CHAPTER I

INTRODUCTION

1·1. Definitions and examples. An equation in which an unknown function appears under one or more signs of integration is called an *integral equation*. For example, the equations

$$y(s)=\int_a^b K(s,t)\,x(t)\,dt \quad (a\leqslant s\leqslant b), \tag{1}$$

$$x(s)=y(s)+\int_a^b K(s,t)\,x(t)\,dt \quad (a\leqslant s\leqslant b), \tag{2}$$

$$x(s)=\int_a^b K(s,t)\,[x(t)]^2\,dt \quad (a\leqslant s\leqslant b), \tag{3}$$

$$x(s)=\int_a^b\int_a^b K(s,t,u)\,x(t)\,x(u)\,dt\,du \quad (a\leqslant s\leqslant b), \tag{4}$$

in each of which $x(s)$ is the unknown function, and all other functions are regarded as given, are integral equations.

Equations (1) and (2) can be written in the form

$$L[x(s)]=y(s), \tag{5}$$

where the expression $L[x(s)]$ containing the unknown function $x(s)$ is *linear* in the sense that

$$L[\lambda_1 x_1(s)+\lambda_2 x_2(s)]=\lambda_1 L[x_1(s)]+\lambda_2 L[x_2(s)]$$

for any constants λ_1 and λ_2. Thus, for equation (1),

$$L[x(s)]=\int_a^b K(s,t)\,x(t)\,dt,$$

and for (2), $$L[x(s)]=x(s)-\int_a^b K(s,t)\,x(t)\,dt.$$

Equations of the type (5), where $L[x(s)]$ is a linear expression in which $x(s)$ appears under one or more signs of integration, are called *linear integral equations*; thus (1) and (2) are linear, whereas (3) and (4) are not. In this book we shall only be concerned with linear equations.

Linear integral equations of the form (1), where the unknown function appears under the sign of integration and nowhere else in the equation, are called equations of the *first kind*. Those of the form (2), where the unknown function appears both under the sign of integration and elsewhere in the equation, are said to be of the *second kind*. In both (1) and (2), the function $K(s,t)$ is called the *kernel*† of the equation; it is defined in the square $a \leqslant s \leqslant b$, $a \leqslant t \leqslant b$ of the (s,t) plane. The equation is completely specified by giving the interval (a,b), the kernel $K(s,t)$ and the function $y(s)$. In this book we shall be mainly concerned with equations of the second kind, which are the most amenable to general theoretical treatment, but we shall have something to say about equations of the first kind in Chapter VIII.

If we take $y(s)=0$ in (2), we obtain the *homogeneous* equation of the second kind

$$x(s)=\int_a^b K(s,t)\,x(t)\,dt \quad (a \leqslant s \leqslant b). \tag{6}$$

When (2) and (6) are considered simultaneously, (6) is called the homogeneous equation *associated* with (2).

Since an integral equation is usually required to hold for all (or almost all) values of the variable in the range of integration, we shall frequently omit the '$(a \leqslant s \leqslant b)$' on the right-hand side of the equation.

It is often convenient to introduce a parameter λ into equations of the second kind, which then assume the form

$$x(s)=y(s)+\lambda\int_a^b K(s,t)\,x(t)\,dt \quad (a \leqslant s \leqslant b).$$

We can then obtain useful information by studying what happens when λ is allowed to vary in the complex plane.

If $K(s,t)=0$ when $s<t$, equation (2) can be written

$$x(s)=y(s)+\int_a^s K(s,t)\,x(t)\,dt \quad (a \leqslant s \leqslant b),$$

with a variable upper limit of integration. Such an equation is called a *Volterra equation* of the second kind. Similarly,

$$y(s)=\int_a^s K(s,t)\,x(t)\,dt \quad (a \leqslant s \leqslant b)$$

is a Volterra equation of the first kind.

† Called 'noyau' in French 'nucleo' in Italian and 'Kern' in German.

The general equations (1) and (2) are sometimes called *Fredholm equations* of the first and second kinds respectively.

1·2. Connexion with differential equations. The theories of ordinary and partial differential equations are a fruitful source of integral equations. We shall sketch here one of the ways in which integral equations can arise from ordinary differential equations.

We begin by considering the first-order differential equation

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

with the initial condition $y(0) = y_0$. If, say, $f(x, y)$ is a continuous function of (x, y), we can integrate (1) from 0 to x, obtaining

$$y(x) = y_0 + \int_0^x f[t, y(t)]\,dt, \tag{2}$$

an integral equation, in general non-linear, for the function $y(x)$. Conversely, any solution $y(x)$ of (2) clearly satisfies both (1) and the initial condition $y(0) = y_0$. This illustrates the general fact that, by going over to integral equations, we can include both the differential equation and the initial conditions in a single equation.

Let us now consider the second-order differential equation

$$\frac{d^2y}{dx^2} = y'' = f(x, y), \tag{3}$$

with the initial conditions $y(0) = y_0$, $y'(0) = y_1$. We then have

$$y'(x) = y_1 + \int_0^x f[u, y(u)]\,du,$$

whence a second integration gives

$$\begin{aligned} y(x) &= y_0 + y_1 x + \int_0^x dt \int_0^t f[u, y(u)]\,du \\ &= y_0 + y_1 x + \int_0^x f[u, y(u)]\,du \int_u^x dt \\ &= y_0 + y_1 x + \int_0^x (x-u)\, f[u, y(u)]\,du. \qquad (4) \end{aligned}$$

The argument is reversible, so that here again the differential equation (3), together with the initial conditions, is equivalent to the single integral equation (4).

We see also that any solution of (3) satisfies an integral equation of the form

$$y(x) = A + Bx + \int_0^x (x-u)\, f[u, y(u)]\, du, \tag{5}$$

the constants A and B being determined by the initial conditions. They may also be determined in other ways; suppose, for instance, that $y(x)$ is required to satisfy a two-point boundary condition, say $y(0) = \alpha$, $y(l) = \beta$. Substituting in (5), we obtain

$$\alpha = y(0) = A,$$
$$\beta = y(l) = A + Bl + \int_0^l (l-u)\, f[u, y(u)]\, du.$$

Hence $$A = \alpha, \quad B = \frac{\beta - \alpha}{l} - \frac{1}{l}\int_0^l (l-u)\, f[u, y(u)]\, du.$$

The function $y(x)$ must therefore satisfy the integral equation

$$y(x) = \alpha + \frac{\beta-\alpha}{l}x + \int_0^x (x-u) f[u, y(u)]\, du - \frac{x}{l}\int_0^l (l-u) f[u, y(u)]\, du,$$

which can be written in the form

$$y(x) = z(x) - \int_0^l K(x,u)\, f[u, y(u)]\, du, \tag{6}$$

where $$z(x) = \alpha + \frac{\beta-\alpha}{l}x,$$

and $$\left.\begin{aligned} K(x,u) &= \frac{u(l-x)}{l} \quad (0 \leqslant u \leqslant x), \\ &= \frac{x(l-u)}{l} \quad (x \leqslant u \leqslant l). \end{aligned}\right\}$$

The argument is again reversible, so that (6) is equivalent to (3) together with the boundary conditions.

If the differential equation is linear, we are led in this way to a linear integral equation of the second kind. For instance, let us start with the linear differential equation

$$y'' + p(x)\, y = \omega(x),$$

which arises in the theory of vibrating strings. Here

$$f[u, y(u)] = \omega(u) - p(u)\, y(u),$$

so that the corresponding integral equation is

$$y(x) = z(x) - \int_0^l K(x, u)\, [\omega(u) - p(u)\, y(u)]\, du$$

$$= w(x) + \int_0^l K(x, u)\, p(u)\, y(u)\, du, \qquad (7)$$

where
$$w(x) = z(x) - \int_0^l K(x, u)\, \omega(u)\, du,$$

which is a known function. We see at once that (7) is a linear integral equation of the second kind. Specializing still further, let us consider the equation

$$y'' + \lambda p(x)\, y = 0,$$

with the boundary conditions $y(0) = y(l) = 0$; this arises in the problem of finding the normal modes of vibration of a string with fixed end-points. We are, of course, interested in finding solutions other than the trivial one $y(x) = 0$. The corresponding integral equation is

$$y(x) = \lambda \int_0^l K(x, u)\, p(u)\, y(u)\, du,$$

a homogeneous linear equation of the second kind.

We shall see in Chapter VII that symmetric kernels, i.e. kernels $K(s, t)$ such that $K(s, t) = K(t, s)$, have important special properties. The kernel $K(x, u)\, p(u)$ of (7) is not itself symmetric, but it is easily seen from its definition that $K(x, u)$ is symmetric. If the function $p(x)$ is positive, as it usually is, we can use this fact to transform (7) into an equation with a symmetric kernel; we write it in the form

$$\sqrt{[p(x)]}\, y(x) = \sqrt{[p(x)]}\, w(x) + \int_0^l \sqrt{[p(x)]}\, K(x, u) \sqrt{[p(u)]} \sqrt{[p(u)]}\, y(u)\, du,$$

i.e.
$$v(x) = g(x) + \int_0^l L(x, u)\, v(u)\, du,$$

where $v(x) = \sqrt{[p(x)]}\, y(x)$, $g(x)$ is a known function, and

$$L(x, u) = \sqrt{[p(x)]}\, K(x, u) \sqrt{[p(u)]},$$

which is clearly symmetric. Symmetric kernels often arise in very much this kind of way in problems of mathematical physics.

We shall not have sufficient space to discuss in this book the detailed consequences of this connexion between differential and integral equations; for a full account of this, and for a discussion of the applications of integral equations in potential theory, the reader is referred to Lovitt (1924) and Courant & Hilbert (1953).

1·3. Continuous functions and $\mathfrak{L}^2$ functions. We shall consider throughout complex-valued functions $x(t)$ of a real variable t, defined in a finite interval $a \leqslant t \leqslant b$. Two classes of functions will mainly concern us: continuous functions and functions of integrable square. A function $x(t)$ belongs to the latter class if it is measurable in the interval (a, b) and

$$\int_a^b |x(t)|^2 \, dt < \infty,$$

the integral being taken in the sense of Lebesgue. Such a function will also be called an $\mathfrak{L}^2$ function.

Since the interval (a, b) will usually be fixed throughout the discussion, we shall often omit the limits of integration.

The restriction to a finite interval (a, b) of the real line is adopted only for reasons of convenience; in the $\mathfrak{L}^2$ case almost everything we shall have to say will be valid, with inessential alterations, for functions defined on any set on which a Lebesgue measure exists, e.g. on an infinite interval, on a set of positive n-dimensional measure in a Euclidean space R^n of n dimensions, or on the surface of a sphere; in the continuous case it may be necessary to restrict the domain of definition to a compact set of finite measure.

If two $\mathfrak{L}^2$ functions $x(t)$ and $y(t)$ are equal for 'almost all' values of t, i.e. except for a set of values of t of Lebesgue measure zero, we shall say that $x(t)$ and $y(t)$ are *equivalent*, and write

$$x(t) =^\circ y(t).$$

If $x(t) =^\circ 0$, we shall call $x(t)$ a *null function*. We shall also use notations such as $x(t) \leqslant^\circ y(t)$ with the obvious meaning. In certain connexions it is customary to regard equivalent functions as being identical; we shall not adopt this convention, but the notion of equivalence will nevertheless play a substantial rôle in our discussions.

The *norm* $\| x \| = \| x \|_c$ of a continuous function $x(t)$ is defined by the equation

$$\| x \|_c = \sup_{a \leqslant t \leqslant b} | x(t) |,$$

the notation 'sup' (for *supremum*) being used to denote the least upper bound. If $x(t)$ is an $\mathfrak{L}^2$ function, we define its $\mathfrak{L}^2$ norm $\| x \| = \| x \|_2$ by the equation

$$\| x \|_2 = \left\{ \int_a^b | x(t) |^2 \, dt \right\}^{\frac{1}{2}}.$$

We note that $\| x \|_c = 0$ if and only if $x(t)$ vanishes identically; on the other hand, $\| x \|_2 = 0$ if and only if $x(t) \overset{\circ}{=} 0$. It will usually be clear from the context which norm is being used.

1·4. The inequalities of Schwarz and Minkowski. We now prove the two fundamental inequalities of the theory of $\mathfrak{L}^2$ functions. The first of these, usually known as *Schwarz's inequality*, is the analogue for integrals of *Cauchy's inequality*† for sequences, which states that if

$$\sum_{n=1}^{\infty} | a_n |^2 < \infty, \quad \sum_{n=1}^{\infty} | b_n |^2 < \infty,$$

then $\sum_{n=1}^{\infty} a_n b_n$ is absolutely convergent, and

$$\left| \sum_{n=1}^{\infty} a_n b_n \right| \leqslant \left(\sum_{n=1}^{\infty} | a_n |^2 \right)^{\frac{1}{2}} \left(\sum_{n=1}^{\infty} | b_n |^2 \right)^{\frac{1}{2}}. \tag{1}$$

THEOREM 1·4·1. *If $x(t)$ and $y(t)$ are $\mathfrak{L}^2$ functions, then $x(t)\,y(t)$ is integrable, and*

$$\left| \int x(t)\, y(t)\, dt \right| \leqslant \| x \| . \| y \|. \tag{2}$$

The function $x(t)\,y(t)$ is clearly measurable. Since

$$\left| \int x(t)\, y(t)\, dt \right| \leqslant \int | x(t) | . | y(t) | \, dt,$$

† Hardy, Littlewood and Pólya (1934), pp. 16, 115.

the existence of the integral on the right implying the existence of that on the left, it is sufficient to prove the result when $x(t)$ and $y(t)$ are real and non-negative. The integrability of $x(t)\,y(t)$ then follows at once from the elementary inequality

$$x(t)\,y(t) \leqslant \tfrac{1}{2}[x(t)]^2 + \tfrac{1}{2}[y(t)]^2.$$

For real λ and μ, the expression

$$\int [\lambda x(t) + \mu y(t)]^2\,dt = \lambda^2 \int [x(t)]^2\,dt + 2\lambda\mu \int x(t)\,y(t)\,dt + \mu^2 \int [y(t)]^2\,dt$$
$$= \alpha\lambda^2 + 2\beta\lambda\mu + \gamma\mu^2,$$

say, is a non-negative definite quadratic form in (λ, μ), so that we must have $\beta^2 \leqslant \alpha\gamma$, which is just the required inequality.

If $x(t)$ and $y(t)$ are $\mathfrak{L}^2$ functions, we define their *inner product*† (x, y) by the equation

$$(x, y) = \int x(t)\,\overline{y(t)}\,dt,$$

where the bar denotes the complex conjugate, as usual. By the theorem we have just proved, (x, y) always exists, and

$$|(x, y)| \leqslant \|x\| . \|y\|.$$

We also have $\quad (y, x) = \overline{(x, y)}, \quad \|x\|^2 = (x, x).$

If $(x, y) = 0$, we say that x and y are *orthogonal* to one another; since $(x, y) = 0$ implies $(y, x) = 0$, the relation of orthogonality is symmetrical.

Our second theorem is a particular case of *Minkowski's inequality*.‡

THEOREM 1·4·2. *If $x(t)$ and $y(t)$ are $\mathfrak{L}^2$ functions, then $x(t) + y(t)$ is an $\mathfrak{L}^2$ function, and*

$$\|x + y\| \leqslant \|x\| + \|y\|. \tag{3}$$

It is again sufficient to prove the result when $x(t)$ and $y(t)$ are real and non-negative. If we square both sides of (3), we obtain in this case

$$\|x\|^2 + 2\int x(t)\,y(t)\,dt + \|y\|^2 \leqslant \|x\|^2 + 2\|x\| . \|y\| + \|y\|^2. \tag{4}$$

† The name 'scalar product' is also used, but tends to be confused with the product of a function by a scalar.

‡ For the general result, see Burkill (1951), p. 66.

Since (4) is an immediate consequence of (2), the result follows at once.

Theorems 1·4·1 and 1·4·2 hold without alteration for functions of more than one variable, the integration being taken over some domain in Euclidean space of the appropriate number of dimensions (and even more generally, in fact); we shall frequently require the two-dimensional case.

For continuous functions, the results corresponding to (2) and (3) are the trivial inequalities

$$\left|\int x(t)\,y(t)\,dt\right| \leqslant (b-a)\,\|x\|_c\,\|y\|_c, \quad \|x+y\|_c \leqslant \|x\|_c + \|y\|_c.$$

It follows from Theorem 1·4·2 that if $x(t)$ and $y(t)$ are $\mathfrak{L}^2$ functions, any linear combination $\lambda x(t)+\mu y(t)$ with constant coefficients is also an $\mathfrak{L}^2$ function. The set of all $\mathfrak{L}^2$ functions therefore forms a complex vector space; the same is clearly true of the set of all continuous functions.

The inner product (x, y) is linear in the first factor and 'anti-linear' in the second, i.e.

$$(\lambda_1 x_1 + \lambda_2 x_2, y) = \lambda_1(x_1, y) + \lambda_2(x_2, y),$$
$$(x, \lambda_1 y_1 + \lambda_2 y_2) = \bar{\lambda}_1(x, y_1) + \bar{\lambda}_2(x, y_2).$$

1·5. Continuous kernels. Let $K(s,t)$ be a continuous function of (s,t) in the square Δ defined by $a \leqslant s \leqslant b$, $a \leqslant t \leqslant b$. If $x(t)$ is a continuous function of t, the function

$$y(s) = \int_a^b K(s,t)\,x(t)\,dt \tag{1}$$

is continuous in $a \leqslant s \leqslant b$. For, given a positive number ϵ, there is a positive number δ, independent of t, such that

$$|K(s,t) - K(s',t)| < \epsilon \quad (|s-s'| < \delta),$$

and, for some $M > 0$, $|x(t)| \leqslant M$ for all t: hence

$$|y(s) - y(s')| \leqslant \int_a^b |K(s,t) - K(s',t)|\,.\,|x(t)|\,dt$$
$$\leqslant \epsilon M(b-a) \quad (|s-s'| < \delta).$$

Equation (1) may be regarded as defining an operator that takes an arbitrary continuous function $x(t)$ into a continuous function $y(s)$, and we shall often write it in the abbreviated form

$$y = Kx. \tag{2}$$

The function $K(s,t)$ is called a *continuous kernel*. If we write

$$\| K \| = \| K \|_c = (b-a) \sup_{(s,t)\in\Delta} | K(s,t) |,$$

we have

$$\| y \|_c \leqslant \| K \|_c \| x \|_c.$$

The operator K is clearly *linear*, in the sense that

$$K(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 K x_1 + \lambda_2 K x_2$$

for arbitrary constants λ_1 and λ_2.

If $H(r,s)$ is also a continuous kernel, defined for $a \leqslant r \leqslant b$, $a \leqslant s \leqslant b$, we can form

$$z = Hy = H(Kx).$$

Then

$$z(r) = \int H(r,s)\,ds \int K(s,t)\,x(t)\,dt$$

$$= \int x(t)\,dt \int H(r,s)\,K(s,t)\,ds = \int L(r,t)\,x(t)\,dt,$$

where

$$L(r,t) = \int H(r,s)\,K(s,t)\,ds,$$

which is clearly a continuous function of (r,t). Thus $H(Kx) = Lx$, where L is a continuous kernel. We write $L = HK$, so that we have

$$(HK)\,x = H(Kx);$$

we may also write $L(s,t) = HK(s,t)$. It is easily seen that this 'operator multiplication' of kernels is associative, i.e.

$$G(HK) = (GH)\,K,$$

and doubly distributive with respect to addition, i.e.,

$$G(H+K) = GH + GK, \quad (G+H)\,K = GK + HK.$$

It is not commutative in general, for we do not as a rule have $HK = KH$. Since

$$| HK(r,t) | \leqslant \int | H(r,s) | . | K(s,t) | \,ds \leqslant \frac{\| H \|_c \| K \|_c}{b-a},$$

we have

$$\| HK \|_c \leqslant \| H \|_c \| K \|_c. \tag{3}$$