Equations of Mathematical Physics

Equations of Mathematical Physics

By

Marian Apostol

Cambridge
Scholars
Publishing

Equations of Mathematical Physics

By Marian Apostol

This book first published 2018

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-1616-4
ISBN (13): 978-1-5275-1616-8

Contents

1 Preface

Mathematics are the Equations of Mathematical Physics. They are based on numbers and mental constructs which we feel to be satisfying and helpful in our endeavour to survive in relation with the world, nature and humans (Planck). The equations are the tools of Theoretical Physics. Both Mathematics and Physics have each their own halo of pseudo-science, a lot of nonsense which goes under the same name as the main scientific core. This is so, because both these sciences are successful and nobody can say what they are in fact. As if they came as strange living beings from an alien, superior world (God's world?).

Wigner would have said, approximately: "The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve"; many believe it is God's gift. But nothing could impose itself upon us, and we would not have accepted such a wonderful gift, without understanding it; except for one thing: our own subjectivity. Mathematics and Physics are our own subjectivity; the difference is made by the quality of our subjectivity. The objectivity is the subjectivity of the "great" men.

Mathematics, as well as Physics, contains a few things which impose themselves upon our mind with necessity, with force, authority, and beauty; or, perhaps, simply, we just recognize them as being quite familiar to us. We do not know, in fact, what they are, nor where they come from. It is in the human soul the perversion to assert our arrogance in matters which we do not understand. Out of this perversion, an immense body of useless, fake, ungracious and, ultimately, pernicious things arise, which go, falsely, under the name of Mathematics or Physics. The most important thing in Mathematics and Physics is to be lucky enough to avoid the nonsense, the monstrous, pathological constructions which bring pain (Landau); such constructions are not even wrong (Pauli). I chose here the basic elements of Mathematics

which I felt are healthy. I have never seen a textbook on Mathematics which was not long, though titles like Methods, Introduction, Course, etc, seem to be meant to convey that only important things were included; nothing more misleading. Mathematics or Physics authors, for their great majority, seem to not compare themselves at all with mathematicians or physicists. I hope this booklet is different, at least by its brevity.

Mathematics has been made by numerous people along many years. I give here a list:

Isaac Newton (1642-1727), Gottfried Wilhelm Leibniz (1646-1714), Jacob Bernoulli (1654-1705), Johann Bernoulli (1667-1748), Brook Taylor (1685-1731), Nicolas Bernoulli (1687-1759), Daniel Bernoulli (1700-1782), Leonhard Euler (1707-1783), Jean le Rond d'Alembert (1717-1783), Joseph-Louis Lagrange (1736-1813), Pierre-Simon Laplace (1749-1827), Adrien-Marie Legendre (1752-1833), Jean Baptiste Joseph Fourier (1768-1830), Carl Friedrich Gauss (1777-1855), Simeon Denis Poisson (1781-1840), Friedrich Bessel (1784-1846), Augustin-Louis Cauchy (1789-1857), George Green (1793-1841), Niels Henrik Abel (1802-1829), Carl Gustav Jacob Jacobi (1804-1851), William Rowan Hamilton (1805-1865), Johann Lejeune Dirichlet (1805-1859), Karl Weierstrass (1815-1897), George Stokes (1819-1903), Arthur Cayley (1821-1895), Hermann von Helmholtz (1821-1894), Leopold Kronecker (1823-1891), Gustav Kirchhoff (1824-1887), William Thomson (Lord Kelvin) (1824-1907), Georg Friedrich Bernhard Riemann (1826-1866), James Clerk Maxwell (1831-1879), Josiah Willard Gibbs (1839-1903), John William Strutt (Lord Rayleigh) (1842-1919), Jean-Gaston Darboux (1842-1917), Ferdinand Georg Frobenius (1849-1917), Oliver Heaviside (1850-1925), Henri Poincare (1854-1912), David Hilbert (1862-1943), Erik Ivar Fredholm (1866-1927), Henri Lebesgue (1875-1941), Peter Debye (1884-1966), Erwin Schrodinger (1887-1961).

They have been decisively aided in their endeavour by many others who remained anonimous.

I include here a list of a few basic books of Mathematics:

E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge Univ. Press, Cambridge (1927); G. N. Watson, *Theory of Bessel Functions*, Cambridge Univ. Press, Cambridge (1922); H. Bateman, *Partial Differential Equations of Mathematical Physics*,

Cambridge, NY (1932); R. Courant and D. Hilbert, *Methods of Mathematical Physics*, vols. 1, 2, Springer, Berlin (1937); A. Sommerfeld, *Partielle Differentialgleichungen der Physik*, *Vorlesungen uber Theoretische Physik*, 3tte Band, Akademische Verlagsgesellschaft, Leipzig (1947); P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, vols. 1, 2, McGraw-Hill, NY (1953); A. N. Tikhonov and A. A. Samarskii, *Equations of Mathematical Physics*, Dover, NY (1963); V. S. Vladimirov, *Equations of Mathematical Physics*, Dekker, NY (1971); M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*, Nat. Bur. Standards, Washington (1964); I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, eds. A. Jeffrey and D. Zwillinger, 6th ed., Academic Press, NY (2000).

The present booklet includes a Chapter on Introductory Elements and Chapters on Differential Equations, Harmonic Oscillator, Laplace, Poisson and Wave Equations, Vector Wave Equations, Quasi-Classical Approximation and Ordinary Differential Equations. More about Complex Variable, Perturbations and Variational Calculus, Integral Equations, or Probabilities, Differential Geometry and Group Theory are specific to various disciplines in Theoretical Physics. Very useful are particular examples of solved problems of Mathematical Physics, which might be included in the present book.

2 Introductory Elements

2.1 Linear Algebra

2.1.1 Vectors

A set $\mathbf{x} = (x_1, x_2, ...x_n)$ of n real numbers, called components, is called an n-dimensional vector. We can add vectors, with usual rules, the null (zero) vector being $0 = (0, 0, ...0)$; we can multiply vectors by scalars λ, μ, etc, and get linear combinations of the form $\lambda\mathbf{x} + \mu\mathbf{y}$; we say that vectors form a vectorial space (with n dimensions), and we can represent vectors as geometrical segments of straight lines pointing from the origin 0 to the point of rectangular coordinates x_i, $i = 1...n$, taken along the corresponding axes. This construction implies a reference frame, and we can rotate vectors (or the reference frame); the scalars remain unchanged under such rotations.

We define an inner (or scalar) product by $\mathbf{xy} = x_i y_i$ (repeating indices being summed), and see that $x^2 = \mathbf{xx} = x_i^2$ is the Pythagora's theorem with $x = \sqrt{x^2}$ the length of the vector; it is also called the norm of the vector $\mathbf{x}$, denoted also by $|\mathbf{x}|$. (The product of a vector by a scalar is an "external" product). The scalar product defines the cosine of the angle α made by two vectors, $\mathbf{xy} = xy\cos\alpha$, which can be seen easily by rotating the vectors such as to superpose one of them along an axis. If $\mathbf{xy} = 0$ the (non-vanishing) vectors $\mathbf{x}$, $\mathbf{y}$ are orthogonal. We have the Schwarz inequality $\mathbf{xy} \leq xy$, since, for instance, $(\lambda\mathbf{x} + \mathbf{y})^2 \geq 0$ as a trinomial in λ. We define the basis vectors $\mathbf{e}_i = (0, 0, ...1, ...0)$, with 1 on the $i - th$ position and 0 in all the other places, and can write $\mathbf{x} = x_i\mathbf{e}_i$. The basis vectors are orthogonal, $\mathbf{e}_i\mathbf{e}_j = \delta_{ij}$, and even orthonormal since $e_i = 1$ for any i; they are unit vectors; $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$ is the Kronecker's symbol. These vectors are said to form a canonical basis. The products $\mathbf{xe}_i/x = x_i/x$ define

the direction cosines of the angles made by vector $\mathbf{x}$ with respect to the references axes.

If the non-vanishing vectors are such that $\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + ...\lambda_m\mathbf{x}_m = 0$ implies all $\lambda_i = 0$, we say that these vectors $\mathbf{x}_{1,...m}$ are linear independent; the basis vectors $\mathbf{e}_i$, $i = 1, ...n$, are linear independent. The basis vectors $\mathbf{e}_i$ and any other (non-vanishing) vector $\mathbf{x}$ are linear dependent, because $\mathbf{x}$ can be written as a linear combination of the basis. This latter statement, concerning the representation of any vector $\mathbf{x}$ by the basis vectors is also called the completeness statement; the orthogonal basis of the n vectors $\mathbf{e}_i$ constructed above is complete. Any set of independent vectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$,... can be orthogonalized (or even orthonormalized) by the Schmidt procedure (Erhard Schmidt (1876-1959)): $\mathbf{v}_1$,$\mathbf{v}_2 + \lambda\mathbf{v}_1$, $\mathbf{v}_3 + \mu\mathbf{v}_1 + \nu\mathbf{v}_2$, etc; requiring the orthogonalization (or orthonormalization) we can find the scalars λ, μ, ν, etc in these combinations.

The construction of vectors described above can be generalized in multiple ways; for instance, including complex numbers, or functions, etc, with a suitable definition of the scalar (inner) product. Functions are represented as linear combinations of infinitely many orthogonal functions; the scalar product of two functions f and g can be defined as the integral of the product f^*g.

2.1.2 Matrices

A (square) matrix $A = (a_{ij})$, i, $j = 1, 2, ...n$ is a set of numbers a_{ij}, called elements, arranged in rows and columns,

$$A = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & & & \\ a_{n1} & a_{n2} & ... & a_{nn} \end{pmatrix} , \tag{2.1}$$

which appears in the system of linear equations

$$A\mathbf{x} = \mathbf{y} \ , \ a_{ij}x_j = y_i \ ; \tag{2.2}$$

it defines a linear transformation, since $A(\lambda\mathbf{x}_1 + \mu\mathbf{x}_2) = \lambda A\mathbf{x}_1 + \mu A\mathbf{x}_2 = \lambda\mathbf{y}_1 + \mu\mathbf{y}_2$. The inversion of the matrix A gives the solution $\mathbf{x}$ of the system of equations (2.1), and it is written formally

as $\mathbf{x} = A^{-1}\mathbf{y}$; we can define summation and multiplication of the matrices, the latter by

$$a_{ij}b_{jk} = c_{ij} \ , \ AB = C \ , \tag{2.3}$$

which implies the multiplication of the rows of A by the columns of B. We can see that the multiplication of matrices is not commutative. We can also define the multiplication of a matrix by a scalar, which implies the multiplication by that scalar of all the elements of the matrix. The inverse A^{-1} is that matrix with the property $A^{-1}A = AA^{-1} = 1$, where the unit matrix 1 has unity for all its diagonal elements and zero for all the others, $1_{ij} = \delta_{ij}$.

The solution of the system of equations (2.2) and, implicitly, the inverse matrix A^{-1} are obtained by elementary calculations, which usually involve determinants. The generalization of such calculations is called the fundamental theorem of linear systems of equations. For $\mathbf{y} \neq 0$ (inhomogeneous system of equations) it may happen that not all of the n equations (2.2) are independent; suppose that only $r \leq n$ are independent, the remaining $n - r$ equations being obtained from the independent ones by multiplication by scalars and addition; they are linear combinations of the others; then, $n - r$ unknowns are free parameters and the system is solved by the r-determinant; r is called the rank of the matrix A (or of the system) and the system is said to be undetermined. We can see that the solution is given by the non-vanishing determinant of the highest order (rank), since if a determinant has one or more lines (or columns) which are linear combinations of the others (*i.e.* they are not independent), then it is vanishing. We can see that the matrix A is inversible, *i.e.* there exists the inverse A^{-1}, if and only if $detA \neq 0$. If $\mathbf{y} = 0$ (homogeneous system), the solution of the system of equations (2.2) is, in general, $\mathbf{x} = 0$, unless the determinant of the matrix A is vanishing ($detA = 0$); in that case, one or more equations are dependent on the others, a corresponding number of unknowns are free (undetermined), and the system is solved again by the remaining non-vanishing determinant of the highest order.

In general, the elements of a matrix A may be complex numbers. $A^* = (a^*_{ij})$ is the conjugate matrix, $A^{'} = (a^{'}_{ij}) = (a_{ji})$ is called the transposed matrix, or the transpose of A, $A^{+} = (a^*_{ji})$ is called the

adjoint matrix, or the adjoint of A; if $A = A^+$, the matrix is said to be hermitian, if $A = A^{'}$, the matrix is symmetric, if a real matrix is such that $AA^{'} = A^{'}A = 1$, it is called orthogonal, if a matrix is such that $AA^+ = A^+A = 1$, the matrix is said to be unitary. The inverse matrix is obtained as $A^{-1} = (\overline{a}_{ij})/detA$, where $\overline{a}_{ij}$ is the co-factor of a_{ij}: we take the transpose, compute the determinant obtained by removing the $i-th$ line and the $j-th$ column in this matrix, multiply this determinant by $(-1)^{i+j}$ and get the co-factor. The absolute value of the determinant of a unitary matrix is equal to unity, the determinant of an orthogonal matrix is equal to ± 1.

2.1.3 Quadratic forms. Diagonalization

A quadratic form is the scalar product of $\mathbf{x}$ by $A\mathbf{x} = \mathbf{y}$, denoted now by $(\mathbf{x}, A\mathbf{x}) = \mathbf{xy} = a_{ij}x_ix_j$; if $\mathbf{x}$ is a complex vector, the definition is $(\mathbf{x}, A\mathbf{x}) = \mathbf{x}^*\mathbf{y} = a_{ij}x_i^*x_j$; the scalar product in this case reads $(\mathbf{x}, \mathbf{y}) = \mathbf{x}^*\mathbf{y} = x_i^*y_i$; we note that $(\mathbf{x}, \mathbf{y})^* = (\mathbf{y}, \mathbf{x})$; obviously, $(\mathbf{x}, A\mathbf{x}) = (A^{'}\mathbf{x}, \mathbf{x})$ or $(\mathbf{x}, A\mathbf{x}) = (A^+\mathbf{x}, \mathbf{x})$, where $A^{'}$ is the transpose and A^+ is the adjoint of A. Similarly, a bilinear form is defined as $(\mathbf{y}, A\mathbf{x}) = a_{ij}y_ix_j$, with similar properties. If the matrix is symmetric or hermitian, the forms are said to be symmetric or hermitian, respectively. We can represent the vectors by columns of their components, then the vector on the left in the scalar product is a row, and the scalar product obeys the multiplication rule of matrices (*i.e.* rows by columns).

In the system of equations

$$A\mathbf{x} = \lambda\mathbf{x} \ , \tag{2.4}$$

$\mathbf{x}$ is called an eigenvector and the scalar λ is called an eigenvalue (or characteristic value) of the matrix A. The system can be solved providing $det(A - \lambda \cdot 1) = 0$, where 1 denotes the unit matrix. This is an algebraic equation of order n in λ, called the characteristic equation, which, in general, has n (complex) solutions for λ, some of them multiple; the multiple eigenvalues are called degenerate eigenvalues. The set of eigenvalues is called the spectrum of A. Therefore, there are n eigenvectors and n eigenvalues, so that we can write

$$A\mathbf{e}^{(i)} = \lambda_i\mathbf{e}^{(i)} \tag{2.5}$$

for the $i-th$ eigenvector $\mathbf{e}^{(i)}$; and we can take the bilinear form

$$(\mathbf{e}^{(j)}, A\mathbf{e}^{(i)}) = (A^{+}\mathbf{e}^{(j)}, \mathbf{e}^{(i)}) = \lambda_i(\mathbf{e}^{(j)}, \mathbf{e}^{(i)}) \ . \tag{2.6}$$

Now, suppose that $A = A^{+}$ is hermitian; then

$$(A^{+}\mathbf{e}^{(j)}, \mathbf{e}^{(i)}) = (A\mathbf{e}^{(j)}, \mathbf{e}^{(i)}) = \lambda_j^{*}(\mathbf{e}^{(j)}, \mathbf{e}^{(i)}) \tag{2.7}$$

and from equations (2.6) and (2.7) we get

$$0 = (\lambda_i - \lambda_j^{*})(\mathbf{e}^{(j)}, \mathbf{e}^{(i)}) \tag{2.8}$$

by subtraction; therefore, for $i = j$ we can see that the eigenvalues are real, while for $i \neq j$ we can see that the eigenvectors are orthogonal; we can take them orthonormal, $(\mathbf{e}^{(j)}, \mathbf{e}^{(i)}) = \delta_{ij}$, or, by components,

$$e_k^{(j)*} e_k^{(i)} = \delta_{ij} \ ; \tag{2.9}$$

we define the matrix $C_{ij} = e_i^{(j)}$, *i.e.* we arrange the eigenvectors as columns in a matrix; the orthogonality equation (2.9) reads

$$C_{kj}^{*} C_{ki} = \delta_{ij} \ \ , \tag{2.10}$$

or $(C^{+})_{jk} C_{ki} = \delta_{ij}$, *i.e.*

$$C^{+}C = 1 \ \ , \tag{2.11}$$

which means that the matrix C is a unitary matrix, since $C^{+} = C^{-1}$; and we have also $CC^{+} = 1$. Equation (2.6) becomes

$$\begin{gathered} e_k^{(j)*} a_{kl} e_l^{(i)} = \lambda_i \delta_{ij} \ , \ C_{kj}^{*} a_{kl} C_{li} = \lambda_i \delta_{ij} \ , \\ (C^{+})_{jk} a_{kl} C_{li} = \lambda_i \delta_{ij} \ , \ C^{+}AC = (\lambda_i \delta_{ij}) = \widetilde{A} \ ; \end{gathered} \tag{2.12}$$

we can see that the hermitian matrix A ($A^{+} = A$) can be brought to the diagonal form denoted $\widetilde{A}$ (with eigenvalues λ_i on the diagonal) by a linear, unitary transformation C ($C^{+} = C^{-1}$), formed by the orthogonal eigenvectors arranged in columns. Usually, the eigenvectors corresponding to a multiple eigenvalue can be orthogonalized, so we have the diagonalizing matrix C. It is said in general that a matrix A is brought to its diagonal form by a linear transformation C which

transforms the equation $A\mathbf{x} = \lambda\mathbf{x}$ into $AC\mathbf{e} = \lambda C\mathbf{e}$, through $\mathbf{x} = C\mathbf{e}$, *i.e.* $C^{-1}AC\mathbf{e} = \widetilde{A}\mathbf{e} = \lambda\mathbf{e}$; for hermitian matrices A the transformation is unitary, $C^{-1} = C^{+}$; for symmetric matrices A (real matrices), the transformation C is orthogonal; in this case $C^{-1} = C^{+} = C^{'}$, where $C^{'}$ is the transpose of C.

The existence of the unitary matrix C, $C_{ij} = e_i^{(j)}$, formed by eigenvectors has far-reaching consequences. From the orthogonality of these vectors $e_k^{(i)*}e_k^{(j)} = \delta_{ij}$ we deduce the unitarity of the matrix C, $C_{ki}^{*}C_{kj} = (C^{+})_{ik}C_{kj} = \delta_{ij}$, $C^{+}C = 1$; as well as $CC^{+}CC^{+} = CC^{+}$, which implies $CC^{+} = 1$. More important, since the eigenvectors are orthonormal, we can take them as canonical basis vectors, so that they satisfy

$$e_i^{(k)*}e_j^{(k)} = \delta_{ij} \ , \tag{2.13}$$

which means $C_{ik}^{*}C_{jk} = (C^{+})_{ki}C_{jk} = C_{jk}(C^{+})_{ki} = \delta_{ij}$, *i.e.* $CC^{+} = 1$ (for complex canonical unit vectors we should use i instead of 1). Then, the components of any vector $\mathbf{x}$ can be written as

$$x_i = \delta_{ij}x_j = e_k^{(i)*}e_k^{(j)}x_j = C_{ik}y_k \ , \ y_k = C_{ki}x_i \ ; \tag{2.14}$$

this means that any component x_i can be written as an expansion of components y_j, and, conversely, these components y_i can be written as an expansion of components x_i; this is due to the existence of the orthogonality relations $e_k^{(i)*}e_k^{(j)} = \delta_{ij}$ and $e_i^{(k)*}e_j^{(k)} = \delta_{ij}$. We note that the above equations can also be read as expansions in eigenvectors, and this possibility is also provided by the orthogonality relations. These orthogonality relations between the eigenvectors make possible the completeness relations, because they ensure the expansions described above.

The vectors are usually generalized to functions, say $\varphi_n(x)$, where n may be discrete or continuous and both x and n may be multicomponent; the scalar product is defined by the integral $\int dx\varphi_n^{*}(x)\varphi_m(x)$. Sometimes the scalar product may include a weight function. The matrices are usually generalized to operators, either differential or integral, or various other forms (if function φ is viewed as a variable, the result of the application of an operator on it is called a functional); we write symbolically $O\varphi$ for the action of an operator O on the function φ. The problem of eigenvectors (eigenfunctions) and eigenvalues read

$O\varphi = \lambda\varphi$. If there is a set of eigenvectors φ_n (usually infinite), which are orthonormal, then we can establish completeness (orthogonality) relations like $(\varphi_n, \varphi_m) = \delta_{nm}$ and $\varphi_n^*(x)\varphi_n(y) = \delta(x-y)$ (possibly with a weight function); in these conditions we can have expansions like $f = f_n\varphi_n(x)$, $f_n = (\varphi_n, f)$. It is worth noting that the role of the labels i, j for the matrices is played here by the labels n and x.

For a symmetric matrix A we have, for a given eigenvalue λ and the corresponding eigenvector x,

$$A\mathbf{x} = \lambda\mathbf{x} \ , \ a_{ij}x_j = \lambda x_i \ , \tag{2.15}$$

or the quadratic form

$$(\mathbf{x}, A\mathbf{x}) = a_{ij}x_ix_j = \lambda x_i^2 = \lambda x^2 \ . \tag{2.16}$$

We can see that the effect of A on one of its eigenvectors $\mathbf{x}$ is to multiply this vector by the corresponding eigenvalue λ (without changing its direction, *i.e.* without rotating it). Now, we can take the variation of equation (2.16) with respect to x_i, and get $a_{ij}x_j = \lambda x_i$, which means that the difference $a_{ij}x_ix_j - \lambda x^2$, *i.e.* the difference $(\mathbf{x}, A\mathbf{x}) - \lambda(\mathbf{x}, \mathbf{x})$, has an extremal, vanishing value for the eigenvector $\mathbf{x}$; it follows that the surface defined by the quadratic form $a_{ij}x_ix_j = const$ for various $\mathbf{x}$ is tangent to the sphere defined locally by λx^2, *i.e.* defined in the vicinity of the eigenvector $\mathbf{x}$. If the (symmetric) matrix A is brought to its diagonal form by the (orthogonal) matrix C, it follows that the quadratic form $(\mathbf{x}, A\mathbf{x})$ for a general $\mathbf{x}$ becomes

$$\begin{aligned} (\mathbf{x}, A\mathbf{x}) = a_{ij}x_ix_j = C_{ik}a_{ij}C_{jl}y_ky_l = \\ = (C^{'})_{ki}a_{ij}C_{jl}y_ky_l = (\widetilde{A})_{kl}y_ky_l = \lambda_k y_k^2 \end{aligned} \tag{2.17}$$

through the transformation $\mathbf{x} = C\mathbf{y}$; it is said that the quadratic form is brought to its principal, or normal, axes, which are defined by the eigenvectors; they define the symmetry axes of the quadratic surface corresponding to the quadratic form (of course, the surface may be degenerate); the transformation C is equivalent with a change of basis, from the original basis in which the vector $\mathbf{x}$ is expressed to the basis formed by the eigenvectors, where the vector $\mathbf{x}$ becomes the vector $\mathbf{y}$. A quadratic form is said to be positive definite if it is positive for any

vector $\mathbf{x}$; for this it should have all the eigenvalues positive. Finally we note that a rotation about an axis is an orthogonal transformation, its determinant being equal to unity; the orthogonal transformations preserve the length of the vectors, $x_i^2 = C_{ij}C_{ik}y_jy_k = (C^{'})_{ji}C_{ik}y_jy_k = y_j^2$.

Since $(A-\lambda_i\cdot 1)\mathbf{e}^{(i)} = 0$, then $(A-\lambda_1\cdot 1)(A-\lambda_2\cdot 1)...(A-\lambda_n\cdot 1)\mathbf{x} = 0$ has solution $\mathbf{x} = \alpha_i\mathbf{e}^{(i)}$, with arbitrary scalars α_i; this means $(A-\lambda_1\cdot 1)(A-\lambda_2\cdot 1)...(A-\lambda_n\cdot 1) = 0$, which is the characteristic equation $f(\lambda) = det(A-\lambda\cdot 1) = (\lambda-\lambda_1)(\lambda-\lambda_2)...(\lambda-\lambda_n) = 0$ written for the matrix A, $f(A) = 0$; a matrix satisfies its own characteristic equation, a result which is called the Hamilton-Cayley theorem.

2.1.4 Bessel inequality

Let $\mathbf{x}$ be a vector and $c_i\mathbf{e}_i$ a linear combination, the coefficients c_i being given by the scalar products $c_i = (\mathbf{e}_i, \mathbf{x})$ of $\mathbf{x}$ with the orthonormalized (basis) vectors $\mathbf{e}_i$; then, since

$$\begin{aligned} |\mathbf{x} - c_i\mathbf{e}_i|^2 = (\mathbf{x} - c_i\mathbf{e}_i, \mathbf{x} - c_i\mathbf{e}_i) = \\ = |\mathbf{x}|^2 - c_i^*(\mathbf{e}_i, \mathbf{x}) - c_i(\mathbf{x}, \mathbf{e}_i) + |c_i|^2 = |\mathbf{x}|^2 - |c_i|^2 \geq 0 \ , \end{aligned} \tag{2.18}$$

we have the Bessel inequality

$$|\mathbf{x}|^2 - |c_i|^2 \geq 0 \ ; \tag{2.19}$$

the equality is called Parseval's equality (Marc-Antoine Parseval (1755-1836)), or, sometimes, the completeness relation, since, in this case, $\mathbf{x}$ can be written as

$$\mathbf{x} = c_i\mathbf{e}_i \ ; \tag{2.20}$$

we have an expansion of $\mathbf{x}$ in a series of $\mathbf{e}_i$ and c_i are the expansion coefficients; the series $|c_i|^2$ is convergent. In general, $c_i\mathbf{e}_i$ is said to approximate $\mathbf{x}$ "in the mean", since the equality (2.20) follows by minimizing the "mean square error" $|\mathbf{x} - c_i\mathbf{e}_i|^2$ with respect to the vectors $\mathbf{e}_i$ or the coefficients c_i.

2.2 Integral Equations

2.2.1 Fredholm equations

Let $K(x, y)$ be a function, called a kernel, $f(x)$ a given function and $\varphi(x)$ an unknown function; the equation

$$f(x) = \varphi(x) - \lambda \int dy K(x, y)\varphi(y) \tag{2.21}$$

is called an integral equation of the second kind (or Fredholm equation); the integration is performed over a fundamental domain $a < x, y < b$; in general, it is convenient to have continuous (and real) functions and finite integrals (non-singular kernels); obviously, the transformation is linear; λ is called the parameter of the equation; equation (2.21) is inhomogeneous; the equation

$$\varphi(x) = \lambda \int dy K(x, y)\varphi(y) \tag{2.22}$$

is homogeneous. Obviously, there is a close resemblance to the algebra of coupled linear equations; the integral in the equations written above is also called an integral transform (a linear transformation). The parameter λ can be viewed as an eigenvalue parameter (in fact, its reciprocal); the homogeneous equation (2.22) looks like an eigenvalue equation, whose independent eigenvectors (eigenfunctions) could, in principle, be taken as being normalized and especially orthonormalized; defining the scalar product of two functions f and g by

$$(f, g) = \int dx f(x)g(x) \ . \tag{2.23}$$

For two eigenfunctions $\varphi_{1,2}$ and a symmetrical kernel we get immediately

$$(\lambda_1^{-1} - \lambda_2^{-1})(\varphi_1, \varphi_2) = 0 \ , \tag{2.24}$$

whence we can see that the eigenfunctions corresponding to distinct eigenvalues are orthogonal.

The equation

$$f(x) = \int dy K(x, y)\varphi(y) \tag{2.25}$$

is called a linear integral equation of the first kind; its solution amounts to an inversion of the kernel; actually, the equation can be viewed as an integral transform *(e.g.*, Fourier, Laplace, etc); so that the unknown function is obtained by the inverse transform. Generalization to hermitian kernels are worthwhile.

2.2.2 Degenerate kernels

Any kernel can have, in principle, an expansion in a set of orthonormal functions, so it can be written as

$$K(x,y) = \alpha_i(x)\beta_i(y) \ ; \tag{2.26}$$

it is called a separable kernel; the integral equation becomes

$$f(x) = \varphi(x) - \lambda\alpha_i(x)\int dy\beta_i(y)\varphi(y) \ \ , \tag{2.27}$$

or

$$\begin{gathered}\int dx\beta_i(x)f(x) = \int dx\beta_i(x)\varphi(x) - \\ -\lambda\int dx\beta_i(x)\alpha_j(x)\int dy\beta_j(y)\varphi(y) \ ;\end{gathered} \tag{2.28}$$

denoting

$$\begin{gathered}f_i = \int dx\beta_i(x)f(x) \ , \ \varphi_i = \int dx\beta_i(x)\varphi(x) \ , \\ K_{ij} = \int dx\beta_i(x)\alpha_j(x) = \int dxdyK(x,y)\beta_i(x)\beta_j(y) \ \ ,\end{gathered} \tag{2.29}$$

we get a system of linear equations

$$f_i = \varphi_i - \lambda K_{ij}\varphi_j \ \ , \tag{2.30}$$

which, in general, is infinitely dimensional. The solution is established (at least in principle) by successive approximations of finite dimensions. This is Fredholm's theory of integral equations.

Let us assume that φ_i are the orthonormal eigenfunctions of the kernel (which is symmetric), with eigenvalues λ_i:

$$\lambda_i\int dyK(x,y)\varphi_i(y) = \varphi_i(x) \ ; \tag{2.31}$$

and let us assume the expansions

$$\begin{aligned} f(x) = f_i\varphi_i(x)\ ,\ \varphi(x) = c_i\varphi_i(x)\ , \\ f_i = (\varphi_i, f) = \int dx\varphi_i(x)f(x)\ ,\ c_i = \int dx\varphi_i(x)\varphi(x)\ ; \end{aligned} \tag{2.32}$$

the integral equation gives

$$c_i = \frac{\lambda_i}{\lambda_i - \lambda} f_i \tag{2.33}$$

and the solution can be represented as

$$\varphi(x) = c_i\varphi_i(x) = f(x) + \frac{\lambda}{\lambda_i - \lambda} f_i\varphi_i(x)\ ; \tag{2.34}$$

this representation is called the resolvent of the integral equation and the series is called the Neumann series. By iterating the integral equation we get a new kernel which can be written as

$$\overline{K}(x, y) = K(x, y) + \lambda \int dz K(x, z)\overline{K}(z, y)\ , \tag{2.35}$$

which implies a Neumann series.

2.2.3 Volterra equation

The equation

$$f(x) = \varphi(x) - \lambda \int_a^x dy K(x, y)\varphi(y) \tag{2.36}$$

is called a Volterra-type equation (Vito Volterra (1860-1940)); usually, transforming it into a differential equation by taking the derivatives is useful in getting the solution. An example is Abel's integral equation

$$f(x) = \int_0^x dy \frac{\varphi(y)}{\sqrt{x - y}}\ , \tag{2.37}$$

whose solution is

$$\varphi(x) = \frac{1}{\pi} \int_0^x dy \frac{f'(y)}{\sqrt{x - y}} \tag{2.38}$$

(for $f(0) = 0$).

2.3 Calculus of Variations

2.3.1 Extrema points

A function $f(x, y, ...)$ of real variables has an extremal value at x_0, y_0, ... given by

$$f_x(x_0, y_0, ...) = f_y(x_0, y_0, ...) = ... = 0 \ , \tag{2.39}$$

where $f_x = \partial f/\partial x$, $f_y = \partial f/\partial y$, etc; if there are constraints like $g_1(x, y, ...) = 0$, $g_2(x, y, ...) = 0$, ... then the extremal value is attained at x_0, y_0, ... given by

$$\begin{gathered} F_x(x_0, y_0, ...) = F_y(x_0, y_0, ...) = ... = 0 \ , \\ \frac{\partial F}{\partial \lambda_1} = \frac{\partial F}{\partial \lambda_2} = ... = 0 \ , \end{gathered} \tag{2.40}$$

where

$$F(x, y, ...) = f(x, y, ...) - \lambda_1 g(x, y, ...) - \lambda_2 g_2(x, y, ...) - ... \tag{2.41}$$

is called the Lagrange functional and $\lambda_{1,2,...}$ are called Lagrange's multipliers. The points x_0, y_0, ... are called the extrema points, or the stationary points. The question as to whether an extrema is a point of maximum, minimum or inflexion (saddle point) depends on the examination of the quadratic form obtained by a series expansion of the function in the vicinity of the extrema (involving the second-order derivatives).

2.3.2 Variational problems

The length of a curve given by $y = y(x)$ is $\sqrt{dx^2 + dy^2} = \sqrt{1 + y'^2}dx$; the problem can be to determine the curve with the shortest distance between two points, *i.e.* looking for the minimum value of

$$\int_{x_1}^{x_2} dx \sqrt{1 + y'^2} \tag{2.42}$$

for various functions $y(x)$; nullifying the variation leads to

$$\begin{aligned} \delta \int_{x_1}^{x_2} dx \sqrt{1+y'^2} = \int_{x_1}^{x_2} dx \frac{y'}{\sqrt{1+y'^2}} \delta(y') = \\ = \frac{y'}{\sqrt{1+y'^2}} \delta y \Big|_{x_1}^{x_2} - \int_{x_1}^{x_2} dx \frac{d}{dx} \left(\frac{y'}{\sqrt{1+y'^2}} \right) \delta y = 0 \ , \end{aligned} \tag{2.43}$$

or

$$\frac{d}{dx} \left(\frac{y'}{\sqrt{1+y'^2}} \right) = \frac{y''}{\sqrt{1+y'^2}} - \frac{y'^2}{(1+y'^2)^{3/2}} y'' = 0 \ , \tag{2.44}$$

since we require the variations δy at the ends $x_{1,2}$ of the interval to vanish; from equation (2.44) we get $y'' = 0$, *i.e.* $y = ax + b$, where a, b are constants, *i.e.* the curve should be a straight line; as expected. The element of area of a surface given by $z = z(x, y)$ is obtained by

$$\sqrt{dx^2 + z_x^2 dx^2} \sqrt{dy^2 + z_y^2 dy^2} = \sqrt{1 + z_x^2 + z_y^2} dx dy \ ; \tag{2.45}$$

the length of a line arc on a surface given by parametric equations $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ is $\sqrt{e + 2fv' + gv'^2} du$, where $e = x_u^2 + y_u^2 + z_u^2$, $f = x_u x_v + y_u y_v + z_u z_v$ and $g = x_v^2 + y_v^2 + z_v^2$; we can ask about the shortest distance on such a surface, *i.e.* a geodesic, which amounts to get the minimum of the integral

$$l = \int_{u_1}^{u_2} du \sqrt{e + 2fv' + gv'^2} \tag{2.46}$$

with respect to the function $v(u)$ which indeed defines a line. The shortest time along a path $y = y(x)$ is given by the minimum of the integral

$$t = \int_{x_1}^{x_2} \frac{\sqrt{1+y'^2}}{v(x, y)} dx \ , \tag{2.47}$$

where v is the velocity (Fermat principle; Pierre Fermat (1601(7)-1665)); if a body falls freely in the gravitational field at the surface of the Earth, along the distance y it acquires the velocity $\sqrt{2gy}$; the

corresponding curve is called a brachistochrone; a curve $y = y^{'}(x)$ gives a surface of revolution

$$2\pi \int_{x_1}^{x_2} dxy\sqrt{1+y^{'2}} \; ; \tag{2.48}$$

we can ask about its minimum value; and so on. Such integrals which include unknown functions are called functionals. Series expansions for the unknown functions, or, in general, sequences approximating the unknown functions can be useful in treating extremal problems. The problem can be generalized to several unknown functions. The equations resulting by taking the variations are called the Euler-Lagrange equations.

2.4 Fourier Transform

2.4.1 Delta function

The δ-function is defined as

$$\delta(x) = \frac{1}{2\pi}\int dk \cdot e^{ikx} \;,\; \delta(k) = \frac{1}{2\pi}\int dx \cdot e^{ikx}, \tag{2.49}$$

for real x and k. We can check that $\delta(0) = \infty$ and $\delta(x \neq 0) = 0$. For the latter, we have

$$\int_{-L}^{+L} dke^{ikx-\mu|k|} = \frac{e^{iLx-\mu L}-1}{ix-\mu} + \frac{1-e^{-iLx-\mu L}}{ix+\mu} \to 0 \tag{2.50}$$

for $L \to \infty$, $\mu \to 0^+$ and $x \neq 0$. In addition,

$$\int dx\delta(x) = 1 \;, \tag{2.51}$$

since

$$\begin{aligned} \int dx \int_{-L}^{+L} dke^{ikx-\mu|k|} &= \int dx \left(\tfrac{1}{ix+\mu} - \tfrac{1}{ix-\mu}\right) = \\ &= \int dx \tfrac{2\mu}{x^2+\mu^2} = \int dx \tfrac{2}{x^2+1} = \\ &= 2\arctan x|_{x=-\infty}^{x=+\infty} = 2\pi \;. \end{aligned} \tag{2.52}$$

We note the convergence factor $e^{-\mu|k|}$ in equation (2.50), the order of the limits $L \to \infty$, $\mu \to 0^+$ and the independence of integral (2.52) of the factor μ.

The properties $\delta(0) = \infty$, $\delta(x \neq 0) = 0$ and $\int dx\delta(x) = 1$ define the δ-function. We can check that

$$\lim_{\mu\to 0} \frac{1}{\pi} \frac{\mu}{x^2 + \mu^2} = \delta(x) \tag{2.53}$$

and

$$\lim_{L\to\infty} \frac{1}{\pi} \frac{\sin Lx}{x} = \delta(x) \ , \tag{2.54}$$

by using the integrals given above. In particular

$$\int dx \frac{\sin x}{x} = \pi \tag{2.55}$$

and

$$\lim_{\Delta\to 0} \frac{1}{\sqrt{\pi\Delta}} e^{-\frac{x^2}{\Delta}} = \delta(x) \ . \tag{2.56}$$

Fourier and Cauchy made an implicit use of the δ-function, which was recognized around 1930 by Dirac, on the occasion of the Quantum Mechanics (Paul Dirac (1902-1984)). Note that

$$\delta(x) = \frac{1}{2\pi} \int dk \cos kx \tag{2.57}$$

by equation (2.49), since $\sin kx$ is an odd function and cancels out the integral.

Since $\delta(x)$ is highly peaked on $x = 0$, we may take the integral

$$\int dx\delta(x)f(x) \tag{2.58}$$

about $x = 0$, for any function $f(x)$, which gives obviously

$$\int dx\delta(x)f(x) = f(0) \int dx\delta(x) = f(0) \ ; \tag{2.59}$$

similarly we have

$$f(x) = \int dx' f(x')\delta(x - x') \ . \tag{2.60}$$

In this sense δ is also called a distribution, *i.e.* a functional which associates a number to a function.

2.4.2 Fourier transform

From equations (2.49) and (2.60) we can write

$$f(x) = \frac{1}{2\pi} \int dk e^{ikx} \left[\int dx' f(x') e^{-ikx'} \right] \ , \tag{2.61}$$

or

$$f(x) = \frac{1}{2\pi} \int dk e^{ikx} g(k) \ , \ g(k) = \int dx f(x) e^{-ikx} \ ; \tag{2.62}$$

$g(k)$ is called the Fourier transform of $f(x)$ and $f(x)$ is called the Fourier transform of $g(k)$; it is also said that $f(x)$ and $g(k)$ are expanded as Fourier integrals (in trigonometric functions e^{ikx}); equation (2.61) is called the completeness equation, while equation (2.49) written as

$$\delta(x - x') = \frac{1}{2\pi} \int dk e^{ik(x-x')} \tag{2.63}$$

is called the orthogonality (or orthonormality) equation of the trigonometric functions; integrals like $\int f^* g$ are scalar products for vectors f and g, which are said to form a Hilbert space because they have a scalar product.

The δ-function and the Fourier transform can directly be extended to several dimensions by, for instance,

$$\begin{gathered} \delta(\mathbf{r}) = \delta(x)\delta(y)\delta(z) \ , \ \delta(\mathbf{r}) = \tfrac{1}{(2\pi)^3} \int d\mathbf{k} e^{i\mathbf{kr}} \ , \\ f(\mathbf{r}) = \tfrac{1}{(2\pi)^3} \int d\mathbf{k} e^{i\mathbf{kr}} g(\mathbf{k}) \ , \ g(\mathbf{k}) = \int d\mathbf{r} f(\mathbf{r}) e^{-i\mathbf{kr}} \ , \end{gathered} \tag{2.64}$$

where $\mathbf{kr} = k_x x + k_y y + k_z z$.

The Fourier integral has an interesting minimum (extremal) property; indeed, if we look for the minimum value of the integral

$$\int dx \left| f(x) - \frac{1}{2\pi} \int dk g(k) e^{ikx} \right|^2 \ , \tag{2.65}$$

then we must set the variation of this integral with respect to the coefficients $g(k)$ equal to zero, *i.e.*

$$\int dx \left[f(x) - \frac{1}{2\pi} \int dk g(k) e^{ikx} \right] e^{-ik'x} = 0 \ ; \tag{2.66}$$