1

A brief discussion of integral equations

The theory and application of integral equations is an important subject within applied mathematics. Integral equations are used as mathematical models for many and varied physical situations, and integral equations also occur as reformulations of other mathematical problems. We begin with a brief classification of integral equations, and then in later sections, we give some of the classical theory for one of the most popular types of integral equations, those that are called *Fredholm integral equations of the second kind*, which are the principal subject of this book. There are many well-written texts on the theory and application of integral equations, and we note particularly those of Hackbusch [249] Hochstadt [272], Kress [325], Mikhlin [380], Pogorzelski [426], Schmeidler [492], Widom [568], and Zabreyko, et al. [586].

1.1. Types of integral equations

This book is concerned primarily with the numerical solution of what are called Fredholm integral equations, but we begin by discussing the broader category of integral equations in general. In classifying integral equations, we say, very roughly, that those integral equations in which the integration domain varies with the independent variable in the equation are *Volterra integral equations*; and those in which the integration domain is fixed are Fredholm integral equations. We first consider these two types of equations, and the section concludes with some other important integral equations.

1.1.1. Volterra integral equations of the second kind

The general form that is studied is

$$x(t) + \int_a^t K(t, s, x(s))\, ds = y(t), \quad t \geq a \tag{1.1.1}$$

The functions $K(t, s, u)$ and $y(t)$ are given, and $x(t)$ is sought. This is a *nonlinear* integral equation, and it is in this form that the equation is most commonly applied and solved. Such equations can be thought of as generalizations of

$$x'(t) = f(t, x(t)), \quad t \geq a, \quad x(a) = x_0 \tag{1.1.2}$$

the *initial value problem for ordinary differential equations*. This equation is equivalent to the integral equation

$$x(t) = x_0 + \int_a^t f(s, x(s))\, ds, \quad t \geq a$$

which is a special case of (1.1.1).

For an introduction to the theory of Volterra integral equations, see R. Miller [384]. The numerical methods for solving (1.1.1) are closely related to those for solving the initial value problem (1.1.2). These integral equations are not studied in this book, and the reader is referred to Brunner and de Riele [96] and Linz [345]. Volterra integral equations are most commonly studied for functions x of one variable, as above, but there are examples of Volterra integral equations for functions of more than one variable.

1.1.2. Volterra integral equations of the first kind

The general nonlinear Volterra integral equation of the first kind has the form

$$\int_a^t K(t, s, x(s))\, ds = y(t), \quad t \geq a \tag{1.1.3}$$

The functions $K(t, s, u)$ and $y(t)$ are given functions, and the unknown is $x(s)$. The general linear Volterra integral equation of the first kind is of the form

$$\int_a^t K(t, s)x(s)\, ds = y(t), \quad t \geq a \tag{1.1.4}$$

For Volterra equations of the first kind, the linear equation is the more commonly studied case. The difficulty with these equations, linear or nonlinear, is that they are "ill-conditioned" to some extent, and that makes their numerical solution more difficult. (Loosely speaking, an *ill-conditioned problem* is one in which small changes in the data y can lead to much larger changes in the solution x.)

A very simple but important example of (1.1.4) is

$$\int_a^t x(s)\, ds = y(t), \quad t \geq a \tag{1.1.5}$$

This is equivalent to $y(a) = 0$ and $x(t) = y'(t)$, $t \geq a$. Thus the numerical solution of (1.1.5) is equivalent to the numerical differentiation of $y(t)$. For a discussion of the numerical differentiation problem from this perspective, see Cullum [149] and Anderssen and Bloomfield [11], and for the numerical solution of the more general equation (1.1.4), see Linz [345].

1.1.3. Abel integral equations of the first kind

An important case of (1.1.4) is the *Abel integral equation*

$$\int_0^t \frac{H(t,s)x(s)}{(t^p - s^p)^\alpha}\, ds = y(t), \quad t > 0 \tag{1.1.6}$$

Here $0 < \alpha < 1$ and $p > 0$, and particularly important cases are $p = 1$ and $p = 2$ (both with $\alpha = \frac{1}{2}$). The function $H(t,s)$ is assumed to be smooth (that is, several times continuously differentiable). Special numerical methods have been developed for these equations, as they occur in a wide variety of applications. For a general solvability theory for (1.1.6), see Ref. [35], and for a discussion of numerical methods for their solution, see Linz [345], Brunner and de Riele [96], and Anderssen and de Hoog [12].

1.1.4. Fredholm integral equations of the second kind

The general form of such an integral equation is

$$\lambda x(t) - \int_D K(t,s)x(s)\, ds = y(t), \quad t \in D, \quad \lambda \neq 0 \tag{1.1.7}$$

with D a closed bounded set in $\mathbf{R}^m$, some $m \geq 1$. The *kernel function* $K(t,s)$ is assumed to be absolutely integrable, and it is assumed to satisfy other properties that are sufficient to imply the Fredholm Alternative Theorem (see Theorem 1.3.1 in §1.3). For $y \neq 0$, we have λ and y given, and we seek x; this is the *nonhomogeneous problem*. For $y = 0$, equation (1.1.7) becomes an *eigenvalue problem*, and we seek both the *eigenvalue* λ and the *eigenfunction* x. The principal focus of the numerical methods presented in the following chapters is the numerical solution of (1.1.7) with $y \neq 0$. In the next two sections we present some theory for the integral operator in (1.1.7).

1.1.5. Fredholm integral equations of the first kind

These equations take the form

$$\int_D K(t,s)x(s)\, ds = y(t), \quad t \in D \tag{1.1.8}$$

with the assumptions on K and D the same as in (1.1.7). Such equations are usually classified as ill-conditioned, because their solution x is sensitive to small changes in the data function y. For practical purposes, however, these problems need to be subdivided into two categories. First, if $K(t, s)$ is a smooth function, then the solution $x(s)$ of (1.1.8) is extremely sensitive to small changes in $y(t)$, and special methods of solution are needed. For excellent introductions to this topic, see Groetsch [241], [242], Kress [325, Chaps. 15–17] and Wing [572]. If however, $K(t, s)$ is a *singular* function, then the ill-conditioning of (1.1.8) is quite manageable; and indeed, much of the theory for such equations is quite similar to that for the second-kind equation (1.1.7). Examples of this type of first-kind equation occur quite frequently in the subject of *potential theory*, and a well-studied example is

$$\int_\Gamma \log|t - s| x(s)\, ds = y(t), \quad t \in \Gamma \tag{1.1.9}$$

with Γ a curve in $\mathbf{R}^2$. This and other similarly behaved first-kind equations will be discusssed in Chapters 7 and 8.

1.1.6. Boundary integral equations

These equations are integral equation reformulations of partial differential equations. They are widely studied and applied in connection with solving boundary value problems for elliptic partial differential equations, but they are also used in connection with other types of partial differential equations.

As an example, consider solving the problem

$$\Delta u(P) = 0, \quad P \in D \tag{1.1.10}$$

$$u(P) = g(P), \quad P \in \Gamma \tag{1.1.11}$$

where D is a bounded region in $\mathbf{R}^3$ with nonempty interior, and Γ is the boundary of D. From the physical setting for (1.1.10)–(1.1.11), there is reason to believe that u can be written as a *single layer potential:*

$$u(P) = \int_\Gamma \frac{\rho(Q)}{|P - Q|} dQ, \quad P \in D \tag{1.1.12}$$

In this, $|P - Q|$ denotes the ordinary Euclidean distance between P and Q. The function $\rho(Q)$ is called a *single layer density function*, and it is the unknown in the equation. Using the boundary condition (1.1.11), it is straightforward to show that

$$\int_\Gamma \frac{\rho(Q)}{|P - Q|} dQ = g(P), \quad P \in \Gamma \tag{1.1.13}$$

This equation is solved for ρ, and then (1.1.12) is used to obtain the solution of (1.1.10)–(1.1.11).

Boundary integral equations can be Fredholm integral equations of the first or second kind, Cauchy singular integral equations (see 1.1.8 below), or modifications of them. In the literature, boundary integral equations are often referred to as *BIE*, and methods for solving partial differential equations via the boundary integral equation reformulation are called *BIE methods*. There are many books and papers written on BIE methods; for example, see Refs. [50], [55], Jaswon and Symm [286], Kress [325, Chaps. 6, 8, 9], Sloan [509] and the references contained therein.

1.1.7. Wiener-Hopf integral equations

These have the form

$$\lambda x(t) - \int_0^\infty k(t - s)x(s)\,ds = y(t), \quad 0 \le t < \infty \tag{1.1.14}$$

Originally, such equations were studied in connection with problems in radiative transfer, and more recently, they have been related to the solution of boundary integral equations for planar problems in which the boundary is only piecewise smooth. A very extensive theory for such equations is given in Krein [322], and more recently, a simpler introduction to some of the more important parts of the theory has been given by Anselone and Sloan [20] and deHoog and Sloan [163].

1.1.8. Cauchy singular integral equations

Let Γ be an open or closed contour in the complex plane. The general form of a Cauchy singular integral equation is given by

$$a(z)\phi(z) + \frac{b(z)}{\pi i}\int_\Gamma \frac{\phi(\zeta)}{\zeta - z} d\zeta + \int_\Gamma K(z,\zeta)\phi(\zeta)\,d\zeta = \psi(z), \quad z \in \Gamma \tag{1.1.15}$$

The functions a, b, ψ, and K are given complex-valued functions, and ϕ is the unknown function. The function K is to be absolutely integrable; and in addition, it is to be such that the associated integral operator is a Fredholm integral operator in the sense of 1.1.4 above. The first integral in (1.1.15) is interpreted as a *Cauchy principal value integral*:

$$\int_\Gamma \frac{\phi(\zeta)}{\zeta - z}\, d\zeta = \lim_{\epsilon \to 0^+} \int_{\Gamma_\epsilon} \frac{\phi(\zeta)}{\zeta - z}\, d\zeta \tag{1.1.16}$$

with $\Gamma_\epsilon = \{\zeta \in \Gamma \mid |\zeta - z| \geq \epsilon\}$. Cauchy singular integral equations occur in a variety of physical problems, especially in connection with the solution of partial differential equations in $\mathbf{R}^2$. Among the best known books on the theory and application of Cauchy singular integral equations are Muskhelishvili [390] and Gakhov [208]; and an important more recent work is that of Mikhlin and Prößdorf [381]. For an introduction to the numerical solution of Cauchy singular integral equations, see Elliott [177], [178] and Prößdorf and Silbermann [438].

1.2. Compact integral operators

The framework we present in this and the following sections is fairly abstract, and it might seem far removed from the numerical solution of actual integral equations. In fact, our framework is needed to understand the behavior of most numerical methods for solving Fredholm integral equations, including the answering of questions regarding convergence, numerical stability, and asymptotic error estimates. The language of functional analysis has become more standard in the past few decades, and so in contrast to our earlier book [39, Part I], we do not develop here any results from functional analysis, but rather state them in the appendix and refer the reader to other sources for proofs.

When $\mathcal{X}$ is a finite dimensional vector space and $\mathcal{A}: \mathcal{X} \to \mathcal{X}$ is linear, the equation $\mathcal{A}x = y$ has a well-developed solvability theory. To extend these results to infinite dimensional spaces, we introduce the concept of a *compact operator* $\mathcal{K}$; and then in the following section, we give a theory for operator equations $\mathcal{A}x = y$ in which $\mathcal{A} = I - \mathcal{K}$.

Definition. *Let $\mathcal{X}$ and $\mathcal{Y}$ be normed vector spaces, and let $\mathcal{K}: \mathcal{X} \to \mathcal{Y}$ be linear. Then $\mathcal{K}$ is* compact *if the set*

$$\{\mathcal{K}x \mid \|x\|_{\mathcal{X}} \leq 1\}$$

has compact closure in $\mathcal{Y}$. This is equivalent to saying that for every bounded sequence $\{x_n\} \subset \mathcal{X}$, the sequence $\{\mathcal{K}x_n\}$ has a subsequence that is convergent to some point in $\mathcal{Y}$. Compact operators are also called completely continuous *operators. (By a set S having* compact closure *in $\mathcal{Y}$, we mean its closure $\bar{S}$ is a compact set in $\mathcal{Y}$.)*

There are other definitions for a compact operator, but the above is the one used most commonly. In the definition, the spaces $\mathcal{X}$ and $\mathcal{Y}$ need not be complete; but in virtually all applications, they are complete. With completeness, some of the proofs of the properties of compact operators become simpler, and we will always assume $\mathcal{X}$ and $\mathcal{Y}$ are complete (that is, Banach spaces) when dealing with compact operators.

1.2.1. Compact integral operators on C(D)

Let D be a closed bounded set in $\mathbf{R}^m$, some $m \geq 1$, and define

$$\mathcal{K}x(t) = \int_D K(t,s)x(s)\,ds, \quad t \in D, \quad x \in C(D) \tag{1.2.17}$$

Using $C(D)$ with $\|\cdot\|_\infty$, we want to show that $\mathcal{K}: C(D) \to C(D)$ is both bounded and compact. We assume $K(t,s)$ is Riemann-integrable as a function of s, for all $t \in D$, and further we assume the following.

K1. $\lim_{h\to 0} \omega(h) = 0$, with

$$\omega(h) \equiv \max_{t,\tau \in D} \max_{|t-\tau|\leq h} \int_D |K(t,s) - K(\tau,s)|\,ds \tag{1.2.18}$$

K2.

$$\max_{t\in D} \int_D |K(t,s)|\,ds < \infty \tag{1.2.19}$$

Using K1, if $x(s)$ is bounded and integrable, then $\mathcal{K}x(t)$ is continuous, with

$$|\mathcal{K}x(t) - \mathcal{K}x(\tau)| \leq \omega(|t-\tau|)\|x\|_\infty \tag{1.2.20}$$

Using K2, we have boundedness of $\mathcal{K}$, with

$$\|\mathcal{K}\| = \max_{t\in D} \int_D |K(t,s)|\,ds \tag{1.2.21}$$

To discuss compactness of $\mathcal{K}$, we first need to identify the compact sets in $C(D)$. To do this, we use the Arzela-Ascoli theorem from advanced calculus. It states that a subset $S \subset C(D)$ has compact closure if and only if (1) S is a uniformly bounded set of functions, and (2) S is an equicontinuous family. Now consider the set $S = \{\mathcal{K}x \mid x \in C(D) \text{ and } \|x\|_\infty \leq 1\}$. This is uniformly bounded, since $\|\mathcal{K}x\|_\infty \leq \|\mathcal{K}\|\|x\|_\infty \leq \|\mathcal{K}\|$. In addition, S is equicontinuous from (1.2.20). Thus S has compact closure in $C(D)$, and $\mathcal{K}$ is a compact operator on $C(D)$ to $C(D)$.

What are the kernel functions K that satisfy K1–K2? Easily, these assumptions are satisfied if $K(t, s)$ is a continuous function of $(t, s) \in D$. In addition, let $D = [a, b]$ and consider

$$\mathcal{K}x(t) = \int_a^b \log|t - s|\, x(s)\, ds \tag{1.2.22}$$

and

$$\mathcal{K}x(t) = \int_a^b \frac{1}{|s - t|^\beta} x(s)\, ds \tag{1.2.23}$$

with $\beta < 1$. These definitions of $\mathcal{K}$ can be shown to satisfy K1–K2, although we will not do so here. Later we will show by other means that these are compact operators.

Another way to show that $K(t, s)$ satisfies K1 and K2 is to rewrite K in the form

$$K(t, s) = \sum_{i=0}^{p} H_i(t, s) L_i(t, s) \tag{1.2.24}$$

for some $p > 0$, with each $L_i(t, s)$ continuous for $a \le t, s \le b$ and each $H_i(t, s)$ satisfying K1–K2. It is left to the reader to show that in this case, K also satisfies K1–K2. The utility of this approach is that it is sometimes difficult to show directly that K satisfies K1–K2, whereas showing (1.2.24) may be easier.

Example. Let $[a, b] = [0, \pi]$ and $K(t, s) = \log|\cos t - \cos s|$. Then rewrite this as

$$K(t, s) = \underbrace{|s - t|^{-\frac{1}{2}}}_{H(t,s)} \underbrace{|s - t|^{\frac{1}{2}} \log|\cos t - \cos s|}_{L(t,s)} \tag{1.2.25}$$

Easily, L is continuous; and from the discussion following (1.2.23), H satisfies K1–K2. Thus K is the kernel of a compact integral operator on $C[0, \pi]$ to $C[0, \pi]$.

1.2.2. Properties of compact operators

Another way of obtaining compact operators is to look at limits of simpler "finite-dimensional operators" in $L[\mathcal{X}, \mathcal{Y}]$, the Banach space of bounded linear operators from $\mathcal{X}$ to $\mathcal{Y}$. This gives another perspective on compact operators, one that leads to improved intuition by emphasizing their relationship to operators on finite dimensional spaces.

Definition. *Let $\mathcal{X}$ and $\mathcal{Y}$ be vector spaces. The linear operator $\mathcal{K} : \mathcal{X} \to \mathcal{Y}$ is a* finite rank operator *if $Range(\mathcal{K})$ is finite dimensional.*

Lemma 1.2.1. Let $\mathcal{X}$ and $\mathcal{Y}$ be normed linear spaces, and let $\mathcal{K} : \mathcal{X} \to \mathcal{Y}$ be a bounded finite rank operator. Then $\mathcal{K}$ is a compact operator.

Proof. Let $\mathcal{R} = Range(\mathcal{K})$. Then $\mathcal{R}$ is a normed finite-dimensional space, and therefore it is complete. Consider the set

$$S = \{\mathcal{K}x \mid \|x\| \le 1\}$$

The set S is bounded by $\|\mathcal{K}\|$. Also $S \subset \mathcal{R}$. Then S has compact closure, since all bounded closed sets in a finite dimensional space are compact. This shows $\mathcal{K}$ is compact. □

Example. Let $\mathcal{X} = \mathcal{Y} = C[a, b]$ with $\|\cdot\|_\infty$. Consider the kernel function

$$K(t, s) = \sum_{i=1}^{n} \beta_i(t)\gamma_i(s) \tag{1.2.26}$$

with each β_i continuous on $[a, b]$ and each $\gamma_i(s)$ absolutely integrable on $[a, b]$. Then the associated integral operator $\mathcal{K}$ is a bounded, finite rank operator on $C[a, b]$ to $C[a, b]$:

$$\mathcal{K}x(t) = \sum_{i=1}^{n} \beta_i(t) \int_a^b \gamma_i(s)x(s)\,ds, \qquad x \in C[a, b] \tag{1.2.27}$$

$$\|\mathcal{K}\| \le \sum_{i=1}^{n} \|\beta_i\|_\infty \int_a^b |\gamma_i(s)|\,ds$$

From (1.2.27), $\mathcal{K}x \in C[a, b]$ and $Range(\mathcal{K}) \subset Span\{\beta_1, \ldots, \beta_n\}$, a finite dimensional space. Kernel functions of the form (1.2.26) are called *degenerate*.

In the following section we will see that the associated integral equation $(\lambda - \mathcal{K})x = y$, $\lambda \ne 0$ is essentially a finite dimensional equation.

Lemma 1.2.2. Let $\mathcal{K} \in L[\mathcal{X}, \mathcal{Y}]$ and $\mathcal{L} \in L[\mathcal{Y}, \mathcal{Z}]$, and let $\mathcal{K}$ or $\mathcal{L}$ (or both) be compact. Then $\mathcal{L}\mathcal{K}$ is compact on $\mathcal{X}$ to $\mathcal{Z}$.

Proof. This is left as an exercise for the reader. □

The following lemma gives the framework for using finite rank operators to obtain similar, but more general compact operators.

Lemma 1.2.3. Let $\mathcal{X}$ and $\mathcal{Y}$ be normed linear spaces, with $\mathcal{Y}$ complete. Let $\mathcal{K} \in L[\mathcal{X}, \mathcal{Y}]$, let $\{\mathcal{K}_n\}$ be a sequence of compact operators in $L[\mathcal{X}, \mathcal{Y}]$, and assume $\mathcal{K}_n \to \mathcal{K}$ in $L[\mathcal{X},\mathcal{Y}]$, i.e., $\|\mathcal{K}_n - \mathcal{K}\| \to 0$. Then $\mathcal{K}$ is compact.

Proof. Let $\{x_n\}$ be a sequence in $\mathcal{X}$ satisfying $\|x_n\| \le 1, n \ge 1$. We must show that $\{\mathcal{K}x_n\}$ contains a convergent subsequence.

Since $\mathcal{K}_1$ is compact, the sequence $\{\mathcal{K}_1 x_n\}$ contains a convergent subsequence. Denote the convergent subsequence by $\{\mathcal{K}_1 x_n^{(1)} \mid n \ge 1\}$, and let its limit be denoted by $y_1 \in \mathcal{Y}$. For $k \ge 2$, inductively pick a subsequence $\{x_n^{(k)} \mid n \ge 1\} \subset \{x_n^{(k-1)}\}$ such that $\{\mathcal{K}_k x_n^{(k)}\}$ converges to a point $y_k \in \mathcal{Y}$. Thus,

$$\lim_{n\to\infty} \mathcal{K}_k x_n^{(k)} = y_k \quad \text{and} \quad \left\{x_n^{(k)}\right\} \subset \left\{x_n^{(k-1)}\right\}, \quad k \ge 1 \tag{1.2.28}$$

We will now choose a special subsequence $\{z_k\} \subset \{x_n\}$ for which $\{\mathcal{K}z_k\}$ is convergent in $\mathcal{Y}$. Let $z_1 = x_j^{(1)}$ for some j, such that $\|\mathcal{K}_1 x_n^{(1)} - y_1\| \le 1$ for all $n \ge j$. Inductively, for $k \ge 2$, pick $z_k = x_j^{(k)}$ for some j, such that z_k is further along in the sequence $\{x_n\}$ than is z_{k-1} and such that

$$\left\|\mathcal{K}_k x_n^{(k)} - y_k\right\| \le \frac{1}{k}, \quad n \ge j \tag{1.2.29}$$

The sequence $\{\mathcal{K}z_k\}$ is a Cauchy sequence in $\mathcal{Y}$. To show this, consider

$$\begin{aligned} \|\mathcal{K}z_{k+p} - \mathcal{K}z_k\| &\le \|\mathcal{K}z_{k+p} - \mathcal{K}_k z_{k+p}\| + \|\mathcal{K}_k z_{k+p} - \mathcal{K}_k z_k\| \\ &\quad + \|\mathcal{K}_k z_k - \mathcal{K}z_k\| \\ &\le 2\|\mathcal{K} - \mathcal{K}_k\| + \|\mathcal{K}_k z_{k+p} - y_k\| + \|y_k - \mathcal{K}_k z_k\| \\ &\le 2\|\mathcal{K} - \mathcal{K}_k\| + \frac{2}{k}, \quad p \ge 1 \end{aligned}$$

The last statement uses (1.2.28)–(1.2.29), noting that $z_{k+p} \in \{x_n^{(k)}\}$ for all $p \ge 1$. Use the assumption that $\|\mathcal{K} - \mathcal{K}_k\| \to 0$ to conclude the proof that $\{\mathcal{K}z_k\}$ is a Cauchy sequence in $\mathcal{Y}$. Since $\mathcal{Y}$ is complete, $\{\mathcal{K}z_k\}$ is convergent in $\mathcal{Y}$, and this shows that $\mathcal{K}$ is compact. □

For almost all function spaces $\mathcal{X}$ of interest, the compact operators can be characterized as being the limit of a sequence of bounded finite-rank operators. This gives a further justification for the presentation of Lemma 1.2.3.

Example. Let D be a closed and bounded set in $\mathbf{R}^m$, some $m \ge 1$. For example, D could be a region with nonempty interior, a piecewise smooth surface, or a piecewise smooth curve. Let $K(t, s)$ be a continuous function of $t, s \in D$. Then