\,q_2)^{1/4}}\,.$

18. $\Big[D\,\dfrac{\partial p}{\partial x}-c\,p\Big]_{x=a}=0.$

19. $D\,\dfrac{d^2p_{\rm st}(x)}{dx^2}-c\,\dfrac{dp_{\rm st}(x)}{dx}=0.$

20. $p_{\rm st}(x)=(c/D)\,e^{c(x+a)/D}.$

21. $\psi(p,t)=e^{-ip^2t/(2m\hbar)}\,\phi(p).$

22. $f_{\rm B}(k,0)=-\dfrac{2m\lambda}{\hbar^2}\displaystyle\int_0^\infty dr\,r^2\,V(r).$

23. $f_{\rm B}(k,\pi)=-\dfrac{m\lambda}{\hbar^2\,k}\displaystyle\int_0^\infty dr\,r\,V(r)\,\sin\,(2kr).$

24. $\phi(\mathbf{k},\mathbf{R},\tau)=c\,\theta(\tau)\,\dfrac{\sin\,c\tau k}{k}\,e^{i\mathbf{k}\cdot\mathbf{R}}.$

25. $J_2\,\cos\,\psi+J_3\,\sin\,\psi.$

26. $0.$

27. $\frac{1}{2}n(n-1)$ and $\frac{1}{2}n(n-1).$

28. n^2 and $n^2-1.$

29. $x+a.$

30. $p+b.$