Integral equations

Many problems of physical interest—for example in the quantum mechanical theory of scattering—are most simply expressed as integral equations. We restrict our discussion to *linear* integral equations, although many interesting equations—especially those arising in Hartree-Fock approximations of many-body systems, or in dispersion relations—are nonlinear and require more powerful mathematical techniques.

A linear integral equation has the form

$$f(x) = g(x) + \lambda \int_a^b dy\, K(x, y)\, f(y) \quad .$$

The function $K(x\, y)$ is called the *kernel* of the equation. There are several easy cases which we now examine:

1. Volterra equations

Integral equations whose upper limit is x are known as Volterra equations. There are two possibilities:

If the the upper limit is x and the kernel does not depend on x, the equation is equivalent to an ordinary differential equation:

$$f(x) = g(x) + \lambda \int_a^x dy\, K(y)\, f(y) \;\Rightarrow\; f'(x) = g'(x) + \lambda\, K(x)\, f(x)$$

which can be reduced to quadratures[1] by standard methods.

If the upper limit is x and the kernel depends on x as well as y, it may still be possible to represent the integral equation as an ordinary differential equation, by differentiating several times with respect to x. In this case the differential equation will be of higher order than unity, and will not in general be reducible to quadratures. Such cases can be solved numerically using the methods of Chapter 5. An example is

$$f(x) = x^2 + \lambda \int_0^x dy\, (1 + xy)\, f(y)$$

for which the corresponding differential equation is

$$f'' = 2 + \lambda\left(1 + x^2\right) f + 3\lambda\, x\, f\,.$$

1. The phrase "reduced to quadratures" means simply that the answer can be expressed as a definite integral of a known function. At worst we evaluate this integral numerically.

If the kernel depends on x in a manner that cannot be dissected out into a differential equation as above, the Volterra equation can nevertheless be solved straightforwardly. It turns out to be straightforward to show that the iterative solution converges for any value of λ. To see this, consider the n'th term of the iterated integral:

$$I_n = \lambda^n \int_a^x d\xi_1 \int_a^{\xi_1} d\xi_2 \ldots \int_a^{\xi_{n-1}} d\xi_n \, K(x, \xi_1)\, K(\xi_1, \xi_2) \ldots K(\xi_{n-1}, \xi_n)\, g(\xi_n) \quad ;$$

assuming the functions $K(x, y)$ and $g(x)$ are bounded,

$$|K(x, y)| \underset{a \le y \le x}{\le} M \,, \qquad |g(x)| \le m \,,$$

we can obtain an upper bound on the integral:

$$|I_n| \le |\lambda|^n M^n m \frac{(x-a)^n}{n!}$$

so the difference between the solution $f(x)$ and the partial sum of the iterated solution is bounded by

$$\left| f(x) - \sum_{k=0}^{n-1} I_k(x) \right| \le |\lambda|^n M^n m \frac{(x-a)^n}{n!} \exp\Big(|\lambda| M (x-a) \Big) \underset{n \to \infty}{\to} 0 \,.$$

That is, the iterated solution definitely converges. If the lower limit of the integral is —say—negative infinity, or if the kernel is not a bounded function, the iterated solution may nevertheless converge, if the kernel is square-integrable. In any event we shall not be interested in Volterra equations for which the kernel is not well-behaved.

2. Equations with a difference kernel

If the integral equation has the form

$$f(x) = g(x) + \lambda \int_a^b dy \, K(x-y)\, f(y)$$

we say it has a *difference* kernel. There are only two cases of interest here:

- the interval of integration is $(a, b) = (-\infty, \infty)$ or $(0, \infty)$;
- the interval is $(a, b) = (0, x)$.

In the first case we can Fourier transform to get

$$\widetilde{f}(k) = \widetilde{g}(k) + \lambda \widetilde{K}(k)\, \widetilde{f}(k)$$

which is solved by simple algebra to obtain the Fourier transform of $f(x)$; from this we can reconstruct the function by taking the inverse Fourier transform, *i.e.* by a quadrature. The only thing we must watch out for is when $\lambda \widetilde{K}(k) = 1$ for values of k in the interval of integration. These correspond to eigenvalues of the equation (which are continuously distributed).

The second case is also straightforward: if

$$f(x) = g(x) + \lambda \int_0^x dy\, K(x-y)\, f(y)$$

we can Laplace transform both sides. Since the Laplace transform of a convolution integral like that above is just the product of the Laplace transforms of the factors,

$$\int_0^\infty dx\, e^{-px} \int_0^x dy\, K(x-y)\, f(y) \equiv \int_0^\infty dx\, e^{-px} K(x) \int_0^\infty dy\, e^{-py} f(y)$$

we can find the Laplace transform of $f(x)$ by algebra and take its inverse to solve the problem. Of course except in certain especially simple cases[2], the inverse Laplace transform must be evaluated by means of a contour integral—but nevertheless the problem has been reduced to numerical quadrature—solvable in at most $O\left(N^2\right)$time[3].

3. Linear equations of Fredholm type

The Fredholm integral equation[4] has the form

$$f(x) = g(x) + \lambda \int_a^b dy\, K(x,y)\, f(y)$$

where a and b are definite limits (that is, not dependent on x). It is useful to write this equation more abstractly, as an operator equation:

$$f = g + \lambda K f\,.$$

Clearly the formal solution is

$$f = (1 - \lambda K)^{-1} g$$

but this is meaningless unless we can define the *resolvent* operator

$$R(\lambda) = (1 - \lambda K)^{-1}.$$

Neumann series

The branch of mathematics called *functional analysis* deals with various methods for defining the resolvent—when it is possible. The first method that comes to mind is to iterate the original equation:

---

2. That is, cases where the Laplace transform can be found in a standard table.
3. If we want to know $f(x)$ on a mesh of N points we must evaluate the integral N times; but each evaluation may require a time proportional to N, hence the total grows as N^2. In certain cases we can approximate the inverse Laplace transform by the Fast Fourier algorithm, in which case the time is of order $N\log N$.
4. See, *e.g.*, F. Smithies, *Integral Equations* (Cambridge University Press, Cambridge, 1958).

$$f_0(x) = g(x)$$

$$f_n(x) = g(x) + \lambda \int_a^b dy\, K(x, y)\, f_{n-1}(y)\,, \quad n=1, 2, \ldots$$

This is equivalent to expanding the resolvent in formal geometric series,

$$R(\lambda) = (1 - \lambda K)^{-1} = 1 + \lambda K + \lambda^2 K^2 + \lambda^3 K^3 + \ldots$$

where by K^2 we mean the convolution

$$K^2(x, y) = \int_a^b d\xi\, K(x, \xi)\, K(\xi, y) \quad .$$

This *Neumann expansion* only works if there is a bounding criterion on the kernel K such that

$$||K|| \leq M < \infty\,;$$

in that case the series converges within the circle (in the complex λ–plane)

$$|\lambda| < \frac{1}{M}\,.$$

Kernels that satisfy the above criterion are called, naturally enough, *bounded*. Of course not all linear operators on a vector space are bounded—the (positive) operator $-\nabla^2$ is unbounded above.

One criterion for boundedness is that the integral

$$||K||^2_{Schmidt} = \int_a^b \int_a^b dy\; |K(x, y)|\; |K(y, x)|$$

be finite. This is sometimes called the *Schmidt norm* of the kernel. A somewhat more useful one is that

$$||K|| \leq \left[\sup_x \frac{1}{\sigma(x)} \int_a^b dy\; |K(x, y)|\; \sigma(y)\right]^{1/2} \left[\sup_y \frac{1}{\sigma(y)} \int_a^b dx\; |K(x, y)|\; \sigma(x)\right]^{1/2}$$

where $\sigma(x)$ is any positive function[5].

If we apply this to the kernel $K(x, y) = |x - y|^{-1/2}$ on the interval $[0, 1]$ (for which the Schmidt norm is clearly infinite), with the choice $\sigma = 1$ (certainly a positive function!) we obtain

$$||K|| \leq \max_{x \in [0, 1]} \int_0^1 \frac{dy}{\sqrt{|x - y|}} = 2\sqrt{2} < \infty\,.$$

5. For example, σ could contain parameters we could vary to reduce the upper bound.

Equations with compact kernels

Kernels with a finite Schmidt norm belong to a wider class known as *compact* kernels. The mathematical definition of compactness is slightly subtle; but one useful criterion for compactness is that a kernel is compact if any power of it is compact.

Integral equations with compact kernels are fundamental in quantum mechanical collision theory—the great progress of the past several decades in treating the quantum mechanical three body problem, for example, derives from new methods of solving such equations numerically.

Here we present three techniques for solving integral equations with compact kernels. The first is the traditional Fredholm method, which is rarely used in practice because it involves complicated multi-dimensional numerical quadrature. The solution to the equation

$$f(x) = g(x) + \lambda \int_a^b dy\, K(x, y)\, f(y)$$

can be written

$$f(x) = g(x) + \lambda \int_a^b dy\, \frac{N(x, y; \lambda)}{D(\lambda)}\, g(y)$$

where both $N(x, y; \lambda)$ and $D(\lambda)$ may be represented as power series in λ, with infinite radii of convergence. Thus the Fredholm theory represents an existence proof of the solution.

An important fact about the solutions of such integral equations is that either the inhomogeneous equation has a solution, or else[6] the homogeneous equation has a solution, but not both at the same time, except in the unlikely instance that at an eigenvalue,

$$\int_a^b dy\, N(x, y; \lambda)\, g(y) = 0 .$$

In order for the homogeneous equation to have a solution, the function $D(\lambda)$—called the *Fredholm determinant*—must vanish. That is, the eigenvalues are the zeros of the Fredholm determinant, just as for a finite system of linear equations the eigenvalues are zeros of the secular equation. Most books on mathematical methods of physics—for example Riley, *et al.*, *Mathematical Methods for Physics and Engineering* (Cambridge U. Press, 1997)—describe how to construct the terms of the series expansion for the Fredholm resolvent,

$$R(x, y; \lambda) = \lambda \frac{N(x, y; \lambda)}{D(\lambda)}$$

so we shall discuss it no further here.

---

6. For a particular value (or values) of λ—that is, the eigenvalues of the kernel K.

A second solution algorithm—good in practice as well as for an existence proof—is the Schmidt method. We use the fact that a compact kernel can be approximated arbitrarily closely in norm by a kernel of finite rank. That is,

$$K(x, y) \approx K_N(x, y) \overset{df}{=} \sum_{k=1}^{N} \varphi_k(x)\, \psi_k^*(y)\,, \qquad \lim_{N \to \infty} ||K - K_N|| = 0\,.$$

In symbolic terms, then, representing

$$K = K_N + \Delta K$$

we have

$$\left(1 - \lambda \Delta K\right) f = g + \lambda \sum_{k=1}^{N} \varphi_k \left(\psi_k\,, f\right)$$

or

$$f = \left(1 - \lambda \Delta K\right)^{-1} g + \lambda \sum_{k=1}^{N} \left(1 - \lambda \Delta K\right)^{-1} \varphi_k \left(\psi_k\,, f\right).$$

Since for large enough N we may be sure that $||\lambda \Delta K|| < 1$, the operator $\left(1 - \lambda \Delta K\right)^{-1}$ can be constructed by Neumann series expansion. The preceding equation can then be replaced by a finite system of linear algebraic equations in the unknowns $f_k = \left(\psi_k\,, f\right)$.

$$f_j = \left(\psi_j\,, \left(1 - \lambda \Delta K\right)^{-1} g\right) + \lambda \sum_{k=1}^{N} \left(\psi_j\,, \left(1 - \lambda \Delta K\right)^{-1} \varphi_k\right) f_k\,;$$

to solve them we must invert the $N \times N$ matrix

$$M_{j,k} = \delta_{j\,k} - \lambda \left(\psi_j\,, \left(1 - \lambda \Delta K\right)^{-1} \varphi_k\right).$$

This is perfectly feasible in general; moreover, in many cases of physical interest only a few terms of the finite-rank approximation are necessary to achieve good accuracy in the solution.

The final method we shall discuss is the direct numerical solution of the equation. That is, we replace the integration by a finite numerical quadrature formula,

$$f(x) \approx g(x) + \lambda \sum_{k=1}^{N} w_k\, K(x, \xi_k)\, f(\xi_k)$$

which again yields a finite system of linear algebraic equations:

$$f(\xi_j) = g(\xi_j) + \lambda \sum_{k=1}^{N} w_k\, K(\xi_j\,, \xi_k)\, f(\xi_k)\,.$$

In this direct method we have no control over the precision except to increase the number of points N and see whether the solution seems to be converging (as it must). To interpolate between the solution points $f(\xi_k)$ we use the formula given above.

It is worth making some concluding remarks about the numerical solution of Fredholm equations. First, if the kernel is at all singular in the range of integration it is worth iterating it enough times that the result is non-singular. That is, formally, we can write

$$f = \left(1 + \lambda K + \lambda^2 K^2 + \ldots \lambda^{n-1} K^{n-1}\right) g + \lambda^n K^n g$$

and invert the (finite) matrix representing $1 - \lambda^n K^n$ rather than that for $1 - \lambda K$.

Second, to find the eigenvalues of the problem it is usually most efficient to evaluate the determinant of the matrix numerically rather than symbolically or as a Fredholm series.[7] In the limit as the rank of K_N becomes large, or equivalently, as the number of points in the numerical quadrature formula becomes large, the determinant of the matrix approximates the Fredholm determinant.

Finally, Gaussian quadrature rules are widely used in solving integral equations, since, for a given precision, they require many fewer points and hence much smaller matrices must be inverted. Since the standard inversion algorithms require time of order N^3, using half the number of points means an eightfold saving in computational effort.

4. Nonlinear integral equations

There is little to say about the numerical solution of nonlinear equations for the simple reason that the only useful mathematical theory (and existence proof) of their solution pertains to a limited class that represent compact mappings. That is, a nonlinear mapping of a function is a kernel $K(f)$ for which $K(\alpha f) \neq \alpha K(f)$; if it is compact, then from every infinite sequence of functions $\{f_k\}$ we can select a subsequence $\{f_{k_r}\}$ such that $\chi_r \overset{df}{=} K\left(f_{k_r}\right)$ converges to a limit. Under these circumstances, the nonlinear equation

$$f = g + \lambda K(f)$$

can be solved by iteration, as long as λ is sufficiently small in magnitude. The proof is based on the *fixed–point theorem*, which states that a compact mapping always possesses a fixed point, *i.e.* a point that is mapped into itself. The standard instance of the fixed-point theorm is stirring a teacup: the surface of the fluid is disarranged from its initial state, with various points being transformed (nonlinearly, in general!) into other points of the surface. But one point is always transformed into itself. For further details, see, *e.g.* Krasnosel'skii[8].

---

7. In fact, all linear equation solvers return the determinant as a matter of course.
8. M. A. Krasnosel'skii, *Topological methods in the theory of nonlinear integral equations* (Pergamon Press, Oxford, 1964). Note that a recent search divulged tens to hundreds of titles related to this topic, so this reference is hardly the last word.