

FREDHOLM, HILBERT, SCHMIDT

Three Fundamental Papers on
Integral Equations

Translated with commentary by
G. W. Stewart

Copyright © 2011 by G. W. Stewart. All rights reserved.

In the first decade of the twentieth century, Ivar Fredholm (1903), David Hilbert (1904), and Erhard Schmidt (1907) published three papers on integral equations that together advanced the topic from studies of special cases to a well developed general theory. This work presents English translations of these papers along with a commentary by the translator.

NOTE TO THE READER

December 15, 2011

This work is a preliminary posting of my translations of papers on integral equations by Ivar Fredholm, David Hilbert, and Erhard Schmidt along with my comments on the papers. A PDF file may be obtained at

<http://www.cs.umd.edu/~stewart/FHS.pdf>

The downloaded paper is for personal use only. Please, no wholesale copying. At present I have no plans for further publication, but that may change.

I would appreciate any comments and errata, which will be used to update this document. For the date of the latest changes go the above URL and look at the date on this page.

I would like to thank Jens Zemke for some encouraging comments at an earlier stage of this project. I owe a special debt to Folkmar Bornemann, whose comments and references have caused me to make some important additions to the commentary. The following people have also contributed errata and comments: Kendall Atkinson, Paul A. Martin, Rhys Ulerich, Rob Schreiber, Van Snyder, and Andre Tits. Thanks!

These papers are a marvelous example of the development of an important area by three outstanding mathematicians. I hope you will enjoy reading them as much as I did translating them.

Pete Stewart

stewart@cs.umd.edu

Commentary on
FREDHOLM, HILBERT, SCHMIDT
Three Fundamental Papers on
Integral Equations
G. W. Stewart

1. Introduction

An integral equation is one in which an unknown function to be determined appears in an integrand. The end of the nineteenth century saw an increasing interest in integral equations, mainly because of their connection with some of the differential equations of mathematical physics. From this work emerged four general forms of integral equations now called Volterra and Fredholm equations of the first and second kinds (a third kind was added to the canon later). Although in principle all four forms can be seen as special cases of the Fredholm equation of the second kind, in fact they have different properties and are usually treated separately.

The purpose of this work is to present translations of three papers by Ivar Fredholm [7, 1903], David Hilbert [12, 1904], and Erhard Schmidt [20, 1907] on the theory of Fredholm equations of the second kind.¹ Although the tenor of Fredholm's and Hilbert's papers had been anticipated in special cases — especially by Carl Neumann and Henri Poincaré — they were the first to treat the problem in full generality, independent of special applications. Schmidt derives and extends the results of Fredholm and Hilbert, but from an entirely different point of view.

A Fredholm equation of the second kind has the form

$$f(s) = g(s) - \lambda \int_0^1 K(s, t)g(t) dt, \quad s \in [0, 1]. \quad (1.1)$$

Here $g(s)$ is the unknown function, $f(s)$ and $K(s, t)$ are known (K is called the kernel of the equation), and λ is a parameter. Our authors assume that $f(s)$ and $K(s, t)$ satisfy certain regularity conditions on, say, $[0, 1]$ and $[0, 1] \times [0, 1]$.² They are all, however, quick to point out that their results apply to more general regions of integration in higher dimensional spaces. They also show conditions under which their regularity conditions can be relaxed.

¹The publication dates can be misleading. Fredholm presented his work in 1900 in a note to the Swedish Academy of Science [8, p. 95], and Schmidt's paper was essentially a reprint of his dissertation, which appeared in 1905.

²Hilbert and Schmidt assume continuity. Fredholm assumes integrability and finiteness. But this assumption is not sufficient for his development, at least under the current definition of integrability,

The contents of the three papers do not line up in a simple progression. Both Fredholm and Hilbert start from the corresponding linear system

$$\hat{f} = (I - \lambda \hat{K})\hat{g}, \quad (1.2)$$

where \hat{K} is a square matrix and \hat{f} and \hat{g} are vectors. But Fredholm, who implicitly takes $\lambda = -1$, is concerned with how to solve the system (1.2) in such a way that the process can be generalized to (1.1). He does not justify his generalization but simply writes down formulas and then shows, quite rigorously, that they work. In the process he treats the right-hand side of (1.1) as an operator on functions, thus ensuring his place among the founders of functional analysis. The crowning glory of his paper is an elegant theory of what happens when (1.1) is “singular,” i.e., when -1 is an eigenvalue of arbitrary multiplicity of $K(s, t)$.

Hilbert, on the other hand, takes \hat{K} to be symmetric and is concerned with generalizing the finite dimensional concept of eigenvalue and eigenvector in such a way that functions can be expanded in terms of eigenfunctions of the kernel $K(s, t)$. (It was Hilbert, by the way, who introduced the terms *Eigenwert* and *Eigenfunktion*.) Unlike Fredholm, he first develops a complete theory for linear systems and eigensystems and then by a limiting process generalizes the theory to (1.1). He is forced to assume that his eigenvalues are not multiple (although he relaxes this assumption toward the end of his paper). There is no significant use of operators.

Schmidt covers the territory mapped by Fredholm and Hilbert (and then some), but with an important difference. Instead of starting with the finite dimensional problem, he works directly with the integral equations. The result is an enormous simplification of the theory. In addition, Schmidt introduces what we would now call the singular value decomposition for unsymmetric kernels and proves an important approximation theorem associated with the decomposition.

In outline this is the story of these three papers. In the rest of this introduction we are going to expand the outline into a guided tour of these papers. We will focus on their major contributions and how they were derived, leaving aside auxiliary matters such as the relaxation of regularity assumptions. At this point it should be stressed that the primary goal of all three authors was understanding, not numerical computation, although occasional comments suggest that the latter was never far from their thoughts.

Bibliographical material on Ivar Fredholm (1866–1927), David Hilbert (1862–1943) and Erhard Schmidt (1876–1959) may be found in the *Dictionary of Scientific Biography* [10] and online in *The MacTutor History of Mathematics* [17]. In addition, Constance Reid [19] has written an excellent full-scale biography of Hilbert. The reader may wish to consult Bernkopf’s paper “The theory of function spaces with particular reference to their origins in integral equation theory” [3] as a supplement to this commentary.

A word on the translations. Our authors wrote their papers in the mathematical prose of the early twentieth century. Fredholm was terse, almost to the point of ob-

scurity, but allowing for French elaboration, his paper translates easily. Hilbert and Schmidt were by no means obscure or prolix, but they made unsparing use of German constructions that have no English equivalents. In these translations I have aimed for the English mathematical prose of the early twenty-first century. This has required rearranging or even splitting of sentences. Paragraphs have been largely left as they stand, even when they are shorter or longer than those of contemporary English. Where it is not anachronistic, I have substituted terser modern terminology for longer phrases. Still, I have tried to hew close to the original and not misrepresent the authors content for the sake of a smooth English style. Bi- or trilingual readers will be able to judge for themselves how well I have succeeded.

In the translations there are footnotes by both the authors and the translator. The former are numbered, the latter use symbols; e.g., *, †, etc.

Epilogue. Although I looked for other translations before undertaking this project, I was nervous about Fredholm's paper. And sure enough, after the first posting Folkmar Bornemann informed me that a translation had appeared in *A Source Book in Classical Analysis* edited by Garrett Birkhoff with the assistance of Uta Merzbach [4]. Although the identity of the translator is not completely clear, the evidence points to Merzbach. A line by line comparison of the translations shows that they differ little, mostly in minor stylistic points. However, I have taken liberty of using the other translation to improve mine here and there.

2. Fredholm

Fredholm's 1903 paper, a followup to a communication informally circulated in 1899,³ is concerned with the integral equation

$$\psi(x) = \varphi(x) + \int_0^1 f(x, y)\varphi(y) dy. \quad (2.1)$$

His approach is to generalize results from the theory of finite linear systems of equation, but the form in which this theory is stated is not the one found in current linear algebra books. We will therefore intersperse this discussion with a summary of the theory of finite dimensional systems of linear equations al la Fredholm. As much as possible we will preserve Fredholm's original notation.

In his paper, Fredholm considers two distinct cases. The first and simplest is when (2.1) always has a unique solution. The second is when (2.1) only has a solution if $\psi(x)$ satisfies certain conditions relating to the kernel $f(x, y)$. In that case the solution is not unique, although the set of solutions can be characterized. The two cases exhaust all possibilities and for that reason are collectively known as the Fredholm alternative.

³In his 1904 paper, Hilbert mentions this paper, saying that it was cited in Fredholm's 1903 paper. But no such citation appears there.

It will be convenient to treat the two cases separately. We will describe the first case in the context of finite dimensional spaces—the spaces of matrices and vectors—followed by a description of how this plays out in his paper. Then we will repeat the process for the second case. The survey concludes with a brief description of the material in the rest of the paper.

We begin by writing (1.1) in operator form as

$$\psi = S_f\varphi = \varphi + K_f\varphi = (I + K_f)\varphi. \quad (2.2)$$

Here K_f stands either for the kernel $f(x, y)$ or for a matrix. In the former case, $K_f\varphi$ represents the integral in (1.1); in the latter, a matrix-vector multiplication. When the concern is the product of operators, say K_f and K_g , the product K_fK_g is equivalent to $\int_0^1 f(x, t)g(t, y) dt$, while K_gK_f is equivalent to $\int_0^1 g(x, t)f(t, y) dt$.

The operation $I + K_f$ appearing in the right-hand side of (2.2) already shows one of the difficulties in generalizing results from linear systems. Specifically, the identity operator cannot be realized as an integral operator with a continuous kernel. Nonetheless, it is an operator, and we can combine it with integral operators, as in (2.2).⁴

Assuming that S_f has a left inverse $S_g = I + K_g$, we can write

$$(I + K_g)(I + K_f) = I, \quad (2.3)$$

or

$$K_f + K_g + K_gK_f = 0. \quad (2.4)$$

Note that if K_g satisfies (2.3), then it satisfies (2.4) and vice versa.

Let us now see what the existence of a left inverse buys us. Suppose that $S_f\varphi = \psi$. Then on multiplying this equation by S_g we get

$$\varphi = S_gS_f\varphi = S_g\psi.$$

Thus if (2.1) has a solution, it is unique and given by $S_g\psi$.

But this does not show that a solution exists. To do this, we need a right inverse satisfying $S_fS_g = I$, or equivalently

$$K_f + K_g + K_fK_g = 0. \quad (2.5)$$

For then if $\varphi = S_g\psi$, then $S_f\varphi = S_fS_g\psi = \psi$, so that $S_g\psi$ is the desired solution.

In finite dimensional spaces the difference between left and right inverses is moot. If one exists, so does the other; moreover, the inverses are unique and equal to each

⁴If actual integrals must be used, one can define the identity function $I(x, y)$ as $\delta(x - y)$, where $\delta(t)$ is the Dirac delta function, a generalized function that is zero when $t \neq 0$ and that satisfies $\int_{-\infty}^{\infty} f(x)\delta(x) = f(0)$. We will use this option in motivating one of Hilbert's definitions (see page 17).

For the moment we will skip §1.6 and move to §§2.7–8, where Fredholm considers the solution of (1.1) in the case where $D_f \neq 0$. He introduces the operator

$$S_f = \varphi(x) \mapsto \varphi(x) + \int_0^1 f(x, s)\varphi(s) ds$$

and observes that operators of this form constitute a group, by which he means only that the class of operators is closed under multiplication. They are not required to have inverses.

Fredholm (§8) now assume that $D_f \neq 0$ and sets

$$g(x, y) = -\frac{D_f \begin{pmatrix} x \\ y \end{pmatrix}}{D_f},$$

which corresponds to our K_g in (2.3). It then follows from the fundamental relations (2.8) and (2.9), that K_f and K_g satisfy (2.4) and (2.5). Fredholm goes on to argue, much as we did above, that $S_g\psi(x)$ is the unique solution of the equation $\psi(x) = S_f\varphi(x)$.

This theorem is a remarkable achievement. In about six pages, Fredholm gives a completely rigorous solution to the equation (1.1). Moreover, he does it by treating the right-hand side of (1.1) as one of a class of operators on functions that can be manipulated in their own right. In doing this, Fredholm made a key step toward abstract functional analysis.

The question remains of how Fredholm derived his formulas. We cannot know for sure, since he does not tell us in his paper. However, in an address to the Swedish Congress of Mathematics (1909) [8, p.95] he said, "...the works of my colleague von Koch on infinite determinants greatly facilitated my research ...". Toward the end of this section we will try to assess what von Koch's work may have contributed to Fredholm's thinking. But for now, we will give an informal derivation of the Fredholm determinant that may be found in books by Lovitt [15, p. 23, ff.], Tricomi [23, §2.5], and others.

Consider first the Fredholm determinant D_f . We begin with an equally spaced mesh on $[0,1]$ consisting of the n points $0 < x_1 < x_2 < \dots < x_n = 1$. Set $h = x_{i+1} - x_i$, $f_{ij} = f(x_i, x_j)$ and

$$D_n = \det \begin{pmatrix} 1 + hf_{11} & hf_{12} & \cdots & hf_{1n} \\ hf_{21} & 1 + hf_{22} & \cdots & hf_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ hf_{n1} & hf_{n2} & \cdots & 1 + hf_{nn} \end{pmatrix} \quad (2.10)$$

Now this determinant can be expanded in the form (see [23, p.66–68])

$$\begin{aligned}
 D_n = & 1 + \sum_{i=1}^n h f_{ii} + \frac{1}{2!} \sum_{i,j=1}^n h^2 \det \begin{pmatrix} f_{ii} & f_{ij} \\ f_{ji} & f_{jj} \end{pmatrix} + \dots \\
 & + \frac{1}{n!} \sum_{i,j,\dots,k=1,n} h^n \det \begin{pmatrix} f_{ii} & f_{ij} & \dots & f_{ik} \\ f_{ji} & f_{jj} & \dots & f_{jk} \\ \vdots & \vdots & & \vdots \\ f_{ki} & f_{kj} & \dots & f_{kk} \end{pmatrix}.
 \end{aligned} \tag{2.11}$$

But as $n \rightarrow \infty$ the k th term in this expansion approaches

$$\frac{1}{k!} \int_0^1 \int_0^1 \dots \int_0^1 f \begin{pmatrix} x_1 & x_2 & \dots & x_k \\ x_1 & x_2 & \dots & x_k \end{pmatrix} dx_1 dx_2 \dots dx_k,$$

in which we have used Fredholm’s abbreviated terminology (2.6). Thus taking a formal limit of the sum (2.11), we get the expression (2.7), which is now seen as a generalization of the matrix determinant to the operator S_f .

The function $D \begin{pmatrix} x \\ y \end{pmatrix}$ is closely related to the adjugate of a matrix. To see this, note that the adjugate A^{adj} of a matrix A is the transpose of the matrix whose (i, j) -element is the cofactor of a_{ij} (that is, $(-1)^{i+j}$ times the determinant of the matrix obtained by striking row i and column j of A). It is well known that

$$A^{\text{adj}}A = AA^{\text{adj}} = \det(A)I. \tag{2.12}$$

It follows that if $\det(A) \neq 0$, then $A^{-1} = A^{\text{adj}}/\det(A)$. In particular, since A^{adj} is a continuous function of the elements of A , if A approaches the singular matrix A_0 , then $\det(A)A^{-1}$ approaches A_0^{adj} .

Returning to our operator $I + K_f$, we have seen that its inverse is

$$I + K_g = I - \frac{D_f \begin{pmatrix} x \\ y \end{pmatrix}}{D_f}.$$

Thus, formally,

$$(I + K_f)^{\text{adj}} = D_f I - D_f \begin{pmatrix} x \\ y \end{pmatrix}.$$

When D_f is zero,

$$(I + K_f)^{\text{adj}} = -D_f \begin{pmatrix} x \\ y \end{pmatrix}. \tag{2.13}$$

There is a minor conundrum in Fredholm's terminology. He calls $D_f \begin{pmatrix} x \\ y \end{pmatrix}$ a minor. Yet we have seen that it represents the adjugate, which is properly called a compound determinant. The problem is resolved by observing that regarded as an operator $D_f \begin{pmatrix} x \\ y \end{pmatrix}$ is indeed the adjugate. But its value for a particular value of x and y represents an element of the adjugate, which, as we saw above, is a minor.

Another property of the adjugate matrix relates to the case where $D_f = 0$. It is known that if A is singular and has rank one less than the order of A then A^{adj} has rank one, and its nonzero columns are nontrivial solutions of the equation $Ax = 0$. If we generalize this fact to the operator S_f , then (2.13) suggests that if $D_f = 0$ and $D_f \begin{pmatrix} x \\ y \end{pmatrix}$ is not identically zero, then for some η the function $\varphi(x) = D_f \begin{pmatrix} x \\ \eta \end{pmatrix}$ satisfies $S_f \varphi(x) = 0$. This fact and its generalizations to the cases where $D_f \begin{pmatrix} x \\ y \end{pmatrix}$ are identically zero is at the heart of Fredholm's treatment of the case $D_f = 0$, to which we now turn.

Once again, it will be informative to consider the finite dimensional case. We will work with the operator S_f , which is now a matrix with determinant D_f .

If $D_f = 0$, then S_f has right and left null spaces Φ and Ψ of common dimension n . Suppose we can find a pseudo-inverse S_g satisfying the following two conditions.⁷

1.

$$S_g S_f = I - \Phi X^*, \quad (2.14)$$

where the columns of Φ form a basis for Φ and $X^* \Phi = I$.

2.

$$S_f S_g = I - Y \Psi^*, \quad (2.15)$$

where the columns of Ψ form a basis for Ψ and $\Phi^* Y = I$.

Then we can draw the following conclusion.

1. If φ is a solution of the equation $S_f \varphi = \psi$, then $\Psi^* \psi = \Psi^* S_f \varphi = 0$. Hence a necessary condition for the equation $S_f \varphi = \psi$ to have a solution is that $\Psi^* \psi = 0$.
2. If $\varphi = S_g \psi$, where ψ satisfies the necessary conditions $\Psi^* \psi = 0$, then $S_f \varphi = S_f S_g \psi = (I - Y \Psi^*) \psi = \psi$, so that φ is a solution of $S_f \varphi = \psi$. Thus the condition that $\Psi^* \psi = 0$ is a sufficient condition for a solution to exist.
3. If φ is a solution of $S_f \varphi = \psi$, then, since $S_f \Phi = 0$, $S_f(\varphi + \Phi u)$ is also a solution of the equation for any u . Thus the set of all solutions is an affine subspace.

⁷The term pseudo-inverse is Fredholm's.

Turning now to Fredholm's solution of the equation $S_f\varphi = \psi$ when $D_f = 0$, he begins (back in §1.6) by introducing the function $f_\lambda(x, y) = \lambda f(x, y)$ and showing that the function $D_{\lambda f}$ is an entire function of λ . Consequently, if for some λ we have $D_{\lambda f} = 0$, then only a finite number of the derivatives of $D_{\lambda f}$ can vanish at λ . But it can be shown that

$$\lambda^n \frac{d^n D_{\lambda f}}{d\lambda^n} = \int_0^1 \cdots \int_0^1 D_{\lambda f} \begin{pmatrix} x_1 \cdots x_n \\ x_1 \cdots x_n \end{pmatrix} dx_1 dx_2 \cdots dx_n,$$

and it follows that not all of the minors $D_{\lambda f} \begin{pmatrix} \xi_1 \cdots \xi_n \\ \eta_1 \cdots \eta_n \end{pmatrix}$ can vanish. In particular for $\lambda = 1$, if $D_f = 0$, then there is a least integer $n > 1$ such that

$$D_f \begin{pmatrix} \xi_1 \cdots \xi_n \\ \eta_1 \cdots \eta_n \end{pmatrix}$$

is not identically zero.

The story now jumps to §2.9, where Fredholm considers the solution of $S_f\varphi = \psi$ when $D_f = 0$. He takes n as in the last paragraph. Making use of the relations among the minors, he first shows that the function

$$D_f \begin{pmatrix} x, \xi_2, \dots, \xi_n \\ \eta_1, \eta_3, \dots, \eta_n \end{pmatrix}$$

is a solution of the homogeneous equation $D_f\varphi = 0$.

He then introduces what he calls the *pseudo-inverse* S_g defined by

$$g(x, y) = - \frac{D_f \begin{pmatrix} x, \xi_1 \cdots \xi_n \\ y, \eta_1 \cdots \eta_n \end{pmatrix}}{D_f \begin{pmatrix} \xi_1 \cdots \xi_n \\ \eta_1 \cdots \eta_n \end{pmatrix}},$$

where the ξ_i and η_i are chosen to make the denominator nonzero. Writing $S_g S_f = I + F$, he shows that

$$F(x, y) = - \sum_{\nu=1}^n f(\xi_\nu, y) \Phi_\nu(x),$$

where the functions

$$\begin{aligned}\Phi_1(x) &= -\frac{D_f \left(\begin{matrix} x, \xi_2, \dots, \xi_n \\ \eta_1, \eta_2, \dots, \eta_n \end{matrix} \right)}{D_f \left(\begin{matrix} \xi_1, \xi_2, \dots, \xi_n \\ \eta_1, \eta_2, \dots, \eta_n \end{matrix} \right)} \\ \Phi_2(x) &= +\frac{D_f \left(\begin{matrix} \xi_1, x, \dots, \xi_n \\ \eta_1, \eta_2, \dots, \eta_n \end{matrix} \right)}{D_f \left(\begin{matrix} \xi_1, \xi_2, \dots, \xi_n \\ \eta_1, \eta_2, \dots, \eta_n \end{matrix} \right)} \\ &\text{etc.}\end{aligned}$$

satisfy $S_f \Phi_i = 0$ ($i = 1, \dots, n$). He then shows that any solution of the homogeneous equation is a linear combination of the Φ_i . Moreover,

$$\int_0^1 f(\xi_\lambda, x) \Phi_\mu(x) dx = \begin{cases} 0 & \text{if } \lambda \neq \mu, \\ 1 & \text{if } \lambda = \mu, \end{cases}$$

which implies that the Φ_i are independent and biorthogonal to the function $f(\xi_i, x)$.

If we set $\Phi(x) = (\Phi_1(x) \cdots \Phi_n(x))$ and $X(y)^* = (f(\xi_1, y), \dots, f(\xi_n, y))$, then

$$F(x, y) = -\Phi(x)X(y)^*,$$

which establishes (2.14). Thus Fredholm is halfway to a proof of his theorem.

This is a good point to pause and ask if the introduction of the equation (2.14) in this commentary is anachronistic. In one sense it is, since the matrix-like operations it implies were unknown to Fredholm. But in another sense it is merely a compact representation of the equations already in Fredholm's paper, a notational device designed to allow the modern reader to easily grasp what Fredholm proved. Taken in this sense, equation (2.14) is just a useful expository tool.

Fredholm also writes down the equation (2.15) (less the biorthogonality condition which he does not need). However, there is a small surprise here. It is evident that (2.15) transposed is the same as (2.14) for the adjoint equation

$$\psi(x) = \varphi(x) + \int_0^1 f(y, x) \varphi(y) dy.$$

Yet Fredholm does not take advantage of this fact. Instead he derives the results he needs piecemeal without direct reference to the adjoint operator, even though he uses it later in the paper. One can only wonder why Fredholm, having discovered the adjoint operator, did not go back to his treatment of the case $D_f = 0$ and clean things up.

Fredholm calls the operator S_g a pseudo-inverse of S_f , and it is of interest to see how they relate to the various generalized inverses that have been introduced since his time. It is easy to show from either (2.15) or (2.14) that $S_f S_g S_f = S_f$. This relation is the first of the four Penrose conditions for the Moore–Penrose generalized inverse of a matrix (see, for example, the preface in [16]). An operator S_g satisfying this condition is called a (1)-generalized inverse. An immediate consequence of this fact is that $S_g S_f$ and $S_f S_g$ are idempotent — that is they are projectors, as (2.14) and (2.15) show. Although there appears to be no direct link between Fredholm’s work and later work on generalized inverses, he should be given credit for exhibiting and naming one of the first examples.

The rest of the paper consists of a potpourri of miscellaneous results. In §3 Fredholm derives an expression for the logarithmic derivative of D_f with respect to the kernel $f(x, y)$. In §4 he uses this result to establish the following product theorem. Let f and g be kernels and let $S_F = S_f S_g$. Then $D_F = D_f D_g$. Here he again introduces the adjoint operator to aid in the proof.

In §5 Fredholm considers some series expansions. Given the resolvent kernel $\varphi(x, y)$, he shows that it can be formally expanded in the form

$$\varphi(\xi, \eta) = \varphi_1(\xi, \eta) - \varphi_2(\xi, \eta) + \varphi_3(\xi, \eta) - \cdots,$$

where

$$\begin{aligned} \varphi_1(\xi, \eta) &= f(\xi, \eta) \\ \varphi_n(\xi, \eta) &= \int_0^1 f(\xi, \tau) \varphi_{n-1}(\tau, \eta) d\tau \quad (n = 2, 3, \dots). \end{aligned} \tag{2.16}$$

He points out that this expansion converges provided f is sufficiently small. This is, of course, the Neumann expansion mentioned above.

Some kernels arising in practice have infinite singularities along the line $x = y$. In §6 Fredholm considers how to solve such equations with kernel $i(x, y)$ where $(x - y)^\alpha i(x, y)$ ($0 < \alpha < 1$) is finite and integrable. His approach relies on the smoothing power of integration. Specifically, he shows that for n large enough, the iterated kernel $i_n(x, y)$ defined in analogy with (2.16) is finite. He then poses an equation associated with i_n and shows how its solution, in some cases, can provide a solution of the original equation. Once again he makes explicit use of the adjoint kernel in his development.

Before concluding this section it is necessary to assess the influence of von Koch, whose paper “On infinite determinants and linear differential equations” [24] was published in 1892. As mentioned above, Fredholm acknowledged a debt to von Koch, but not in his original paper.

Von Koch’s paper is in two parts, the first of which concerns determinants of matrices of infinite order and the solution of linear systems involving such matrices. In the first section of the first part, he gives conditions on his infinite matrices under which the determinant and its minors are well defined as limits of finite subdeterminants. In the

second section he gives formulas involving minors for the solutions of the homogeneous system, showing in the process that the space of such solutions must be of finite rank. However, he dismisses the inhomogeneous system by writing, “For the case of an infinite system of linear inhomogeneous equations, . . . one can easily establish a corresponding theorem by similar considerations.”

Although all of von Koch’s results have finite-dimensional versions, it is nonetheless clear that Fredholm’s development owes much to von Koch’s especially in his treatment of the inhomogeneous equation. Fredholm’s notation is clearly derived from von Koch’s as are his formulas for the Ψ ’s. Fredholm’s approach to establishing the finiteness of the space of solutions is analogous to von Koch’s: namely, start with the existence of a least minor that is in some sense nonzero. On the other hand, there is no trace of the Fredholm alternative in von Koch.

Although Fredholm’s determinants and minors are analogous to von Koch’s in that they both satisfy a number of analogous relations, their constructions are quite different—a little like constructions in ℓ_2 and L_2 or, for that matter, Heissenberg’s matrix mechanics and Schrödinger’s equation. It is true that von Koch gives a series for the determinant that is analogous to Fredholm’s formula (and which was already known in a finite-dimensional) setting. But it von Koch’s series is not a springboard to the Fredholm formula. This inclines me to believe that Fredholm must have used techniques resembling those described on page 6.

All this granted, Fredholm’s achievement is nothing less than impressive, and what impresses most is its completeness. He took the problem of solving a nonsymmetric integral equation of the second kind and by his alternative not only showed when solutions exist but also characterized their nonuniqueness. In the process he showed the finiteness of independent solutions of the homogeneous equation and introduced operators into the analysis. He also has a claim to be one of the founding fathers of the subject of generalized inverses. But his equation was just that—a single equation. By parameterizing the equation, Hilbert would show that there was more to be said on the subject, at least when the kernel is symmetric.

3. Hilbert

Hilbert’s paper appeared in 1904 just a year after Fredholm’s paper, seemingly too short a time for the latter to have much influenced the former. But, as Hilbert notes, Fredholm communicated his results in 1899, and, as we have seen, Fredholm says he presented the results to the Swedish Academy of Science in 1900. Moreover, according to Hermann Weyl [25] E. Holmgren gave “a lecture in 1901 in Hilbert’s seminar dealing with the now classical paper of Fredholm’s on *integral equations*, then but recently published.” Thus Hilbert had ample time to assimilate Fredholm’s work. Although Hilbert’s debt to Fredholm is evident, Hilbert was concerned not so much with solving integral equations

of the second kind as with developing a theory of eigenvalues and eigenfunctions, and especially of the expansion of other functions in series of eigenfunctions. In this respect, Hilbert’s contribution was broader than Fredholm’s. But in other respects, it was more circumscribed. In particular, Hilbert was forced to assume that his kernel was symmetric and that its eigenvalues were simple.

There is also a difference in the way they treat the original finite-dimensional linear system. By the time Fredholm started his paper, the original system had vanished, leaving only the names “determinant” and “minor” to suggest the origins of his constructions. Hilbert, on the other hand, develops the algebraic problem—that is the finite-dimensional theory of linear equations and eigensystems—in such a way that one can pass directly to the infinite dimensional case by taking limits. Hilbert writes his integral equation in the form

$$f(s) = \varphi(s) - \lambda \int_a^b K(s, t)\varphi(t) dt, \tag{3.1}$$

a notation that hints, at least to the present-day reader, of the eigenvalues and eigenfunctions to come. The kernel K is assumed to be continuous and symmetric [i.e., $K(s, t) = K(t, s)$]. Hilbert variously takes the interval of integration to be $[a, b]$ or $[0, 1]$. For uniformity we will use the latter.

Hilbert immediately discretizes this problem by introducing the following notation

$$K_{pq} = K\left(\frac{p}{n}, \frac{q}{n}\right) \quad (p, q = 1, 2, \dots, n)$$

$$Kxy = K_{11}x_1y_1 + K_{12}x_1y_2 + K_{21}x_2y_1 + \dots + K_{nn}x_ny_n$$

$$= \sum_{pq} K_{pq}x_py_q, \quad (K_{pq} = K_{qp}),$$

$$\varphi_p = \varphi\left(\frac{p}{n}\right), \quad f_p = f\left(\frac{p}{n}\right), \quad (p = 1, 2, \dots, n),$$

$$Kx_1 = K_{11}x_1 + K_{12}x_2 + \dots + K_{1n}x_n,$$

$$Kx_2 = K_{21}x_1 + K_{22}x_2 + \dots + K_{2n}x_n,$$

$$\dots\dots\dots$$

$$Kx_n = K_{n1}x_1 + K_{n2}x_2 + \dots + K_{nn}x_n,$$

$$[x, y] = x_1y_1 + x_2y_2 + \dots + x_ny_n.$$

Thus the algebraic problem approximating (3.1) is

$$\begin{aligned} f_1 &= \varphi_1 - \ell K\varphi_1, \\ &\dots\dots\dots \\ f_n &= \varphi_n - \ell K\varphi_n, \end{aligned} \tag{3.2}$$

where ℓ the equivalent of λ in (3.1).

Hilbert now introduces the two determinants

$$d(\ell) = \begin{vmatrix} 1 - \ell K_{11} & -\ell K_{12} & \cdots & -\ell K_{1n} \\ -\ell K_{21} & 1 - \ell K_{22} & \cdots & -\ell K_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -\ell K_{n1} & -\ell K_{n2} & \cdots & 1 - \ell K_{nn} \end{vmatrix}$$

and

$$D \left(\ell, \begin{matrix} x \\ y \end{matrix} \right) = \begin{vmatrix} 0 & x_1 & x_2 & \cdots & x_n \\ y_1 & 1 - \ell K_{11} & -\ell K_{12} & \cdots & -\ell K_{1n} \\ y_2 & -\ell K_{21} & 1 - \ell K_{22} & \cdots & -\ell K_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y_n & -\ell K_{n1} & -\ell K_{n2} & \cdots & 1 - \ell K_{nn} \end{vmatrix}. \quad (3.3)$$

Note that $D \left(\ell, \begin{matrix} x \\ y \end{matrix} \right)$ is a bilinear function of x and y . These two determinants satisfy the relation

$$d(\ell)[x, y] + D \left(\ell, \begin{matrix} x \\ y \end{matrix} \right) - \ell D \left(\ell, \begin{matrix} x \\ Ky \end{matrix} \right) = 0,$$

which is analogous the two Fredholm relations (2.8) and (2.9). But because of symmetry Hilbert requires only one relation to Fredholm's two.

Hilbert now shows by elementary manipulations that if $d(\ell) \neq 0$ then the solution φ of (3.2) satisfies

$$[\varphi, y] = -\frac{D \left(\ell, \begin{matrix} f \\ y \end{matrix} \right)}{d(\ell)}. \quad (3.4)$$

In other words, φ is the vector of coefficients of y in the bilinear form on the right-hand side of (3.4).

Hilbert's derivation of (3.4) requires only thirteen lines of text and equations, and each step in the argument is clear. Yet to modern eyes it is unsatisfying because it gives no insight into why (3.4) solves the algebraic problem. Here we will give a different and perhaps more enlightening derivation.

The key observation is that

$$D \left(\ell, \begin{matrix} x \\ y \end{matrix} \right) = -y^*(I - \ell K)^{\text{adj}}x, \quad (3.5)$$

where $(I - \ell K)^{\text{adj}}$ is the adjugate matrix of $I - \ell K$. This can easily be verified by expanding (3.3) along the first row and column. Now supposing $d(\ell) \neq 0$, we have $d(\ell)^{-1}(I - \ell K)^{\text{adj}} = (I - \ell K)^{-1}$. Hence

$$-D \left(\ell, \begin{matrix} f \\ y \end{matrix} \right) / d(\ell) = y^*(I - \ell K)^{-1}f = [y, \varphi],$$

with the same uniform convergence as above. He then sets

$$\Delta(\lambda; s, t) = \lambda \left\{ \Delta \left(\lambda, \begin{matrix} x \\ y \end{matrix} \right) \right\}_{\substack{x(r)=K(r,s) \\ y(r)=K(r,t)}} - \delta(\lambda)K(s, t), \quad (3.8)$$

and shows that

$$\delta(\lambda)K(s, t) + \Delta(\lambda; s, t) - \lambda \int_0^1 \Delta(\lambda; s, r)K(t, r) dr = 0.$$

Finally, assuming that $\delta(\lambda) \neq 0$, he defines

$$\mathbb{K}(s, t) = -\frac{\Delta(\lambda; s, t)}{\delta(\lambda)}, \quad (3.9)$$

from which it follows that

$$K(s, t) = \mathbb{K}(s, t) - \lambda \int_0^1 \mathbb{K}(s, r)K(t, r) dr.$$

In particular, if $\varphi(x)$ satisfies the Fredholm integral equation (3.1), then

$$\varphi(s) = f(s) + \lambda \int_0^1 \mathbb{K}(s, t)f(t) dt.$$

Hilbert calls $\mathbb{K}(s, t)$ the solution function (*die lösende Funktion*) for the kernel $K(s, t)$ because it solves the equation (3.1). Today it is known as the resolvent kernel and is widely used in functional analysis. Later in the paper, Hilbert will establish the important resolvent identity

$$\mathbb{K}(\mu; s, t) - \mathbb{K}(\lambda + \mu; s, t) = \lambda \int_0^1 \mathbb{K}(\lambda + \mu; r, s)\mathbb{K}(\mu; r, t) dr.$$

Hilbert provides little motivation for his definition of $\Delta(\lambda; s, t)$ in (3.8), and it is instructive to see, informally, how it might be derived. Hilbert is looking for an operator \mathbb{K} such that

$$(I - \lambda K)^{-1} = I + \lambda \mathbb{K}.$$

If we formally expand $(I - \lambda K)^{-1}$ in a Neumann series we find that

$$\mathbb{K} = K^1 + \lambda K^2 + \lambda^2 K^3 + \dots, \quad (3.10)$$

where K^i is the iterated kernel defined as in (2.16).

Now from (3.5) one might expect that in the limit

$$\Delta \left(\lambda, \begin{matrix} p(x) \\ q(y) \end{matrix} \right) = -\delta(\lambda) \left(\int_0^1 \int_0^1 q(y) [I - \lambda K(x, y)]^{-1} p(x) dx dy \right).$$

Once again taking the Neumann expansion of the inverse

$$\begin{aligned} \Delta \left(\lambda, \begin{matrix} p(x) \\ q(y) \end{matrix} \right) &= -\delta(\lambda) \left(\int_0^1 \int_0^1 q(y) [K^0(x, y) + \lambda K^1(x, y) \right. \\ &\quad \left. + \lambda^2 K^2(x, y) + \dots] p(x) dx dy \right). \end{aligned} \quad (3.11)$$

Here we take $K^0(x, y)$ to be $d(x - y)$, where d is the Dirac delta function (see footnote 4 on page 4). If we now set $p(x) = K(x, s)$ and $q(y) = K(y, t)$ in (3.11), we get

$$\begin{aligned} \Delta \left(\lambda, \begin{matrix} K(x, s) \\ K(y, t) \end{matrix} \right) &= -\delta(\lambda) \left(\int_0^1 \int_0^1 [K(x, s) K^0(x, y) K(y, t) \right. \\ &\quad \left. + \lambda K^1(x, s) K(x, y) K(y, t) + \lambda K(x, s) K^2(x, y) K(y, t) + \dots] dx dy \right). \end{aligned}$$

But $\int_0^1 \int_0^1 K(x, s) K^i(x, y) K(y, t) dx dy = K^{i+2}(s, t)$. It follows that

$$\Delta \left(\lambda, \begin{matrix} K(x, s) \\ K(y, t) \end{matrix} \right) - \delta(\lambda) K(s, t) = -\delta(\lambda) (K^1(s, t) + \lambda K^2(s, t) + \lambda^2 K^3(s, t) + \dots).$$

Thus if we define $\Delta(\lambda; s, t)$ by (3.8) and $\mathbb{K}(s, t)$ by (3.9), the expansion of the latter agrees with (3.10).

Hilbert concludes this part with an expression for the derivative of $\delta'(\lambda)$:

$$\delta'(\lambda) = \int_0^1 \Delta(\lambda; s, s) ds. \quad (3.12)$$

The second part of the passage to the limit concerns the spectrum of $K(s, t)$. As it turns out, the wherewithal to define eigenvalues and eigenfunctions has already been developed in the first part. This is not as surprising as it might at first seem. Recall that in Fredholm's theory if $D_f = 0$ then $\varphi(x) = D_f \begin{pmatrix} x \\ \eta \end{pmatrix}$ is a solution of the homogeneous equation $S_g \varphi(x) = 0$. But this means that $\varphi(x)$ is an eigenfunction of $K(x, y)$ with eigenvalue -1 . Since Hilbert has just recapitulated part of Fredholm's theory, he should also be able to define eigenvalues and eigenfunctions straightaway.

Hilbert begins by showing that $\delta(\lambda)$ has no complex zeros. He then shows that if the zeros $\ell^{(h)}$ of $d(\ell)$ are suitably ordered and the zeros $\lambda^{(h)}$ of $\delta(\lambda)$ are ordered in the

same way, then $\lim_{n \rightarrow \infty} n\ell^{(h)} = \lambda^{(h)}$. He quickly points out that this theorem does not imply that $\delta(\lambda)$ has any zeros, since the $\ell^{(h)}$ may diverge.

Hilbert then defines the eigenvalues of $K(s, t)$ to be the zeros of $\delta(\lambda)$. He also assumes that any eigenvalue λ is of multiplicity one and hence that $\delta'(\lambda) \neq 0$.

Hilbert now turns to the definition of the eigenfunction corresponding to an eigenvalue $\lambda^{(h)}$. From (3.12) and the fact that $\lambda^{(h)}$ is a simple eigenvalue, we get

$$\int_0^1 \Delta(\lambda; s, s) ds \neq 0.$$

Since $\Delta(\lambda; s, s)$ is continuous, there must be a value s^* such that $\Delta(\lambda; s^*, s^*) \neq 0$. Hilbert then defines

$$\varphi^{(h)}(s) = \left| \sqrt{\frac{\lambda^{(h)}}{\Delta(\lambda^{(h)}; s^*, s^*)}} \right| \Delta(\lambda^{(h)}; s, s^*),$$

and shows that $\varphi^{(h)}(s)$ has the properties of an eigenfunction corresponding to $\lambda^{(h)}$.

These properties simplify if we work with the normalized eigenfunctions

$$\psi^{(h)}(s) = \frac{\varphi^{(h)}(s)}{\left| \sqrt{\int_0^1 (\varphi^{(h)}(s))^2 ds} \right|}.$$

Specifically,

$$\begin{aligned} \frac{\Delta\left(\lambda^{(h)}, \begin{matrix} x \\ y \end{matrix}\right)}{\lambda^{(h)}\delta'(\lambda^{(h)})} &= \int_0^1 \psi^{(h)}(s)x(s) ds \cdot \int_0^1 \psi^{(h)}(s)y(s) ds \\ \int_0^1 (\psi^{(h)}(s))^2 ds &= 1, \\ \int_0^1 \psi^{(h)}(s)\psi^{(k)}(s) ds &= 0 \quad (h \neq k), \\ \psi^{(h)}(s) &= \lambda^{(h)} \int_0^1 K(s, t)\psi^{(h)}(t) dt. \end{aligned}$$

Thus the $\psi^{(h)}$ form an orthonormal system of eigenfunctions for $K(s, t)$.

Finally, Hilbert turns to the generalization of (3.7), and arrives at the following formula:

$$\begin{aligned} \int_0^1 \int_0^1 K(s, t)x(s)y(t) ds dt &= \frac{1}{\lambda^{(1)}} \int_0^1 \psi^{(1)}(s)x(s) ds \cdot \int_0^1 \psi^{(1)}(s)y(s) ds \\ &+ \frac{1}{\lambda^{(2)}} \int_0^1 \psi^{(2)}(s)x(s) ds \cdot \int_0^1 \psi^{(2)}(s)y(s) ds + \dots \end{aligned}$$

The convergence is uniform in the sense that the error in the m th approximation is bounded by

$$\frac{1}{2|\lambda^{(m+1)}|} \left(\int_0^1 (x(s))^2 ds + \int_0^1 (y(s))^2 ds \right).$$

Thus if $x(s)$ and $y(s)$ are uniformly bounded, the convergence, whatever $x(s)$ and $y(s)$, is bounded by a common multiple of $1/|\lambda^{(m+1)}|$, which approaches zero at a rate depending only on $K(s, t)$.⁸

In the introduction to his paper, Hilbert asserts that his theory will deliver a simple proof of the existence of eigenvalues of a symmetric kernel. In the first part of Section IV, he attempts to justify his claim. He does this by using his theory to prove that $K(s, t)$ has a finite number, say m , of eigenvalues if and only if $K(s, t)$ can be written in the form

$$K(s, t) = \frac{1}{\lambda^{(1)}} \psi^{(1)}(s) \psi^{(1)}(t) + \cdots + \frac{1}{\lambda^{(m)}} \psi^{(m)}(s) \psi^{(m)}(t). \quad (3.13)$$

How this implies the existence of eigenvalues is not clear, since the above expression is established under the hypothesis that eigenvalues exist. Possibly Hilbert means that if we set $m = 0$ in (3.13) then we get $K(s, t) = 0$ by the convention that an empty sum is zero. This would lead to the conclusion that only a zero kernel can fail to have eigenvalues. But a convention is just a convention, and in special cases, like this one, it must be proved that it is applicable—something Hilbert does not do. Although Schmidt mentions Hilbert's existence proof in the introduction to his own paper, he nonetheless finds it necessary to include a different proof in the body of his paper. Moreover, Courant and Hilbert included an explicit existence proof in their celebrated *Methoden der mathematischen Physik I* [5, pp 104–107].

The remainder of Section IV is devoted the expansion of arbitrary [*wilkürlich*] functions in terms of eigenfunctions. By arbitrary Hilbert does not mean just any function; rather he seeks conditions which a function must satisfy to be expandable. The material in this section is straightforward, and we will simply summarize the results.

As usual, Hilbert begins with a kernel $K(s, t)$ whose normalized eigenfunctions are $\psi^{(1)}(s), \psi^{(2)}(s), \dots$. He first shows that any function that can be represented in the form

$$f(s) = \int_0^1 \int_0^1 K(r, t) K(s, t) h(r) dr dt \quad (3.14)$$

has the expansion

$$\begin{aligned} f(s) &= c_1 \psi^{(1)}(s) + c_2 \psi^{(2)}(s) + \cdots, \\ c_m &= \int_0^1 f(s) \psi^{(m)}(s) ds, \end{aligned} \quad (3.15)$$

where the convergence is absolute and uniform.

⁸Here Hilbert assumes that the number of eigenvalues is infinite. Since $\delta(\lambda)$ is an entire function, this implies that the eigenvalues cannot have an upper bound; i.e., they must increase to infinity.

Seeking more general conditions, Hilbert proves two more theorems, both of which depend on the result (3.14) and (3.15). First, he defines $K(s, t)$ to be closed if for any function $g(s)$ that is not identically zero there is a t such that $\int_0^1 K(s, t)g(s)ds \neq 0$. He then shows that if $h(s)$ is closed and

$$\int_0^1 h(s)\psi^{(m)}(s) ds = 0, \quad m = 1, 2, \dots$$

then $h(s)$ is identically zero. He then goes on to show that if $K(s, t)$ is closed and $f(s)$ is such that the series

$$c_1\psi^{(1)}(s) + c_2\psi^{(2)}(s) + \dots ,$$

$$c_m = \int_0^1 f(s)\psi^{(m)}(s) ds,$$

converges uniformly, then it converges to $f(s)$.

Finally, Hilbert introduces the notion of a general kernel. Specifically, a kernel is general if for any function $g(s)$ and any $\epsilon > 0$ there is a function $h(t)$ such that

$$\int_0^1 \left(g(s) - \int_0^1 K(s, t)h(t) dt \right)^2 ds < \epsilon.$$

He then shows that if $K(s, t)$ is general and $f(s)$ can be represented in the form $f(s) = \int_0^1 K(s, t)h(t) dt$, then

$$f(s) = c_1\psi^{(1)}(s) + c_2\psi^{(2)}(s) + \dots ,$$

$$c_m = \int_0^1 f(s)\psi^{(m)}(s) ds,$$

and the series converges absolutely and uniformly. Schmidt will show later that the condition that the kernel be general is unnecessary for this result.

The remainder of the paper is devoted to an application and to extensions of the theory. The application is to the constrained minimization or maximization of the form

$$J(x) = \int_0^1 \int_0^1 K(s, t)x(s)x(t) ds dt$$

subject to certain constraints.

There are two extensions. The first shows that the continuity restrictions can be relaxed to permit restricted discontinuities of order less than $\frac{1}{2}$ — loosely speaking, discontinuities at which the growth of $K(s, t)$ is bounded by a multiple of $1/d^\alpha$ ($\alpha < \frac{1}{2}$), where d is the distance to the discontinuity. The technique is to patch the discontinuities on strips of width ϵ in such a way that the earlier proofs remain valid and then take the limit as $\epsilon \rightarrow 0$.

The second extension removes the restriction that the eigenvalues of $K(s, t)$ be simple. Hilbert starts with a single multiple eigenvalue, and once again his approach is to modify the kernel, this time to split the multiple eigenvalue into n distinct eigenvalues, where n is the multiplicity of the original eigenvalues. The modification is then allowed to go to zero in such a way that the modified eigenfunctions remain orthonormal. Repeating the process for each multiple eigenvalue, he obtains a full system of eigenfunctions for the original kernel.

Hilbert's contribution to the theory of integral equations was not just a matter of introducing eigenvalues and eigenfunctions. After all, they had already appeared in the study of differential equations, though not by those names. Hilbert also recognized their connection with expansions of "arbitrary" functions, a far reaching extension of Fourier series. The cost of this extension was the requirement that the kernel be symmetric. Moreover, his results are not as general as they might be — as is true of much pioneering work. His student Schmidt would have something to say about that.

4. Schmidt

Reading Erhard Schmidt after reading Fredholm and Hilbert is like walking into a familiar room through a new door. The same furniture is there in the same places; but the viewpoint is different, and one sees things in a fresh light. Schmidt's approach is to dispense with the relation between integral equations and finite-dimensional linear systems and to treat the integral equation on its own terms. The result is a striking simplification of the the Fredholm–Hilbert theory. The style of Schmidt's paper is also different. Fredholm, who distilled a great deal of technical detail into a small set of formulas, writes as if to put the results down as fast as his pen will permit. Hilbert often appears to be doing the research at the same time as the writing. His it-is-easy-to-see's are not all that easy, and one suspects he would have had some difficulty explaining them to a class a year later. Schmidt is all order and clarity, and his paper has the conciseness of someone who knows exactly what he is doing.

Schmidt's paper is divided into five chapters preceded by an introduction. Like Fredholm, Schmidt subdivides his paper into sections which are numbered consecutively regardless of chapter boundaries. Unless otherwise noted, all functions and kernels are continuous on $[a, b]$ and $[a, b] \times [a, b]$.

The first chapter consists of preliminary results. Schmidt introduces a (possibly infinite) set of functions ψ_1, ψ_2, \dots that are orthogonal and normalized in the sense that

$$\int_a^b \psi_\mu(x)\psi_\nu(x) dx = \begin{cases} 1 & \text{if } \mu = \nu, \\ 0 & \text{if } \mu \neq \nu. \end{cases}$$

(Following current terminology we will call such a sequence orthonormal.) He states

Bessel's identity

$$\int_a^b (f(x) - \sum_{\nu=1}^{\nu=n} \psi_\nu(x) \int_a^b f(y)\psi_\nu(y) dy)^2 = \int_a^b (f(x))^2 dx - \sum_{\nu=1}^{\nu=n} (\int_a^b f(y)\psi_\nu(y) dy)^2,$$

and the resulting inequality

$$\sum_{\nu=1}^{\nu=n} (\int_a^b f(y)\psi_\nu(y) dy)^2 \leq \int_a^b (f(x))^2 dx,$$

noting that the latter implies that the sum $\sum_{\nu=1}^{\nu=\infty} (\int_a^b f(y)\psi_\nu(y) dy)^2$ converges. Finally he uses Bessel's inequality to derive the Schwarz inequality $(\int_a^b f(x)\varphi(x) dx)^2 \leq \int_a^b (f(x))^2 dx \cdot \int_a^b (\varphi(x))^2 dx$. This inequality is also associated with the names Cauchy and Bunyakovsky.

In §2, Schmidt establishes a convergence result that will be used later in proving the convergence of the expansion of a function in terms of eigenfunctions. Suppose $Q(z, x)$ is a function that is integrable with respect to x and such that for some constant A we have $\int_a^b (Q(z, x))^2 dx \leq A$, for all $z \in [a, b]$. Then if the continuous functions $\psi_\nu(x)$ are orthonormal, $f(x)$ is integrable, and $\int_a^b (f(x))^2$ is finite, then the series

$$\sum_{\nu=1}^{\nu=\infty} \int_a^b f(y)\psi_\nu(y) dy \cdot \int_a^b Q(z, x)\psi_\nu(x) dx = \sum_{\nu=1}^{\nu=\infty} U_\nu(z) \tag{4.1}$$

converges absolutely and uniformly for $z \in [a, b]$.

This result is remarkable for its generality, requiring, instead of continuity, only the integrability $Q(z, x)$ with respect to x and the integrability of $f(x)$.⁹

In §3, Schmidt turns to the orthogonalization of sequences of independent functions. The method is constructive, and Schmidt proves that the method breaks down at step m if and only if the first m functions are linearly dependent.

The method has a discrete analogue for vectors, and both methods are called the Gram–Schmidt algorithm. It has an interesting history. Formulas that were equivalent to the discrete Gram–Schmidt algorithm were first given by Laplace [14] in a supplement to his *Théorie Analytique des Probabilités*. However, Laplace, who was concerned with estimating the variability of least squares parameters, had no notion of orthogonality, and, in fact, discarded the orthogonalized vectors as soon as they were computed. J. P. Gram [11, 1883] certainly had the general concept of orthogonal functions and their relation to least squares approximation, but he did not use the formulas of the Gram–Schmidt algorithm. Instead he gave determinantal expressions for his orthogonal

⁹However, the proof uses the Schwarz inequality, which Schmidt has stated only for continuous $f(x)$.

functions. Schmidt, as we have seen, knew what orthogonality was and produced the orthogonalization formulas that we now call the Gram–Schmidt algorithm. In §3, he makes a gracious nod to Gram, saying, “In essence, J. P. Gram presented the same formulas . . .” Presumably he meant that each of their formulas produced the same polynomials. One might sum up by saying: Laplace had the formulas but no notion of orthogonality; Gram had orthogonality but not the formulas; and Schmidt had both.

In Chapter II, entitled *On the Linear Symmetric Integral Equation*, Schmidt begins with the homogeneous equation

$$\varphi(s) = \lambda \int_a^b K(s, t)\varphi(t) dt, \quad (4.2)$$

where the kernel $K(s, t)$ is symmetric. This reverses the order in Fredholm and Hilbert, who begin with the inhomogeneous equation, and emphasizes the kernel itself rather than the integral equation. Schmidt adopts Hilbert’s nomenclature for nontrivial solutions of (4.2), calling $\varphi(s)$ an eigenfunction of $K(s, t)$ corresponding to the eigenvalue λ . He then shows that two eigenfunctions corresponding to two distinct eigenvalues are orthogonal and that all eigenvalues must be real. Finally, he shows that for any eigenvalue there is a corresponding real eigenfunction.

In §5, he treats the system of eigenvalues and eigenfunctions as a whole. He defines the multiplicity of an eigenvalue as the number of linearly independent eigenfunctions corresponding to it. Using Bessel’s inequality, he shows that if λ is an eigenvalue of multiplicity n then

$$n \leq \lambda^2 \int_a^b \int_a^b (K(s, t))^2 ds dt.$$

Hence any particular eigenvalue has finite multiplicity.

Schmidt now defines a full orthonormal system of a kernel as a set of orthonormal functions $\varphi_1(s), \varphi_2(s), \dots$ such that any eigenfunction can be represented as a finite linear combination of elements of the set. Such a system may be obtained by orthonormalizing a set of linearly independent eigenfunctions corresponding to each eigenvalue. He then shows that

$$\int_a^b \int_a^b (K(s, t))^2 ds dt \geq \sum_{\rho} \frac{1}{\lambda_{\rho}^2},$$

where ρ ranges over a finite set of indices. This immediately implies that if $K(s, t)$ has an infinite number of eigenvalues, then their magnitudes must be unbounded.

In §6 Schmidt introduces the iterated kernels defined by

$$\begin{aligned} K^1(s, t) &= K(s, t), \\ K^{i+1}(s, t) &= \int_a^b K(s, r)K^i(r, t) dr, \quad i = 1, 2, \dots \end{aligned}$$

He shows that the iterated kernels are nonzero and that if λ is an eigenvalue of $K(s, t)$ then λ^i is an eigenvalue of $K^i(s, t)$. Moreover, he shows that any eigenvalue of $K^i(s, t)$ must be the i th power of an eigenvalue of $K(s, t)$ and that a full orthonormal system for $K(s, t)$ is a full orthonormal system for $K^i(s, t)$. It should be noted that the converse of this last statement is not true, even in finite dimensions, as can be shown by considering the matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. However, he does show that any eigenfunction of $K^i(s, t)$ is a linear combination of no more than two eigenfunctions of $K^1(s, t)$.

Recall that the iterated kernels were introduced by Fredholm as terms in a Neumann series for the resolvent kernel. Schmidt has two entirely different reasons for introducing them. First $K^4(s, t)$ is particularly well behaved, and second iterated kernels can be used to prove the existence of eigenfunctions for any nonzero kernel. Schmidt announces this result in a single statement in §7—each kernel $K(s, t)$ that does not vanish identically has at least one eigenfunction—but the proof is deferred to §11.

The expansion results in §§8–9 are the core of this chapter. Throughout, the functions $\varphi_1(s), \varphi_2(s), \dots$ form a full orthonormal system for the continuous kernel $K(s, t)$, and $\lambda_1, \lambda_2, \dots$ are the corresponding eigenvalues. Schmidt’s object is to give conditions on a function $g(s)$ under which

$$g(s) = \sum_{\nu} \varphi_{\nu}(s) \int_a^b g(t) \varphi_{\nu}(t) dt. \tag{4.3}$$

Because $g(s)$ is to be expressed as a linear combination of the eigenfunctions of $K(s, t)$, a good guess for a condition is that there is a function $p(t)$ such that

$$g(s) = \int_a^b K(s, t) p(t) dt. \tag{4.4}$$

Schmidt first proves that if the series

$$\sum_{\nu} \frac{\varphi_{\nu}(s) \varphi_{\nu}(t)}{\lambda_{\nu}} \tag{4.5}$$

converges uniformly then it converges to $K(s, t)$. Thus the uniform convergence of (4.5) implies a spectral representation of $K(s, t)$, which, it turns out, can be used to establish (4.3) under the condition (4.4). Unfortunately, the desired uniform convergence need not occur, and Schmidt is forced to take a detour.

The detour consists in proving that

$$K^4(s, t) = \sum_{\nu} \frac{\varphi_{\nu}(s) \varphi_{\nu}(t)}{\lambda_{\nu}^4}$$

and that the series is absolutely and uniformly convergent. He then uses this result to prove that for any continuous function $h(s)$ if

$$\int_a^b K(s, t)h(t) dt = 0 \quad (4.6)$$

then

$$\int_a^b h(s)\varphi_\nu(s) ds = 0, \quad \nu = 1, 2, \dots \quad (4.7)$$

and conversely.

Schmidt then proves the main expansion theorem.

Let the continuous function $g(s)$ be represented by the equation

$$g(s) = \int_a^b K(s, t)p(t) dt,$$

where $p(t)$ is a continuous function. Then

$$\begin{aligned} g(s) &= \sum_\nu \varphi_\nu(s) \int_a^b g(t)\varphi_\nu(t) dt = \sum_\nu \frac{\varphi_\nu(s)}{\lambda_\nu} \int_a^b p(t)\varphi_\nu(t) dt \\ &= \sum_\nu \int_a^b K(s, t)\varphi_\nu(t) dt \int_a^b p(t)\varphi_\nu(t) dt, \end{aligned} \quad (4.8)$$

and the series on the right converges absolutely and uniformly.

In proving this theorem, Schmidt points out that the results of §2 [see (4.1)] combined with the third series in (4.8) imply that the series converges absolutely and uniformly. It remains only to prove that it converges to $g(s)$, which he does by using the equivalence of (4.6) and (4.7).

Finally, in a very short paragraph, Schmidt derives Hilbert's expansion of a quadratic form

$$\int_a^b \int_a^b K(s, t)q(s)p(t) ds dt = \sum_\nu \frac{1}{\lambda_\nu} \int_a^b q(s)\varphi_\nu(s) ds \int_a^b p(t)\varphi_\nu(t) dt. \quad (4.9)$$

There is a significant reversal in the direction of proof between Hilbert and Schmidt. Hilbert proves (4.9) as a limiting case of the corresponding theorem for quadratic forms and then proceeds to prove, at some length, the expansion theorem (4.8). In the process he must assume that the kernel $K(s, t)$ is "general" in the sense that any continuous function $x(s)$ can be approximated to arbitrary accuracy in the form $\int_a^b K(s, t)h(t) dt$. This is a very restrictive condition. It leaves out, for example, any kernel with only a finite number of eigenvalues. Schmidt, on the other hand, proves the expansion theorem

first, without the assumption of generality, and then derives Hilbert's expansion of a quadratic form as an obvious corollary.

Almost as an afterthought, Schmidt turns in §10 to the inhomogeneous integral equation and derives, for symmetric kernels, Fredholm's results.

Section 11 is essentially an appendix in which the existence of eigenfunctions of an operator is established. To exhibit the idea behind the proof, assume that we have the expansion

$$K^{2n}(s, t) = \sum_{\nu} \frac{\varphi_{\nu}(s)\varphi_{\nu}(t)}{\lambda_{\nu}^{2n}}$$

and suppose that $\lambda_1^{2n} = \dots = \lambda_m^{2n}$ are the smallest eigenvalues in the above sum. Multiplying by λ_1^{2n} , we can write

$$\lambda_1^{2n} K^{2n}(s, t) = \sum_{\nu=1}^m \varphi_{\nu}(s)\varphi_{\nu}(t) + \sum_{\nu=n+1} \frac{\varphi_{\nu}(s)\varphi_{\nu}(t)}{(\lambda_{\nu}/\lambda_1)^{2n}}.$$

Since $|\lambda_{\nu}/\lambda_1| > 1$, the second sum in the above equation goes to zero as $n \rightarrow \infty$ and hence we can write

$$\lim_{n \rightarrow \infty} \lambda_1^{2n} K^{2n}(s, t) = \sum_{\nu=1}^m \varphi_{\nu}(s)\varphi_{\nu}(t) \equiv u(s, t).$$

If we now chose t_1 so that $\varphi(s) = u(s, t_1)$ is nonzero, then $\varphi(s)$ is an eigenfunction of $K^2(s, t)$, which by the results of §6 implies that $K(s, t)$ has an eigenfunction.

Note that the above argument requires that all the eigenvalues of smallest magnitude have the same sign, which is why one must start with $K^2(s, t)$ rather than $K(s, t)$. The above argument is, of course, not a proof since it assumes what Schmidt is trying to prove: the existence of eigenfunctions. But Schmidt treats the kernels $K^{2n}(s, t)$ directly to arrive at the function $u(s, t)$. In the introduction to his paper, Schmidt says that his proof is "based on a famous proof of H. A. Schwarz, which in the language of Fredholm's formula amounts to solving the equation $\delta(\lambda) = 0$ by Bernoulli's method." Today we call this the power method.

Finally in §12, Schmidt, like Fredholm and Hilbert before him, relaxes the continuity conditions on which his results were grounded. Specifically, he shows that his results remain intact if the kernel satisfies two conditions.

I. The point set in the s, t -plane consisting of the points of discontinuity of $K(s, t)$ (and which is therefore closed) has outer content zero on any line $s = \text{const}$.

II. $\int_a^b (K(s, t))^2 dt$ is defined and finite for $a \leq s \leq b$ and represents a continuous function of s .

This is far more general than the conditions of Fredholm and Hilbert, and the use of outer content points to the nascent theory of measure and integration.

In Chapter III, Schmidt turns his attention to unsymmetric kernels. He begins in §13 with the nonhomogeneous equation

$$f(s) = \varphi(s) - \int_a^b K(s, t)\varphi(t) dt, \quad (4.10)$$

with the goal of reproducing Fredholm's theory. As we have seen in the discussion of Fredholm's paper, a key problem is to compute bases for the left and right null spaces of the right hand side of (4.10). Schmidt shows that they are eigenspaces of the symmetric kernels $K(s, t) + K(t, s) + \int_a^b K(s, r)K(t, r) dr$ and $K(s, t) + K(t, s) + \int_a^b K(r, s)K(r, t) dr$. (However, he does not show that the dimensions of these spaces are the same.) He then goes on to prove Fredholm's theorems by reducing the solution to that of a symmetric kernel.

Returning to eigensystems, in §14 Schmidt defines two functions $\varphi(s)$ and $\psi(s)$ to be an adjoint pair of eigenfunctions of K if they satisfy the relations

$$\varphi(s) = \lambda \int_a^b K(s, t)\psi(t) dt$$

and

$$\psi(s) = \lambda \int_a^b K(t, s)\varphi(t) dt.$$

The value λ is the corresponding eigenvalue. He shows that $\varphi(s)$ and $\psi(s)$ are eigenfunctions of the symmetric kernels

$$\overline{K}(s, t) = \int_a^b K(s, r)K(t, r) dr \quad \text{and} \quad \underline{K}(s, t) = \int_a^b K(r, s)K(r, t) dr$$

corresponding to the eigenvalue λ^2 . He goes on to show that $\varphi(s)$ and $\psi(s)$ can be taken to be real valued functions and λ to be taken to be positive. He then shows that "The adjoint functions of a full normalized orthogonal system of the kernel $\overline{K}(s, t)$ form a full normalized orthogonal system of the kernel $\underline{K}(s, t)$, and conversely."

In §16, Schmidt establishes a generalization of his expansion theorem for symmetric kernels. Specifically, if $g(s) = \int_a^b K(t, s)h(t) dt$ for some continuous function $h(t)$, then

$$\begin{aligned} g(s) &= \sum_{\nu} \psi_{\nu}(s) \int_a^b g(t)\psi_{\nu}(t) dt \\ &= \sum_{\nu} \frac{\psi_{\nu}(s)}{\lambda_{\nu}} \int_a^b h(t)\varphi_{\nu}(t) dt \\ &= \sum_{\nu} \int_a^b K(t, s)\varphi_{\nu}(t) dt \int_a^b h(t)\varphi_{\nu}(t) dt. \end{aligned}$$

He establishes a similar theorem involving the eigenfunctions $\psi_\nu(s)$ for a function $g(s)$ that can be represented in the form $g(s) = \int_a^b K(t, s)h(t) dt$. Finally, he shows that

$$\int_a^b K(s, t)q(t) dt = \sum \frac{\varphi_\nu(s)}{\lambda_\nu} \int_a^b q(t)\psi_\nu(t) dt,$$

which “corresponds to the canonical decomposition of a bilinear form.”

This development is a magnificent tour de force, marred only by an unfortunate choice of nomenclature. For it is clear that Schmidt’s construction amounts to a generalization of the singular value decomposition of Beltrami [2, 1873] and Jordan [13, 1874], as the quote in the last paragraph shows. Today eigenvalues are a part of the spectrum of an operator, self-adjoint or not. The term singular value was first used by Bateman [1, 1908] for the reciprocals of Schmidt’s eigenvalues. The reader is referred to [22], where these matters are treated in detail.

The climax of the paper is Schmidt’s approximation theorem, which is established in Chapter IV. Let $\varphi_i(s)$ and $\psi_i(s)$ ($i = 1, 2, \dots$) be a full adjoint system of orthonormal functions for $K(s, t)$. Consider an approximation to $K(s, t)$ of the form

$$\sum_{i=1}^m p_i(s)q_i(t)$$

and let

$$M_m = \int_a^b \int_a^b \left(K(s, t) - \sum_{i=1}^m p_i(s)q_i(t) \right)^2 ds dt.$$

Then M_m is minimized by the approximation

$$\sum_{\nu=1}^{\nu=m} \frac{\varphi_\nu(s)\psi_\nu(t)}{\lambda_\nu}. \tag{4.11}$$

The proof of this theorem is tricky, and most people (your translator included) get it wrong the first time around. The reader is referred to [22] for a paraphrase of Schmidt’s proof in modern terminology.

In §19, Schmidt shows that the quantity M_m , evaluated at (4.11), approaches zero as m approaches ∞ , an early example of L_2 convergence of a sequence of operators.

Schmidt’s approximation theorem in the finite-dimensional case was rediscovered by C. Eckart and G. Young in 1936 [6] and is sometimes erroneously called the Eckart–Young theorem.

We will pass by the last chapter, which is devoted to expansion theorems that are unrelated the material on integral equations.

With Schmidt's paper the subject of integral equations came of age. To this day, the results of Fredholm, Hilbert, and Schmidt occupy a large part of textbooks on integral equations. There was, of course, much more to come. Just as integral equations had a large role in the development of functional analysis so did functional analysis enhance the further development of integral equations. For example, continuous kernels became compact operators. Hilbert and Schmidt themselves were part of this development. While Schmidt was transcribing his dissertation, Hilbert launched an investigation of the theory of infinite systems of linear equations that was motivated by applications to integral equations. Schmidt picked up Hilbert's work and developed many of the fundamental concepts of what are now known as (ℓ_2) Hilbert spaces. But that is another story.

References

- [1] H. Bateman. A formula for the solving function of a certain integral equation of the second kind. *Messenger of Mathematics*, Series 2, 37:179–187, 1908.
- [2] E. Beltrami. Sulle funzioni bilineari. *Giornale di Matematiche ad Uso degli Studenti Delle Universita*, 11:98–106, 1873. An English translation by D. Boley is available in Technical Report 90–37, Department of Computer Science, University of Minnesota, Minneapolis, 1990.
- [3] M. Bernkopf. The development of function spaces with particular reference to their origins in integral equation theory. *Archive for History of Exact Sciences*, 3:1–96, 1966.
- [4] G. Birkhoff, editor. *A Source Book in Classical Analysis*. Harvard University Press, Cambridge, Massachusetts, 1973. Uta Merzback Assistant.
- [5] R. Courant and D. Hilbert. *Methoden der mathematischen Physik*. Springer-Verlag, Berlin, second edition, 1931.
- [6] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936.
- [7] I. Fredholm. Sur une classe d'équations fonctionnelles. *Acta Mathematica*, 27:365–390, 1903.
- [8] I. Fredholm. Les équations intégrales linéaires. In *Compte rendu du Congrès des mathématiciens tenu à Stockholm 1909, (1910)*, pages 92–100, 1909. Also in [9, pp. 123–131].

- [9] I. Fredholm. *Oevres Complètes de Ivar Fredholm*. Publiées sous les auspices de le Kungliga Svenska Vetenskapsakademien par l'Institute Mittag-Leffler, Malmö, 1955.
- [10] C. C. Gillispie. *Dictionary of Scientific Biography*. Charles Scribner's Sons, New York, 1970.
- [11] J. P. Gram. Über die Entwicklung reeller Functionen in Reihen mittelst der Methode der kleinsten Quadrate. *Journal für die reine und angewandte Mathematik*, 94:41–73, 1883.
- [12] D. Hilbert. Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. *Nachrichten von der Königl. Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 1904:49–91, 1904.
- [13] C. Jordan. Mémoire sur les formes bilinéaires. *Journal de Mathématiques Pures et Appliquées, Deuxième Série*, 19:35–54, 1874.
- [14] P. S. Laplace. *Théorie Analytique des Probabilités*. Courcier, Paris, third edition, 1820.
- [15] W. V. Lovitt. *Linear Integral Equations*. McGraw–Hill, New York, 1924. Dover edition published in 1950,.
- [16] M. Z. Nashed, editor. *Generalized Inverses and Applications*. Academic Press, New York, 1976.
- [17] J. J. O'Connor and E. F. Robertson. The MacTutor History of Mathematics Archive. www-history.mcs.st-andrews.ac.uk/index.html.
- [18] D. Porter and D. S. G. Stirling. *Integral Equations*. Cambridge University Press, Cambridge, 1990.
- [19] Constance Reid. *Hilbert*. Springer Verlag, Berlin, 1970.
- [20] E. Schmidt. Zur Theorie der linearen und nichtlinearen Integralgleichungen. I Teil. Entwicklung willkürlichen Funktionen nach System vorgeschriebener. *Mathematische Annalen*, 63:433–476, 1907.
- [21] F. Smithies. *Integral Equations*. Cambridge University Press, Cambridge, England, 1958.
- [22] G. W. Stewart. On the early history of the singular value decomposition. *SIAM Review*, 35:551–566, 1993.
- [23] F. G. Tricomi. *Integral Equations*. Interscience, New York, 1957.

- [24] H. von Koch. Sur les déterminants infinis et les équations différentielles linéaires. *Acta Mathematica*, 16:217–295, 1892/1893.
- [25] Hermann Weyl. David hilbert: 1862–1943. *Obituary Notices of Fellows of th Royal Society*, 4:547–553, 1944.

ON A CLASS OF FUNCTIONAL EQUATIONS

BY

IVAR FREDHOLM
STOCKHOLM

In certain works¹ ABEL was concerned with the problem of determining a function $\varphi(x)$ in such a way that it satisfies the functional equation

$$\int f(x, y)\varphi(y) dy = \psi(x), \quad (\text{a})$$

where $f(x, y)$ and $\psi(x)$ are given functions. ABEL solved some special cases of this functional equation, and he seems to be the first to have recongnized its importance. For this reason I propose to call the functional equation (a) an *Abelian functional equation*.

In this note I will be chiefly concerned not with the Abelian equation but with the functional equation

$$\varphi(x) + \int_0^1 f(x, y)\varphi(y) dy = \psi(x), \quad (\text{b})$$

which is closely related to the Abelian equation.

In fact, if in place of $f(x, y)$ and $\varphi(x)$ we substitute $\frac{1}{\lambda}f(x, y)$ and $\lambda\varphi(x)$, then equation (b) becomes

$$\lambda\varphi(x) + \int_0^1 f(x, y)\varphi(y) dy = \psi(x), \quad (\text{c})$$

which can be transformed into equation (a) by setting $\lambda = 0$. Thus the solution of (a) can be regarded as implicitly subsumed by the solution of (b).

Regarding equation (b), it seems to me that it deserves the special attention of mathematicians because most of the problems of mathematical physics that lead to

¹*Magazin for Naturvidenskaberne, Kristiania 1823 and Oeuvres complètes*

linear differential equations can be transformed into functional equations of the form (b) or of the form

$$\begin{aligned} \varphi(x_1 \dots x_n) + \int \dots \int f(x_1 \dots x_n, \xi_1 \dots \xi_n) \varphi(\xi_1 \dots \xi_n) d\xi_1 \dots d\xi_n \\ = \psi(x_1 \dots x_n). \end{aligned}$$

To see this one only has to recall the Dirichlet problem in which one tries to represent the unknown potential between two plates in terms of the potential on the plates, as well as analogous problems in the theory of magnetism and the theory of elasticity.

The first attempt to solve an instance of equation (b) was made by NEUMANN. Specifically, NEUMANN's famous method for the solution of the *Dirichlet* problem consisted in the expansion of $\varphi(x)$ in increasing powers of the parameter $\frac{1}{\lambda}$. But NEUMANN's expansion, which always converges in the case of DIRICHLET's problem, may not converge in the general case.

In an important work² M. VOLTERRA successfully applied the method of NEUMANN to the functional equation

$$\varphi(x) + \int_0^x f(x, y) \varphi(y) dy = \psi(x). \quad (d)$$

In the same work M. VOLTERRA also made clear the close connection of equation (d) with the Abelian equation

$$\int_0^x f(x, y) \varphi(y) dy = \psi(x).$$

The equation I propose to study in the present work includes M. VOLTERRA's equation as a special case. For one immediately obtains equation (c) by supposing that the function $f(x, y)$ in (b) is zero for $y > x$.

In what follows the function $f(x, y)$ will be subject to a certain condition. I will assume that $f(x, y)$ is such that for α less than one the function $(x - y)^\alpha f(x, y)$ is bounded and integrable. Thus I am not going to treat equation (b) in full generality. But the applications of equation (b) to mathematical physics justify the condition I have imposed on the function. I will return to these applications in another paper.

§ 1. On the construction and properties of the determinant of the fundamental functional equation.

1. Suppose that $f(x, y)$ is a finite function that is integrable with respect to the variables x and y , either individually or together. For definiteness we will suppose that x and y are positive and less than one.

² *Annali di Matematica*, 1886

have a limit superior. Then obviously we can write

$$\left| f \begin{pmatrix} x_1 \dots x_n \\ x_1 \dots x_n \end{pmatrix} \right| \leq \sqrt{n^n} A^n (x_1 - x_2)^\alpha (x_2 - x_3)^\alpha \dots (x_{n-1} - x_n)^\alpha.$$

But in determining the maximum it is clearly sufficient to consider variables that satisfy the condition

$$x_1 > x_2 > \dots > x_n,$$

since the left-hand side is a symmetric function of the variables x_1, \dots, x_n .

In this case the maximum value of the product

$$(x_1 - x_2)(x_2 - x_3) \dots (x_{n-1} - x_n)$$

is equal to

$$\frac{1}{n^n}.$$

Consequently,

$$\frac{1}{n!} \int_0^1 \dots \int_0^1 f \begin{pmatrix} x_1 \dots x_n \\ x_1 \dots x_n \end{pmatrix} dx_1 dx_2 \dots dx_n < \frac{(n^n)^{\frac{1}{2}-\alpha}}{n!} A^n.$$

4. In the same way that we demonstrated the validity of the expression for D_f , we can demonstrate the validity of the following expressions, which I will call the minors of D_f :

$$\begin{aligned} & D_f \begin{pmatrix} \xi_1, & \xi_2, & \dots, & \xi_n \\ \eta_1, & \eta_2, & \dots, & \eta_n \end{pmatrix} \\ &= f \begin{pmatrix} \xi_1, & \xi_2, & \dots, & \xi_n \\ \eta_1, & \eta_2, & \dots, & \eta_n \end{pmatrix} + \int_0^1 f \begin{pmatrix} \xi_1 \dots \xi_n, & x \\ \eta_1 \dots \eta_n, & x \end{pmatrix} dx \\ & \quad + \frac{1}{2} \int_0^1 \int_0^1 f \begin{pmatrix} \xi_1 \dots \xi_n, & x_1, & x_2 \\ \eta_1 \dots \eta_n, & x_1, & x_2 \end{pmatrix} dx_1 dx_2 + \dots \\ &= \sum_{\nu=0}^{\infty} \frac{1}{\nu!} \int_0^1 \dots \int_0^1 f \begin{pmatrix} \xi_1 \dots \xi_n, & x_1 \dots x_\nu \\ \eta_1 \dots \eta_n, & x_1 \dots x_\nu \end{pmatrix} dx_1 \dots dx_\nu. \end{aligned} \tag{3}$$

5. The minors satisfy some important relations, which we will now derive. On expanding the determinant

$$f \begin{pmatrix} \xi_1, & \xi_2, & \dots, & \xi_n, & x_1 \dots x_\nu \\ \eta_1, & \eta_2, & \dots, & \eta_n, & x_1 \dots x_\nu \end{pmatrix}$$

along the elements of its first row we find that

$$\begin{aligned}
& f \begin{pmatrix} \xi_1, & \xi_2, & \dots, & \xi_n, & x_1 \dots x_\nu \\ \eta_1, & \eta_2, & \dots, & \eta_n, & x_1 \dots x_\nu \end{pmatrix} \\
&= f(\xi_1, \eta_1) f \begin{pmatrix} \xi_2 \dots \xi_n, & x_1 \dots x_\nu \\ \eta_2 \dots \eta_n, & x_1 \dots x_\nu \end{pmatrix} \\
&- f(\xi_1, \eta_2) f \begin{pmatrix} \xi_2, & \xi_3 \dots \xi_n, & x_1 \dots x_\nu \\ \eta_1, & \eta_3 \dots \eta_n, & x_1 \dots x_\nu \end{pmatrix} + \dots \\
&- (-1)^n f(\xi_1, \eta_n) f \begin{pmatrix} \xi_2 \dots \xi_n, & x_1 \dots x_\nu \\ \eta_1 \dots \eta_{n-1}, & x_1 \dots x_\nu \end{pmatrix} \\
&+ (-1)^n f(\xi_1, x_1) f \begin{pmatrix} \xi_2 \dots \xi_n, & x_1 \dots x_\nu \\ \eta_1 \dots \eta_{n-1}, & \eta_n \dots x_\nu \end{pmatrix} - \dots \\
&- (-1)^{n+\nu} f(\xi_1, x_\nu) f \begin{pmatrix} \xi_2 \dots \xi_n, & x_1 \dots x_\nu \\ \eta_1 \dots \eta_{n-1}, & \eta_n \dots x_{\nu-1} \end{pmatrix}.
\end{aligned}$$

If we multiply the two sides of this identity by $dx_1 \dots dx_\nu$ and integrate between 0 and 1, we get the formula

$$\begin{aligned}
& \int_0^1 \dots \int_0^1 f \begin{pmatrix} \xi_1 \dots \xi_n, & x_1 \dots x_\nu \\ \eta_1 \dots \eta_n, & x_1 \dots x_\nu \end{pmatrix} dx_1 \dots dx_\nu \\
&= f(\xi_1, \eta_1) \int_0^1 \dots \int_0^1 f \begin{pmatrix} \xi_2 \dots \xi_n, & x_1 \dots x_\nu \\ \eta_2 \dots \eta_n, & x_1 \dots x_\nu \end{pmatrix} dx_1 \dots dx_\nu \\
&- f(\xi_1, \eta_2) \int_0^1 \dots \int_0^1 f \begin{pmatrix} \xi_2, & \xi_3 \dots \xi_n, & x_1 \dots x_\nu \\ \eta_1, & \eta_3 \dots \eta_n, & x_1 \dots x_\nu \end{pmatrix} dx_1 \dots dx_\nu + \dots \\
&- \nu \int_0^1 \dots \int_0^1 f(\xi_1, \tau) f \begin{pmatrix} \tau, & \xi_2 \dots \xi_n, & x_2 \dots x_{\nu-1} \\ \eta_1, & \eta_2 \dots \eta_n, & x_1 \dots x_{\nu-1} \end{pmatrix} d\tau dx_1 \dots dx_{\nu-1}.
\end{aligned}$$

If we now multiply by $\frac{1}{\nu!}$ and take the sum for $\nu = 0$ to $\nu = \infty$, we obtain the very important formula

$$\begin{aligned}
& D_f \begin{pmatrix} \xi_1 \dots \xi_n \\ \eta_1 \dots \eta_n \end{pmatrix} + \int_0^1 f(\xi_1, \tau) D_f \begin{pmatrix} \tau, & \xi_2 \dots \xi_n \\ \eta_1, & \eta_2 \dots \eta_n \end{pmatrix} d\tau \\
&= f(\xi_1, \eta_1) D_f \begin{pmatrix} \xi_2 \dots \xi_n \\ \eta_2 \dots \eta_n \end{pmatrix} - f(\xi_1, \eta_2) D_f \begin{pmatrix} \xi_2, & \xi_3 \dots \xi_n \\ \eta_1, & \eta_3 \dots \eta_n \end{pmatrix} + \dots.
\end{aligned} \tag{4}$$

In the same way, if we start by expanding the determinant by the elements of the

first column, we get the formula

$$\begin{aligned} & D_f \begin{pmatrix} \xi_1 \cdots \xi_n \\ \eta_1 \cdots \eta_n \end{pmatrix} + \int_0^1 f(\tau, \eta_1) D_f \begin{pmatrix} \xi_1, & \xi_2 \cdots \xi_n \\ \tau, & \eta_2 \cdots \eta_n \end{pmatrix} d\tau \\ &= f(\xi_1, \eta_1) D_f \begin{pmatrix} \xi_2 \cdots \xi_n \\ \eta_2 \cdots \eta_n \end{pmatrix} - f(\xi_2, \eta_1) D_f \begin{pmatrix} \xi_1, & \xi_2 \cdots \xi_n \\ \eta_2, & \eta_3 \cdots \eta_n \end{pmatrix} + \cdots \text{etc.} \end{aligned} \quad (5)$$

For the case $n = 1$ these two formulas become

$$D_f \begin{pmatrix} \xi \\ \eta \end{pmatrix} + \int_0^1 f(\xi, \tau) D_f \begin{pmatrix} \tau \\ \eta \end{pmatrix} d\tau = f(\xi, \eta) D_f, \quad (41)$$

$$D_f \begin{pmatrix} \xi \\ \eta \end{pmatrix} + \int_0^1 f(\tau, \eta) D_f \begin{pmatrix} \xi \\ \tau \end{pmatrix} d\tau = f(\xi, \eta) D_f. \quad (51)$$

6. If we replace $f(x, y)$ by $\lambda f(x, y)$ in D_f , we find that $D_{\lambda f}$ can be expanded in a power series in λ , which, because of H. HADAMARD'S lemma, converges for all values of λ . Thus $D_{\lambda f}$ is an entire function of λ .

Recalling the definition of D_f and its minors, we immediately get the relations

$$\lambda^n \frac{d^n D_{\lambda f}}{d\lambda^n} = \int_0^1 \cdots \int_0^1 D_{\lambda f} \begin{pmatrix} x_1 \cdots x_n \\ x_1 \cdots x_n \end{pmatrix} dx_1 dx_2 \cdots dx_n, \quad (6)$$

which hold for $n = 1, 2, 3, \dots$

These relations allow us to obtain an important result. Specifically, since $D_{\lambda f}$ is an entire function of λ , each root of the equation

$$D_{\lambda f} = 0$$

is necessarily of finite multiplicity.

Consequently, one cannot find a value of λ for which $D_{\lambda f}$ and all its derivatives are zero.

In particular, if for $\lambda = 1$ we have $D_{\lambda f} = D_f = 0$, then we can always find a minor of D_f of least order that is not identically zero.

§ 2. On a class of functional transformations and their inversion.

7. Let us now consider the functional equation

$$\varphi(x) + \int_0^1 f(x, s)\varphi(s) ds = \psi(x), \quad (7)$$

where $\varphi(x)$ is an unknown function and $\psi(x)$ is a finite, integrable function.

Regarding equation (7) as transforming the function $\varphi(x)$ into a new function $\psi(x)$, I will write the above equation as

$$S_f\varphi(x) = \psi(x) \quad (7)$$

and say that the transformation S_f belongs to $f(x, y)$.

*The transformations (7) form a group.** Specifically, consider another transformation S_g belonging to the function $g(x, y)$ which satisfies the same conditions of integrability etc. as $f(x, y)$.

Then it is easily seen that one can set

$$S_g\psi(x) = S_gS_f\varphi(x) = S_F\varphi(x),$$

where

$$F(x, y) = g(x, y) + f(x, y) + \int_0^1 g(x, t)f(t, y) dt.$$

Regarding the inversion of (7), there are two possible cases: D_f is nonzero or $D_f = 0$.

8. First suppose that the determinant D_f is nonzero and set

$$g(x, y) = -\frac{D_f \begin{pmatrix} x \\ y \end{pmatrix}}{D_f}.$$

Then from equation (5₁) we see that F is identically zero. Consequently, the equation

$$S_gS_f\psi(x) = \psi(x)$$

is an identity, and the transformation S_g is the inverse of S_f . Thus, if there exists a solution of equation (7), it is unique and is given by the equation

$$\varphi(x) = S_g\psi(x).$$

On the other hand, if in equation (7) we substitute $S_g\psi(x)$ in place of $\varphi(x)$, we get

$$S_f\varphi(x) = S_fS_g\psi(x) = S_F\psi(x),^\dagger$$

where by equation (4₁) F is again equal to zero.

Consequently, we can state the following theorem.

*By a *group* Fredholm means a set of operators that is closed under multiplication. The existence of inverses is not implied.

†The function F in S_F is not the same as the F defined above.

If the determinant D_f of a functional equation of the form

$$\varphi(x) + \int_0^1 f(x, s)\varphi(s) ds = \psi(x),$$

where $f(x, s)$ and $\psi(x)$ are bounded integrable functions, is different from zero, then there is a unique function $\varphi(x)$ satisfying the equation.

This function is given by the equation

$$\varphi(x) = \psi(x) - \int_0^1 \frac{D_f \begin{pmatrix} x \\ y \end{pmatrix}}{D_f} \psi(y) dy.$$

9. Let us now consider the case where D_f is zero.

We have seen that in this case there is a minor of D_f of least order that is not identically zero.

Let

$$D_f \begin{pmatrix} \xi_1 \dots \xi_n \\ \eta_1 \dots \eta_n \end{pmatrix}$$

be that minor. Because all minors of lower order are zero, the formula (4) can be written

$$D_f \begin{pmatrix} \xi_1 \dots \xi_n \\ \eta_1 \dots \eta_n \end{pmatrix} + \int_0^1 f(\xi_1, \tau) D_f \begin{pmatrix} \tau, & \xi_2 \dots \xi_n \\ \eta_1, & \eta_2 \dots \eta_n \end{pmatrix} d\tau = 0.$$

That is to say that

$$\varphi(x) = D_f \begin{pmatrix} x, & \xi_2 \dots \xi_n \\ \eta_1, & \eta_2 \dots \eta_n \end{pmatrix}$$

is a solution of the homogeneous equation

$$\varphi(x) + \int_0^1 f(x, y)\varphi(y) dy = 0. \quad (7')$$

To find all such solutions, let S_f denote the transformation belonging to f and let φ be a solution of the equation

$$S_f \varphi(x) = 0.$$

Let us call S_g the *pseudo-inverse* transformation of S_f if

$$g(x, y) = -\frac{D_f \begin{pmatrix} x, & \xi_1 \dots \xi_n \\ y, & \eta_1 \dots \eta_n \end{pmatrix}}{D_f \begin{pmatrix} \xi_1 \dots \xi_n \\ \eta_1 \dots \eta_n \end{pmatrix}},$$

where the parameters ξ and η are chosen in such a way that the denominator is nonzero—something that by hypothesis is always possible.

Then

$$S_g S_f \varphi(x) = S_F \varphi(x) = 0,$$

where

$$F(x, y) = f(x, y) + g(x, y) + \int_0^1 g(x, \tau) f(\tau, y) d\tau.$$

But by equation (5) we have

$$\begin{aligned} F(x, y) = \\ \frac{1}{D_f \begin{pmatrix} \xi_1 \dots \xi_n \\ \eta_1 \dots \eta_n \end{pmatrix}} \left[f(\xi_1, y) D_f \begin{pmatrix} x, & \xi_2 \dots \xi_n \\ \eta_1, & \eta_2 \dots \eta_n \end{pmatrix} - f(\xi_2, y) D_f \begin{pmatrix} \xi_1, & x, & \xi_3 \dots \xi_n \\ \eta_1, & \eta_2, & \eta_3 \dots \eta_n \end{pmatrix} - \right. \\ \left. \dots - (-1)^n f(\xi_n, y) D_f \begin{pmatrix} \xi_1 \dots x \\ \eta_1 \dots \eta_n \end{pmatrix} \right]. \end{aligned} \quad (9)$$

Alternatively, using an abbreviated notation, we have

$$F(x, y) = - \sum_{\nu=1}^n f(\xi_\nu, y) \Phi_\nu(x). \quad (10)$$

Now $\varphi(x)$ satisfies the equation

$$S_F \varphi(x) = 0,$$

and hence we have

$$\begin{aligned} \varphi(x) = - \int_0^1 F(x, y) \varphi(y) dy &= \sum_{\nu=1}^n \Phi_\nu(x) \int_0^1 f(\xi_\nu, y) \varphi(y) dy \\ &= \sum_{\nu=1}^n A_\nu \Phi_\nu(x). \end{aligned} \quad (11)$$

It is readily verified that this expression satisfies the equation

$$S_f \varphi(x) = 0$$

whatever the coefficients A_ν .

The functions Φ_1, \dots, Φ_n are linearly independent. For equation (4) shows that

$$\int_0^1 f(\xi_\lambda, x) \Phi_\mu(x) dx = \begin{cases} 0 & \text{if } \lambda \neq \mu, \\ 1 & \text{if } \lambda = \mu. \end{cases}$$

Hence the hypothesis that there is a linear relation

$$a_1\Phi_1 + a_2\Phi_2 + \cdots + a_n\Phi_n = 0$$

among the Φ_ν leads to the contradiction

$$\int_0^1 \sum_{\nu=1}^n a_\nu f(\xi_\nu, x) \cdot \sum_{\nu=1}^n a_\nu \Phi_\nu(x) dx = \sum a_\nu^2 = 0.$$

Thus, not only are the functions Φ_ν linearly independent, but so are the functions $f(\xi_\nu, x)$. We may sum up the results we have obtained in the following theorem.

A necessary and sufficient condition that there exists a nonzero solution of the equation

$$S_f\varphi(x) = 0$$

is that $D_f = 0$. If n is the order of the first minor of D_f that is different from zero, the above equation has n linearly independent solutions.

Let us now look for conditions for the existence of a solution of the equation

$$S_f\varphi(x) = \psi(x),$$

under the hypothesis that $D_f = 0$ and the minors of order less than n are zero.

First we must establish a certain formula. Since the function

$$\alpha(x) = D_f \begin{pmatrix} x, & a_2 \dots a_n \\ b_1, & b_2 \dots b_n \end{pmatrix}$$

satisfies the equation

$$S_f\alpha(x) = 0,$$

$\alpha(x)$ is a linear combination of the functions Φ_ν . Recalling that $\alpha(x)$ also satisfies the equation

$$S_F\alpha(x) = 0$$

or alternatively the equation

$$\alpha(x) = - \int_0^1 F(x, y)\alpha(y) dy,$$

we immediately obtain the following expression for $\alpha(x)$:

$$\alpha(x) = - \sum_{\nu=1}^n \alpha(\xi_\nu)\Phi_\nu(x). \tag{12}$$

Proceeding analogously with the function

$$\beta(x) = D_f \begin{pmatrix} a_1, & a_2 \dots a_n \\ x, & b_2 \dots b_n \end{pmatrix}$$

we get the expression

$$\beta(x) = - \sum_{\nu=1}^n \beta(\xi_\nu) \Psi_\nu(x), \quad (13)$$

where for brevity we have set

$$\Psi_1(x) = - \frac{D_f \begin{pmatrix} \xi_1, & \xi_1 \dots \xi_n \\ x, & \eta_2 \dots \eta_n \end{pmatrix}}{D_f \begin{pmatrix} \xi_1 \dots \xi_n \\ \eta_1 \dots \eta_n \end{pmatrix}},$$

and so on. It can be shown that the n functions Ψ are linearly independent.

Let us now return to the equation in question and integrate it after multiplying by

$$D_f \begin{pmatrix} a_1, & a_2 \dots a_n \\ x, & b_2 \dots b_n \end{pmatrix} dx.$$

Then we find that

$$\begin{aligned} \int_0^1 \varphi(x) D_f \begin{pmatrix} a_1, & a_2 \dots a_n \\ x, & b_2 \dots b_n \end{pmatrix} dx + \int_0^1 \int_0^1 \varphi(y) f(x, y) D_f \begin{pmatrix} a_1, & a_2 \dots a_n \\ x, & b_2 \dots b_n \end{pmatrix} dx dy \\ = \int_0^1 \psi(x) D_f \begin{pmatrix} a_1, & a_2 \dots a_n \\ x, & b_2 \dots b_n \end{pmatrix} dx. \end{aligned}$$

But by equation (5),* we see that the left-hand side is zero whatever the function $\varphi(x)$.

Hence $\psi(x)$ must satisfy the relation

$$\int_0^1 \psi(x) D_f \begin{pmatrix} a_1, & a_2 \dots a_n \\ x, & b_2 \dots b_n \end{pmatrix} dx = 0, \quad (15)$$

whatever the parameters a and b . The number of these constraints appears to be infinite, but because of equation (13) this number reduces to n ; namely, the n equations

$$\int_0^1 \psi(x) \Psi_\nu(x) dx = 0 \quad (\nu = 1, \dots, n). \quad (15')$$

*Equation (4) in the original, which appears to be a mistake.

Let us assume that these constraints are satisfied and see if, in this case, a solution of equation (7) exists.

If, to this end, we apply the transformation S_g to the two sides of equation (7), we have

$$S_g S_f \varphi(x) = S_F \varphi(x) = S_g \psi(x).$$

But

$$S_F \varphi(x) = \varphi(x) - \sum_{\nu=1}^n A_\nu \Phi_\nu(x).$$

Thus

$$\varphi(x) = S_g \psi(x) + \sum_{\nu=1}^n A_\nu \Phi_\nu(x).$$

Let us now see if the above value satisfies equation (7). To do this it is sufficient to see if $\varphi(x) = S_g \psi(x)$ satisfies equation (7), since the second term is a solution of the homogeneous equation and can be discarded. We have

$$S_f \varphi(x) = S_f S_g \psi(x) = S_G \psi(x),$$

where, by equation (4) and the definition of the functions Ψ_ν , we have

$$G(x, y) = - \sum_{\nu=1}^n f(x, \eta_\nu) \Psi_\nu(y).$$

Consequently from equation (15) we find that

$$\int_0^1 G(x, y) \psi(y) dy = 0,$$

and hence

$$S_G \psi(x) = \psi(x)$$

and

$$S_f \varphi(x) = \psi(x).$$

Thus equation (15) gives necessary and sufficient conditions for the equation

$$S_f \varphi(x) = \psi(x)$$

to have a solution.

10. The system of equations

$$\varphi_\lambda(x) + \int_0^1 \sum_{\nu=1}^n f_{\lambda\nu}(x, y) \varphi_\nu(y) dy = \psi_\lambda(x) \quad (\lambda = 1 \dots n) \quad (16)$$

can be reduced to a single equation of the original type.

To show this, define a function $F(x, y)$ for values between 0 and n by the n^2 conditions

$$F(x, y) = f_{\lambda\nu}(x - \lambda + 1, y - \nu + 1), \quad \text{for } 0 < \frac{x - \lambda + 1}{y - \nu + 1} < 1.$$

Also define a function Ψ by the n conditions

$$\Psi(x) = \psi_\lambda(x - \lambda + 1) \quad \text{for } 0 < x - \lambda + 1 < 1.$$

Then if the determinant of the equation

$$\Phi(x) + \int_0^n F(x, y)\Phi(y) dy = \Psi(x) \quad (17)$$

is nonzero, the equation has a unique solution $\Phi(x)$. If we then define the functions φ_λ by the conditions

$$\Phi(x) = \phi_\lambda(x - \lambda + 1) \quad \text{for } 0 < x - \lambda + 1 < 1.$$

we see that these functions satisfy the original system.

We also see that this is the only solution that can satisfy the original system. For otherwise there would be another function satisfying (17), which is impossible.

§ 3. On the first variation of the determinant D_f .

11. Let us first calculate the first variation of

$$f \begin{pmatrix} x_1 \dots x_n \\ x_1 \dots x_n \end{pmatrix}.$$

If we denote the sequence of values x_1, x_2, \dots, x_n with the exception of x_λ by

$$x_1, x_2 \dots (x_\lambda) \dots x_n,$$

we can write

$$\delta f \begin{pmatrix} x_1 \dots x_n \\ x_1 \dots x_n \end{pmatrix} = \sum_{\lambda, \mu} (-1)^{\lambda+\nu} f \begin{pmatrix} x_1 \dots (x_\lambda) \dots x_n \\ x_1 \dots (x_\nu) \dots x_n \end{pmatrix} \delta f(x_\lambda, x_\mu).$$

Multiply the two sides by dx_1, \dots, dx_n and integrate from 0 to 1. If we note that names of the variables are immaterial, we clearly have

$$\begin{aligned} & \delta \int_0^1 \dots \int_0^1 f \begin{pmatrix} x_1 \dots x_n \\ x_1 \dots x_n \end{pmatrix} dx_1 \dots dx_n \\ &= n \int_0^1 \dots \int_0^1 f \begin{pmatrix} x_1, & x_2 \dots x_{n-1} \\ x_1, & x_2 \dots x_{n-1} \end{pmatrix} \delta f(x, x) dx dx_1 \dots dx_{n-1} \\ &-n(n-1) \int_0^1 \dots \int_0^1 f \begin{pmatrix} y, & x_1 \dots x_{n-2} \\ x, & x_1 \dots x_{n-2} \end{pmatrix} \delta f(x, y) dx dy dx_1 \dots dx_{n-2}. \end{aligned}$$

Multiplying by $\frac{1}{n!}$ and taking the sum from $n = 1$ to ∞ , we get

$$\delta D_f = \int_0^1 D_f \delta f(x, x) dx - \int_0^1 \int_0^1 D_f \begin{pmatrix} x \\ y \end{pmatrix} \delta f(x, y) dx dy$$

or

$$\delta \log \delta D_f = \int_0^1 \delta f(x, x) dx - \int_0^1 \int_0^1 \frac{D_f \begin{pmatrix} x \\ y \end{pmatrix}}{D_f} \delta f(x, y) dx dy.$$

Obviously

$$\delta f(x, y) - \int_0^1 \frac{D_f \begin{pmatrix} x \\ t \end{pmatrix}}{D_f} \delta f(t, y) dt = S_f^{-1} \delta f(x, y).$$

Hence we can also write

$$\delta \log D_f = \int_0^1 [S_f^{-1} \delta f(x, y)]_{x=y} dx. \quad (18)$$

By introducing the notation

$$T_f$$

for the transformation

$$\varphi(x) + \int_0^1 f(y, x) \varphi(y) dy$$

we get another expression for the logarithmic variation of D_f : namely,

$$\delta \log D_f = \int_0^1 [T_f^{-1} \delta f(x, y)]_{x=y} dx. \quad (18')$$

§ 4. The product theorem.

12 To establish the product theorem, consider the two transformations

$$S_f \varphi(x) = \varphi(x) + \int_0^1 f(x, y) \varphi(y) dy,$$

$$S_g \varphi(x) = \varphi(x) + \int_0^1 g(x, y) \varphi(y) dy.$$

If we write the product of the two transformations as

$$S_f S_g = S_F,$$

then we have

$$F(x, y) = f(x, y) + g(x, y) + \int_0^1 f(x, t)g(t, y) dt.$$

If we likewise consider the transformations

$$T_f\varphi(x) = \varphi(x) + \int_0^1 f(y, x)\varphi(y) dy,$$

$$T_g\varphi(x) = \varphi(x) + \int_0^1 g(y, x)\varphi(y) dy,$$

then we have

$$T_gT_f = S_G,$$

where

$$G(x, y) = f(y, x) + g(y, x) + \int_0^1 f(y, t)g(t, x) dt = F(y, x).$$

We have seen that

$$\delta \log \delta D_F = \int_0^1 \delta F(x, x) dx - \int_0^1 \int_0^1 \frac{D_F \begin{pmatrix} y \\ x \end{pmatrix}}{D_F} \delta F(x, y) dx dy,$$

a formula that can also be written as in (18) thus:

$$\delta \log D_F = \int_0^1 [(S_f S_g)^{-1} \delta F(x, y)]_{x=y} dx. \quad (19)$$

Moreover,

$$\delta \log D_F = \int_0^1 [(T_g T_f)^{-1} \delta F(x, y)]_{x=y} dx. \quad (20)$$

Now

$$\begin{aligned} \delta F(x, y) &= \delta f(x, y) + \delta g(x, y) + \int_0^1 [f(x, t)\delta g(t, y) + g(t, y)\delta f(x, t)] dt \\ &= T_g \delta f(x, y) + S_f \delta g(x, y). \end{aligned}$$

Hence on substituting this expression in (19) and (20), we find that

$$\begin{aligned} \delta \log D_F &= \int_0^1 [(T_g T_f)^{-1} T_g \delta f(x, y) + (S_f S_g)^{-1} S_f \delta g(x, y)]_{x=y} dx \\ &= \int_0^1 [T_f^{-1} \delta f(x, y) + S_g^{-1} \delta g(x, y)]_{x=y} dx \end{aligned}$$

or

$$\delta \log D_F = \delta \log D_f + \delta \log D_g.$$

It follows that

$$\log D_F - \log D_f - \log D_g$$

does not depend on the functions f and g . Finally, because for $f = g = 0$, we have $D_F = D_f = D_g = 1$, we have the following theorem:

$$D_F = D_f D_g. \quad (21)$$

§ 5. Various expansions.

13. We have seen that the function

$$\varphi(\xi, \eta) = \frac{D_f \left(\begin{smallmatrix} \xi \\ \eta \end{smallmatrix} \right)}{D_f}$$

satisfies the equation

$$\varphi(\xi, \eta) + \int_0^1 f(\xi, \tau) \varphi(\tau, \eta) d\tau = f(\xi, \eta). \quad (4_1)$$

Let us look for an expansion of the function $\varphi(\xi, \eta)$ in the form

$$\varphi(\xi, \eta) = \varphi_1(\xi, \eta) - \varphi_2(\xi, \eta) + \varphi_3(\xi, \eta) - \cdots, \quad (22)$$

where $\varphi_n(\xi, \eta)$ is of dimension n with respect to f .

Substituting this series in equation (4₁) and equating to zero the sum of terms having the same dimensions with respect to f , we get the following equations:

$$\begin{aligned} \varphi_1(\xi, \eta) &= f(\xi, \eta) \\ \varphi_n(\xi, \eta) &= \int_0^1 f(\xi, \tau) \varphi_{n-1}(\tau, \eta) d\tau \quad (n = 2, 3, \dots). \end{aligned}$$

From this it follows that

$$\varphi_n(\xi, \eta) = \int_0^1 \cdots \int_0^1 f(\xi, \tau_1) f(\tau_1, \tau_2) \cdots f(\tau_{n-1}, \eta) d\tau_1 \cdots d\tau_{n-1}.$$

The resulting expansion converges provided the upper bound of f is sufficiently small.

Now recall the formula (6), which for $n = 1$ we may write as

$$\lambda \frac{d \log D_{\lambda f}}{\eta} = \int_0^1 \varphi(\xi, \xi) d\xi.$$

If we replace $\varphi(\xi, \xi)$ by the representation (22), we have the formula

$$\begin{aligned} \log D_{\lambda f} &= \lambda \int_0^1 f(x, x) dx - \frac{\lambda^2}{2} \int_0^1 \int_0^1 f(x, y) f(y, x) dx dy + \cdots \\ &= - \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n} \int_0^1 \cdots \int_0^1 f(x_1, x_2) f(x_2, x_3) \cdots f(x_{n-1}, x_n) f(x_n, x_1) dx_1 \cdots dx_n. \end{aligned}$$

In particular, if the series on the right converges for $\lambda = 1$,

$$\log D_f = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \int_0^1 \cdots \int_0^1 f(x_1, x_2) f(x_2, x_3) \cdots f(x_{n-1}, x_n) f(x_n, x_1) dx_1 \cdots dx_n.$$

§ 6. The case where $f(x, y)$ becomes unbounded in such a way that $(x - y)^\alpha f(x, y)$ remains bounded.

Let $f(x, y)$ be a bounded integrable function and $i(x, y)$ be a function such that $(x - y)^\alpha i(x, y)$ is bounded and integrable. Suppose that D_f is zero along with its minors up to order n . If, in addition,

$$S_f S_i = S_i S_f,$$

then obviously

$$S_i \Phi_\lambda(x) = \sum_{\mu=1}^n p_{\lambda\mu} \Phi_\mu(x) \quad (\lambda = 1, \dots, n), \quad (23)$$

where $\Phi_1(x) \dots \Phi_n(x)$ are the n linearly independent solutions of the equation

$$S_f \varphi(x) = 0.$$

If we let

$$T_f \varphi(x) = \varphi(x) + \int_0^1 f(y, x) \varphi(y) dy,$$

then

$$T_i \Psi_\lambda(x) = \sum_{\mu=1}^n q_{\lambda\mu} \Psi_\mu(x) \quad (\lambda = 1, \dots, n), \quad (24)$$

where $\Psi_1(x) \dots \Psi_n(x)$ are the n linearly independent solutions of the equation

$$T_f \Psi(x) = 0.$$

I claim that the determinant of the coefficients $p_{\lambda\mu}$ is equal to that of the coefficients $q_{\lambda\mu}$.

I will prove the assertion under the assumption that the determinant of the quantities

$$c_{\lambda\mu} = \int_0^1 \Phi_\lambda(x)\Psi_\mu(x) dx$$

is nonzero. Clearly, a simple continuity argument allows the proposition to be extended to the case where this determinant is zero.

Noting that we have the identity

$$\int_0^1 \Psi(x)S_i\Phi(x) dx \int_0^1 \Phi(x)T_i\Psi(x) dx$$

and taking into account equations (23) and (24), we get

$$\sum_{\nu=1}^n c_{\nu\mu}p_{\lambda\nu} = \sum_{\nu=1}^n c_{\lambda\nu}q_{\mu\nu},$$

from which the desired result follows immediately.

15. Denote by $i(x, y)$ a function to which the transformation S_i belongs. We are going to look for conditions under which an inverse transformation for S_i exists, supposing that $i(x, y)$ becomes unbounded in such a way that $(x - y)^\alpha i(x, y)$ remains bounded. Here α is a number less than one.

If we set

$$i_\nu(x, y) = \int_0^1 \cdots \int_0^1 i(x, t_1)i(t_1, t_2) \cdots i(t_{\nu-1}, y) dt_1 \cdots dt_{\nu-1}$$

and

$$k(x, y) = -i(x, y) + i_2(x, y) - \cdots + (-1)^{n-1}i_{n-1}(x, y),$$

then we have

$$S_k S_i = S_i S_k = S_f,$$

where

$$f(x, y) = (-1)^{n-1}i_n(x, y)$$

If n is chosen so that

$$n > \frac{1}{1 - \alpha},$$

then $i_n(x, y)$ can no longer become unbounded.

To show this, we note that one can write

$$\int_0^1 \frac{dt}{|x - t|^\alpha |t - y|^\beta} < \frac{\Psi(\alpha, \beta)}{|x - y|^{\alpha + \beta - 1}}, \quad (25)$$

where $\Psi(\alpha, \beta)$ is a function that is bounded as long as

$$0 < \alpha < 1, \quad 0 < \beta < 1, \quad \alpha + \beta < 1.$$

The inequality (25) can easily be established by making a the change of variable

$$t = x + (y - x)s$$

in the above integral. The repeated application of inequality (25) to the inequality

$$|i(x, y)| < \frac{a}{|x - y|^\alpha}$$

easily leads to the result that

$$|i_\nu(x, y)| < \frac{a_\nu}{|x - y|^{\nu\alpha - \nu + 1}},$$

as long as

$$\nu\alpha - \nu + 1 < 0,$$

which is to say that

$$\nu > \frac{1}{1 - \alpha}.*$$

If

$$\frac{a}{1 - \alpha} - 1 < n - 1 < \frac{a}{1 - \alpha},$$

we have

$$|i_n(x, y)| < \int_0^1 \frac{a_{n-1}\alpha dt}{|x - t|^{(n-1)\alpha - n + 2}|t - y|^\alpha}. \quad (25)$$

From this inequality it follows that $i_n(x, y)$ has a finite upper bound.

16. The results obtained above can be extended almost immediately to the more general transformations

$$S_i\varphi(x_1 \dots x_n) = \varphi(x_1 \dots x_n) + \int_0^1 \dots \int_0^1 i(x_1 \dots x_n; y_1 \dots y_n)\varphi(y_1 \dots y_n) dy_1 \dots dy_n$$

by postulating that $i(x_1 \dots x_n; y_1 \dots y_n)$ becomes unbounded in such a way that

$$r^\alpha i(x_1 \dots y_1 \dots)$$

remains bounded. Here α is a suitably chosen number less than n , and r is the distance between the points whose Cartesian coordinates are $x_1 \dots x_n$ and $y_1 \dots y_n$.

*Fredholm writes $<$ here.

Specifically, we have

$$\sum_{\nu=1}^n (x_\nu - y_\nu)^2 > n \sqrt[n]{\prod_{\nu=1}^n (x_\nu - y_\nu)^2}.$$

or

$$r \geq \sqrt[n]{n} \prod_{\nu=1}^n |x_\nu - y_\nu|^{\frac{1}{n}}.$$

Hence there is a number a such that

$$|i| \leq \frac{a}{\prod_{\nu=1}^n |x_\nu - y_\nu|^{\frac{a}{n}}}.$$

We define the functions i_ν in the same way as before; i.e., we set

$$i_\nu(x_1 \dots x_n) = \int_0^1 \dots \int_0^1 i(x_1 \dots x_n; t_1 \dots t_n) i_{\nu-1}(t_1 \dots t_n) dt_1 \dots dt_n.$$

By an argument analogous to the one used in the preceding case, we establish the inequality

$$|i_\lambda(x_1 \dots x_n, y_1 \dots y_n)| < \frac{a_\nu}{\{\prod |x_\nu - y_\nu|\}^{\frac{\lambda a}{n} - \lambda + 1}},$$

and from this inequality we conclude that if $\lambda > \frac{1}{1 - \frac{a}{n}}$ then i_λ does not become unbounded.

17. To shorten the presentation in showing how these results apply to the solution of the equation

$$S_i \varphi(x) = \psi(x),$$

I will restrict myself to the case where i depends only on two variables.

If we apply the transformation S_k to the two sides of the above equation, we have

$$S_k S_i \varphi(x) = S_f \varphi(x) = S_k \psi(x).$$

Here f and $S_k \psi(x)$ are bounded functions and clearly are also integrable. Hence we can apply the procedures described in Section 2 to the equation

$$S_f \varphi(x) = S_k \psi(x). \tag{26}$$

To treat the most general case, let us suppose that D_f is zero along with its minors up to order n . We will use the notation of § 2.

Applying the *pseudo-inverse* transformation of S_f to both sides of equation (7), we have

$$S_g S_f \varphi(x) = S_F \varphi(x) + S_g S_k \psi(x),$$

or

$$\varphi(x) = S_g S_k \psi(x) = \sum_{\nu=1}^n c_k \Phi_\nu(x).$$

If there is a solution of the given equation, then one can determine the coefficients in such a way that $S_i \varphi(x)$ is equal to $\psi(x)$.

18. Among the cases where this determination is possible, there is one that seems to me to merit attention. It is the case where the equation

$$S_i \varphi(x) = 0$$

has only the solution

$$\varphi(x) = 0.$$

We obviously have

$$S_i S_f = S_f S_i.$$

Hence

$$S_i \Phi_\lambda(x) = \sum_{\mu=1}^n p_{\lambda\mu} \Phi_\mu(x).$$

Since the functions Φ_μ are linearly independent and the equation

$$S_i \varphi(x) = 0$$

has only the solution $\phi(x) = 0$, the determinant of the coefficients $p_{\lambda\mu}$ is nonzero.

Since the determinant of the $p_{\lambda\mu}$ is nonzero, the determinant of the $q_{\lambda\mu}$ is also nonzero. It follows that the equation

$$T_i \varphi(x) = 0$$

has only the solution $\varphi(x) = 0$ and we have

$$\left. \begin{array}{l} S_k \Phi_\lambda = 0, \\ T_k \Psi_\lambda = 0. \end{array} \right\} \quad (\lambda = 1, \dots, n) \quad (27)$$

Given all this, if we set

$$\varphi_0(x) = S_g S_k \psi(x),$$

we have

$$\begin{aligned} S_f \varphi_0(x) &= S_f S_g S_k \psi(x) = S_G S_k \psi(x) \\ &= S_k \psi(x) - \sum_{\nu=1}^n f(x, \eta_\nu) \int_0^1 \psi_\nu(x) S_k \psi(x) dx. \end{aligned}$$

But

$$\int_0^1 \psi_\nu(x) S_k \psi(x) dx = \int_0^1 \psi(x) T_k \psi_\nu(x) dx = 0.$$

for all ν . Consequently,

$$S_f \varphi_0(x) - S_k \varphi(x) = 0,$$

or

$$S_k(S_i \varphi_0(x) - \psi(x)) = 0.$$

From this we conclude that

$$S_i \varphi_0(x) = \psi(x) + \sum_{\nu=1}^n a_\nu \Phi_\nu(x),$$

where the a_ν are known quantities.

If we now set

$$\varphi(x) = \varphi_0(x) + \sum_{\nu=1}^n c_\nu \Phi_\nu(x),$$

we get

$$S_i \varphi(x) = \psi(x) + \sum_{\nu=1}^n a_\nu \Phi_\nu(x) + \sum_{\nu=1}^n \sum_{\lambda=1}^n p_{\lambda\nu} c_\lambda \Phi_\nu(x).$$

But because the determinant of the coefficients $p_{\lambda\nu}$ is nonzero, we can obviously determine the c_ν so that

$$S_i \varphi(x) = \psi(x).$$

QED

Foundations of a General Theory of Linear Integral Equations

(First Communication)

by

David Hilbert at Göttingen

Presented in the session of March 5

Let $K(s, t)$ be a function of the real variables s and t . Let $f(s)$ be a given function of s and let $\varphi(s)$ be taken as a function of s to be determined. Each of the variables s and t range over the interval from a to b . Then

$$f(s) = \int_a^b K(s, t)\varphi(t) dt$$

is called an *integral equation of the first kind* and

$$f(s) = \varphi(s) - \lambda \int_a^b K(s, t)\varphi(t) dt$$

is called an *integral equation of the second kind*. Here λ denotes a parameter. The function $K(s, t)$ is called the *kernel of the integral equation*.

GAUSS was first led to a particular integral equation by boundary value problems in potential theory. The term “integral equation” was first used by P. DU BOIS-RAYMOND.¹ The first method for solving the integral of the second kind was due to C. NEUMANN.² In his method the function $\varphi(s)$ appears directly as infinite series in powers of the parameter λ whose coefficients are certain functions of s defined by multiple integrals. FREDHOLM³ found a different formula for the solution of the integral equation of the second kind, in which he was able to represent $\varphi(s)$ as a quotient, whose numerator is an everywhere convergent power series in λ with certain coefficients that depend on s , while its denominator turns out to be a power series in λ with numerical coefficients. At my suggestion KELLOGG⁴ produced a direct proof of the agreement of the formulas of C. NEUMANN and FREDHOLM. In the special case of a certain boundary

¹Bemerkungen über $\Delta z = 0$. Journ. f. Math. Bd. 103 (1888).

²Ueber die Methode des arithmetische Mittels. Leipz. Abh. Bd. 13 (1887).

³Sur une classe d'équations fonctionnelles. Acta mathematica Bd. 27 (1903). Also the 1899 paper cited there concerning the same topic.

⁴Zur Theorie der Integralgleichungen. Gött. Nachr. 1902.

value problem in potential theory, POINCARÉ⁵ was the first to introduce the parameter λ and was also the first to show that the solution must necessarily be in the form of the quotient of two everywhere convergent power series in λ . In a later paper in these proceedings I will treat a third method for the solution of the integral equation of the second kind that is at the same time also applicable to the integral equation of the first kind. VOLTERRA⁶ has obtained the solution of special integral equations. In certain cases the integral equation of the first kind can be reduced to one of the second kind by a method pointed out by me.⁷

A closer examination of the topic lead me to the realization that a systematic development of a general theory of linear of linear integral equations is of the greatest importance for analysis as a whole, in particular for the theory of the definite integral and the theory of the expansion of arbitrary functions in infinite series and in addition for the theory of linear differential equations as well as potential theory and the calculus of variation. In a series of papers I intend to give a new treatment of the question of the solution of integral equations, and above all to explore the interrelations and general properties of their solutions. In this connection, I generally make an assumption that is essential for my results—namely that the the kernel $K(s, t)$ of the integral equation is a *symmetric* function of the variables s and t . In particular, in this first paper I obtain formulas that yield the expansion of an arbitrary function in certain special functions that I call *eigenfunctions*. This result subsumes as special cases the expansion in trigonometric, Bessel, Kugel, Lamé and Sturm function, as well as the expansion in those functions of several variables that POINCARÉ first established in his investigations of certain boundary value problems of potential theory. My investigations will show that the theory of the expansion of arbitrary functions by no means requires the introduction of ordinary or partial differential equations; rather it is the *integral equation* that constitutes the necessary basis and the natural starting point for a theory of expansion in series. The most remarkable result is that the expansion of a function by the eigenfunctions belonging to a integral equation of the *second* kind turns out to depend on the solvability of the corresponding integral equation of the *first* kind.

At the same time, the problem of the *existence* of eigenfunctions receives a new and more nearly perfect solution. In the special case of the boundary value problems of potential theory, H. WEBER,⁸ as is well known, first tried to prove the existence of eigenfunctions on the basis of the DIRICHLET–THOMSON minimum principle. Then POINCARÉ actually produced the proof for this special problem by using a method

⁵Sur les équations de la physique mathématique. Rendiconti del circolo di Palermo t. 8 (1894). La méthode de Neumann et le problème de Dirichlet. Acta mathematica Bd. 20 (1896–97).

⁶Sopra alcune questioni di inversione de integrali definite. Annali di matematica s. 2 t. 25 (1897)

⁷Cf. KELLOGG, Zur Theorie der Integralgleichungen. Inaugural-Dissertation, Göttingen 1902, as well as Math. Ann. Bd. 58

⁸Ueber die Integration der partiellen Differentialgleichung $\Delta u + k^2 u = 0$. Math. Ann. Bd. 1 (1868).

developed by H. A. SCHWARTZ. Not only does the existence of eigenfunctions in the most general case follow from an application of my theorems, but at the same time my theory also gives in simple terms necessary and sufficient conditions for the existence of infinitely many eigenfunctions. This result is basically predicated on the fact that I do not, as has previously been done, proceed from an initial proof of the existence of eigenvalues. On the contrary, I first establish a general expansion theorem (pp. 74–75), and from this I can easily derive the conditions for the existence of eigenvalues and eigenfunctions.

The strategy that I use in this first paper is the following. I start with an algebraic problem, namely the problem of the orthogonal transformation a quadratic form in n variables into a sum of squares. Then by a rigorous process of passing to the limit for $n = \infty$, I arrive at the solution of the transcendental problem that was to be treated.⁹ The basic idea has been used by many others as a heuristic expedient, in particular by LORD RAYLEIGH.¹⁰ I have turned it into a rigorous principle.

For the sake of greater intelligibility and a shorter presentation in this communication, I have confined myself strictly to the case of an integral equation with a *simple* integral. However, the strategy and results are also valid when the simple integral in the integral equation given above is replaced by a *double* or *multiple* integral, so that K denotes a symmetric function of two *sequences* of variables.

I.

Solution of the Algebraic Problem

Let $K(s, t)$, $f(s)$, and $\varphi(s)$ have the meaning given at the beginning of this paper. For simplicity, however, we take the interval of the variables s and t to be the interval from 0 to 1. In addition, let $K(s, t)$ be a *symmetric* function of s and t . Furthermore, by n we will understand a fixed positive integer and introduce the following abbreviations.

$$K_{pq} = K\left(\frac{p}{n}, \frac{q}{n}\right) \quad (p, q = 1, 2, \dots, n)$$

$$Kxy = K_{11}x_1y_1 + K_{12}x_1y_2 + K_{21}x_2y_1 + \dots + K_{nn}x_ny_n$$

$$= \sum_{p,q} K_{pq}x_py_q, \quad (K_{pq} = K_{qp}),$$

$$\varphi_p = \varphi\left(\frac{p}{n}\right), \quad f_p = f\left(\frac{p}{n}\right), \quad (p = 1, 2, \dots, n),$$

⁹I have presented the basic idea of this strategy repeatedly in my seminar and in lectures during W.-S. 1900–1901.

¹⁰Cf. RAYLEIGH, *The Theory of Sound*, 2. ed. London 1894–1986 and PÖCKELS-KLEIN, *Ueber die partielle Differentialgleichung $\Delta u + K^2u = 0$ und deren Auftreten in der mathematischen Physik*. Leipzig 1891.

If we denote by $D \left(\ell, \begin{smallmatrix} x \\ Ky \end{smallmatrix} \right)$ the determinant that results from replacing every occurrence of y_p in $D \left(\ell, \begin{smallmatrix} x \\ y \end{smallmatrix} \right)$ by

$$Ky_p = K_{p1}y_1 + K_{p2}y_2 + \cdots + K_{pn}y_n,$$

then we have the easily following easily verified identity in x, y , and ℓ :

$$d(\ell)[x, y] + D \left(\ell, \begin{smallmatrix} x \\ y \end{smallmatrix} \right) - \ell D \left(\ell, \begin{smallmatrix} x \\ Ky \end{smallmatrix} \right) = 0. \quad (3)$$

Our problem now consists of determining the n unknowns $\varphi_1, \varphi_2, \dots, \varphi_n$ from the equations (1) or (2); that is, to find a linear form

$$[\varphi, y] = \varphi_1y_1 + \varphi_2y_2 + \cdots + \varphi_ny_n$$

that satisfies the equation

$$[f, y] = [\varphi, y] - \ell[K\varphi, y]$$

identically in y . Since from

$$K_{pq} = K_{qp}$$

it necessarily follows that

$$[K\varphi, y] = [\varphi, Ky],$$

the equation to be satisfied is equivalent to the equation

$$[f, y] = [\varphi, y] - \ell[\varphi, Ky].$$

It is immediately clear from (3) that this equation is solved by the formula

$$[\varphi, y] = -\frac{D \left(\ell, \begin{smallmatrix} f \\ y \end{smallmatrix} \right)}{d(\ell)}. \quad (4)$$

Thus if ℓ has the property that $d(\ell) \neq 0$, the coefficients of the linear form (4) are the values of the unknowns $\varphi_1, \varphi_2, \dots, \varphi_n$ that we have been seeking.

It is well known that the roots of the equation

$$d(\ell) = 0$$

are all real. We denote them by

$$\ell^{(1)}, \ell^{(2)}, \dots, \ell^{(n)}$$

and assume that they are distinct from one another.

If we denote by $d_{11}(\ell), \dots, d_{nn}(\ell)$ the minors of $d(\ell)$ with respect to its n diagonal elements and if $d'(\ell)$ is the derivative of $d(\ell)$ with respect to ℓ , then the equation

$$d_{11}(\ell) + \dots + d_{nn}(\ell) = nd(\ell) - \ell d'(\ell)$$

holds identically in ℓ . Hence for $\ell = \ell^{(h)}$ it follows that

$$d_{11}(\ell^{(h)}) + \dots + d_{nn}(\ell^{(h)}) = -\ell^{(h)} d'(\ell^{(h)}). \quad (5)$$

According to our assumptions $d'(\ell^{(h)})$ cannot be zero. Hence the minors on the left-hand side certainly cannot all be zero; that is, the homogeneous equations

$$\begin{aligned} \varphi_1 - \ell K \varphi_1 &= 0 \\ \dots\dots\dots & \\ \varphi_n - \ell K \varphi_n &= 0 \end{aligned} \quad (6)$$

for $\ell = \ell^{(h)}$ has a certain solution system

$$\varphi_1 = \varphi_1^{(h)}, \dots, \varphi_n = \varphi_n^{(h)}$$

that is uniquely determined up to a common factor of these n quantities. By (3) the coefficients of y_1, \dots, y_n in the expression

$$D \left(\ell^{(h)}, \begin{matrix} x \\ y \end{matrix} \right)$$

must be solutions of the homogeneous equations (6) that are independent of x . Hence we have the proposition

$$D \left(\ell^{(h)}, \begin{matrix} x \\ y \end{matrix} \right) = [\psi^{(h)}, x][\varphi^{(h)}, y],$$

where the first factor in the right-hand side represents a linear form in x_1, \dots, x_n . From this and the symmetry of the expression on the left-hand side, it follows on exchanging x and y that

$$D \left(\ell^{(h)}, \begin{matrix} x \\ y \end{matrix} \right) = C[\varphi^{(h)}, x][\varphi^{(h)}, y],$$

where C is a constant that is independent of x and y . If we choose the the common factor mentioned above suitably, we get

$$D \left(\ell^{(h)}, \begin{matrix} x \\ y \end{matrix} \right) = \pm[\varphi^{(h)}, x][\varphi^{(h)}, y]. \quad (7)$$

From this equation, by comparing coefficients of the products

$$x_1 y_1, \dots, x_n y_n$$

on both sides, we get the special formula

$$d_{11}(\ell^{(h)}) + \dots + d_{nn}(\ell^{(h)}) = \mp [\varphi^{(h)}, \varphi^{(h)}]. \quad (8)$$

Hence from (5)

$$[\varphi^{(h)}, \varphi^{(h)}] = \pm \ell^{(h)} d'(\ell^{(h)}), \quad (h = 1, 2, \dots, n), \quad (9)$$

and thence from (7)

$$\frac{D\left(\ell^{(h)}, \frac{x}{y}\right)}{\ell^{(h)} d'(\ell^{(h)})} = \frac{[\varphi^{(h)}, x][\varphi^{(h)}, y]}{[\varphi^{(h)}, \varphi^{(h)}]}, \quad (h = 1, 2, \dots, n). \quad (10)$$

Equation (9) shows that one must take the top or bottom sign in (7) and (8) according as $\ell^{(h)} d'(\ell^{(h)})$ turns out to be positive or negative. We can write the equations (6) as an identity in x as follows:

$$[\varphi^{(h)}, x] = \ell^{(h)} [\varphi^{(h)}, Kx]. \quad (11)$$

Since for unequal indices $\ell^{(h)}$ and $\ell^{(k)}$ are different, we obtain from (11) the relation

$$[\varphi^{(h)}, \varphi^{(k)}] = 0, \quad (h \neq k).$$

Finally, to obtain the connection with the theory of the orthogonal transformation of a quadratic form, we start with the expression

$$\frac{D\left(\ell, \frac{x}{y}\right)}{d(\ell)}.$$

Since the numerator is a function of degree $n - 1$ in ℓ and the denominator is of degree n , from the rule for expansion in partial fractions and the use of (10) we get the formula

$$\begin{aligned} \frac{D\left(\ell, \frac{x}{y}\right)}{d(\ell)} &= \frac{D\left(\ell^{(1)}, \frac{x}{y}\right)}{d'(\ell^{(1)})} \frac{1}{\ell - \ell^{(1)}} + \dots + \frac{D\left(\ell^{(n)}, \frac{x}{y}\right)}{d'(\ell^{(n)})} \frac{1}{\ell - \ell^{(n)}} \\ &= \frac{[\varphi^{(1)}, x][\varphi^{(1)}, y]}{[\varphi^{(1)}, \varphi^{(1)}]} \frac{\ell^{(1)}}{\ell - \ell^{(1)}} + \dots + \frac{[\varphi^{(n)}, x][\varphi^{(n)}, y]}{[\varphi^{(n)}, \varphi^{(n)}]} \frac{\ell^{(n)}}{\ell - \ell^{(n)}}, \end{aligned}$$

which is satisfied identically in x , y , and ℓ . For $\ell = 0$, this formula becomes

$$[x, y] = \frac{D(\ell^{(1)}, x)}{\ell^{(1)} d'(\ell^{(1)})} + \cdots + \frac{D(\ell^{(n)}, x)}{\ell^{(n)} d'(\ell^{(n)})} \quad (12)$$

$$= \frac{[\varphi^{(1)}, x][\varphi^{(1)}, y]}{[\varphi^{(1)}, \varphi^{(1)}]} + \cdots + \frac{[\varphi^{(n)}, x][\varphi^{(n)}, y]}{[\varphi^{(n)}, \varphi^{(n)}]}. \quad (13)$$

If we replace y by the linear combination Ky , in view of (11) we get

$$Kxy = [Kx, y] = [x, Ky] = \frac{D(\ell^{(1)}, x)}{(\ell^{(1)})^2 d'(\ell^{(1)})} + \cdots + \frac{D(\ell^{(n)}, x)}{(\ell^{(n)})^2 d'(\ell^{(n)})} \quad (14)$$

$$= \frac{[\varphi^{(1)}, x][\varphi^{(1)}, y]}{\ell^{(1)}[\varphi^{(1)}, \varphi^{(1)}]} + \cdots + \frac{[\varphi^{(n)}, x][\varphi^{(n)}, y]}{\ell^{(n)}[\varphi^{(n)}, \varphi^{(n)}]}. \quad (15)$$

In addition, we append the particular formulas that result from the last two sets of formulas by setting y equal to x :

$$[x, x] = \frac{D(\ell^{(1)}, x)}{\ell^{(1)} d'(\ell^{(1)})} + \cdots + \frac{D(\ell^{(n)}, x)}{\ell^{(n)} d'(\ell^{(n)})} \quad (16)$$

$$= \frac{[\varphi^{(1)}, x]^2}{[\varphi^{(1)}, \varphi^{(1)}]} + \cdots + \frac{[\varphi^{(n)}, x]^2}{[\varphi^{(n)}, \varphi^{(n)}]}.$$

$$Kxx = \frac{D(\ell^{(1)}, x)}{(\ell^{(1)})^2 d'(\ell^{(1)})} + \cdots + \frac{D(\ell^{(n)}, x)}{(\ell^{(n)})^2 d'(\ell^{(n)})} \quad (17)$$

$$= \frac{[\varphi^{(1)}, x]^2}{\ell^{(1)}[\varphi^{(1)}, \varphi^{(1)}]} + \cdots + \frac{[\varphi^{(n)}, x]^2}{\ell^{(n)}[\varphi^{(n)}, \varphi^{(n)}]}.$$

II.

Solution of the Transcendental Problem

Recalling the significance of the quantities K_{pq} , which were constructed from the function $K(s, t)$ at the beginning of Section I, we now assume that $K(s, t)$ is a symmetric *continuous* function of the variables s and t in the interval of interest, namely $[0, 1]$. Our strategy requires the rigorous passage to the limit $n = \infty$. The transcendental problem of solving the integral equation of the second kind

$$f(s) = \varphi(s) - \lambda \int_0^1 K(s, t) dt$$

corresponds to the preliminary algebraic problem that we solved in Section I. In this Section II we essentially confine ourselves to using our strategy to derive the formulas required to solve the integral equation, formulas that were first given by *Fredholm*.

If we expand $d(\ell)$ in powers of ℓ as follows

$$d(\ell) = 1 - d_1\ell + d_2\ell^2 - \dots \pm d_n\ell^n$$

and let h stand for any the indices $1, 2, \dots, n$, then we have

$$d_h = \sum_{(p_1, p_2, \dots, p_h)} \begin{vmatrix} K_{p_1 p_1} & K_{p_1 p_2} & \dots & K_{p_1 p_h} \\ K_{p_2 p_1} & K_{p_2 p_2} & \dots & K_{p_2 p_h} \\ \dots & \dots & \dots & \dots \\ K_{p_h p_1} & K_{p_h p_2} & \dots & K_{p_h p_h} \end{vmatrix}, \left(\begin{matrix} p_1 < p_2 < p_3 < \dots < p_h \\ p_1, p_2, \dots, p_h = 1, 2, \dots, n \end{matrix} \right).$$

The sum on the right-hand side consists of $\binom{n}{h}$ determinants. According to a well-known theorem¹¹ the absolute value of any of these determinants cannot exceed the bound $\sqrt{h^h}K^h$, where K denotes the maximum of the absolute values of the function values $K(s, t)$. From this we find that

$$|d_h| \leq \binom{n}{h} \sqrt{h^h} K^h \leq \frac{\sqrt{h^h}}{h!} (nK)^h \leq \left(\frac{neK}{\sqrt{h}} \right)^h.$$

In other words,

$$\frac{|d_h|}{n^h} \leq \left(\frac{eK}{\sqrt{h}} \right)^h. \tag{18}$$

On the other hand, we can easily show that if h is fixed, in the limit as n increases without bound

$$\lim_{n \rightarrow \infty} \frac{d_h}{n^h} = \delta_h, \tag{19}$$

where δ_h stands for the multiple integral

$$\delta_h = \frac{1}{h!} \int_0^1 \dots \int_0^1 \begin{vmatrix} K(s_1, s_1) & K(s_1, s_2) & \dots & K(s_1, s_h) \\ \dots & \dots & \dots & \dots \\ K(s_h, s_1) & K(s_h, s_2) & \dots & K(s_h, s_h) \end{vmatrix} ds_1 \dots ds_h.$$

From (18) and (19) it follows that

$$|\delta_h| \leq \left(\frac{eK}{\sqrt{h}} \right)^h. \tag{20}$$

¹¹HADAMARD, Bulletin des sciences mathématiques (2) XVII (1893).

We now introduce the power series first given by FREDHOLM

$$\delta(\lambda) = 1 - \delta_1\lambda + \delta_2\lambda^2 - \delta_3\lambda^3 + \dots,$$

which because of (20) is everywhere convergent. We can now establish the following lemma.

LEMMA 1. As n increases without bound, the expression $d\left(\frac{\lambda}{n}\right)$ converges to $\delta(\lambda)$, and the convergence is uniform for all values of λ whose absolute value lies below an arbitrarily chosen positive bound Λ . In the same sense, the expression $\frac{1}{n}d'\left(\frac{\lambda}{n}\right)$ converges to $\delta'(\lambda)$.

To prove this lemma, we will suppose, on the contrary, that there exists a positive quantity ϵ such that for infinitely many integers n and corresponding values λ with absolute value not greater than Λ we have

$$\left|d\left(\frac{1}{n}\right) - \delta(\lambda)\right| > \epsilon.$$

We now choose an integer m so large that the following conditions are satisfied. For all λ whose absolute value is not greater than Λ ,

$$|\delta_{m+1}\lambda^{m+1} - \delta_{m+2}\lambda^{m+2} + \dots| \leq \frac{\epsilon}{3}. \quad (21)$$

Further, the inequalities

$$m > (2eK\Lambda)^2 \quad (22)$$

and

$$\frac{1}{2^m} < \frac{\epsilon}{3} \quad (23)$$

are to be satisfied. Then in view of (18) and (22) for each n we surely have

$$\begin{aligned} d\left(\frac{1}{n}\right) &= 1 - \frac{d_1}{n}\lambda + \dots \pm \frac{d_m}{n^m}\lambda^m \mp \frac{d_{m+1}}{n^{m+1}}\lambda^{m+1} \pm \dots \pm \frac{d_n}{n^n}\lambda^n \\ &= 1 - \frac{d_1}{n}\lambda + \dots \pm \frac{d_m}{n^m}\lambda^m \pm \frac{\vartheta}{2^m} \quad (0 \leq \vartheta \leq 1). \end{aligned}$$

By (23)

$$\left|d\left(\frac{1}{n}\right) - \left(1 - \frac{d_1}{n}\lambda + \dots \pm \frac{d_m}{n^m}\lambda^m\right)\right| < \frac{\epsilon}{3}. \quad (24)$$

The integer m having been determined in this way, we choose the integer n so large that

$$\left|\left(1 - \frac{d_1}{n}\lambda + \dots \pm \frac{d_m}{n^m}\lambda^m\right) - (1 - \delta_1\lambda + \delta_2\lambda^2 - \dots \pm \delta_m\lambda^m)\right| < \frac{\epsilon}{3}. \quad (25)$$

Because of equation (19) such a choice of n is always possible. The inequalities (21), (24), and (25) now show that the absolute value of the difference between $d\left(\frac{\lambda}{n}\right)$ and $\delta(\lambda)$ must be less than ϵ . This result contradicts our assumption, and hence Lemma 1 is proved.

In order to discover how the passage to the limiting transcendental problem works out for the determinants $D\left(\ell, \frac{x}{y}\right)$, we will take $x(s)$ and $y(s)$ to be two arbitrary continuous functions of the variable s in the interval 0 to 1 and substitute

$$x_p = x\left(\frac{p}{n}\right), \quad y_p = y\left(\frac{p}{n}\right)$$

into into the determinants $D\left(\ell, \frac{x}{y}\right)$. We then expand these determinants in powers of ℓ as follows:

$$D\left(\ell, \frac{x}{y}\right) = D_1\left(\frac{x}{y}\right) - D_2\left(\frac{x}{y}\right)\ell + D_3\left(\frac{x}{y}\right)\ell^2 - \dots \pm D_{n-1}\left(\frac{x}{y}\right)\ell^{n-1}.$$

It is easily seen that in the limit with infinitely increasing n and fixed h

$$\lim_{n \rightarrow \infty} \frac{D_h\left(\frac{x}{y}\right)}{n^h} = \Delta_h\left(\frac{x}{y}\right)$$

where $\Delta_h\left(\frac{x}{y}\right)$ denotes the multiple integral

$$\Delta_h\left(\frac{x}{y}\right) = \frac{1}{h!} \int_0^1 \dots \int_0^1 \begin{vmatrix} 0 & x(s_1) & x(s_2) & \dots & x(s_h) \\ y(s_1) & K(s_1, s_1) & K(s_1, s_2) & \dots & K(s_1, s_h) \\ \dots & \dots & \dots & \dots & \dots \\ y(s_h) & K(s_h, s_1) & K(s_h, s_2) & \dots & K(s_h, s_h) \end{vmatrix}, ds_1 \dots ds_h$$

If we now introduce the everywhere convergent power series

$$\Delta\left(\lambda, \frac{x}{y}\right) = \Delta_1\left(\frac{x}{y}\right) - \Delta_2\left(\frac{x}{y}\right)\lambda + \Delta_3\left(\frac{x}{y}\right)\lambda^2 - \dots,$$

we get the following lemma by a proof analogous to the previous one.

LEMMA 2. *As n increases without bound the expression $\frac{1}{n}D\left(\frac{\lambda}{n}, \frac{x}{y}\right)$ converges to $\Delta\left(\lambda, \frac{x}{y}\right)$. Moreover, this convergence is uniform for all λ whose absolute value lies below an arbitrarily chosen bound Λ .*

Thus it is seen that $\Delta\left(\lambda, \frac{x}{y}\right)$ is a power series in λ whose coefficients depend on the arbitrary functions $x(s)$ and $y(s)$.

We now proceed to determine the limit in the formula (3) for $n = \infty$.

Bearing in mind that according to the abbreviations introduced at the beginning Ky_p is

$$\begin{aligned} Ky_p &= K_{p1}y_1 + K_{p2}y_2 + \cdots + K_{pn}y_n \\ &= K\left(\frac{p}{n}, \frac{1}{n}\right)y\left(\frac{1}{n}\right) + K\left(\frac{p}{n}, \frac{2}{n}\right)y\left(\frac{2}{n}\right) + \cdots + K\left(\frac{p}{n}, \frac{n}{n}\right)y\left(\frac{n}{n}\right), \end{aligned}$$

we obtain the following formula by the same method that led to Lemmas 1 and 2:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\lambda}{n^2} D\left(\frac{\lambda}{n}, \frac{x}{Ky}\right) &= \lim_{n \rightarrow \infty} \frac{\lambda}{n} D\left(\frac{\lambda}{n}, \frac{x}{\frac{1}{n}Ky}\right) \\ &= \lambda \left\{ \Delta\left(\lambda, \frac{x}{\bar{y}}\right) \right\}_{\bar{y}(s) = \int_0^1 K(s,t)y(t) dt} \\ &= \lambda \int_0^1 \left\{ \Delta\left(\lambda, \frac{x}{\bar{y}}\right) \right\}_{\bar{y}(s) = K(s,t)} \cdot y(t) dt. \end{aligned}$$

Therefore if in formula (3) we set $\ell = \frac{\lambda}{n}$ and divide the formula by n , then the limit as n grows without bound becomes

$$\delta(\lambda) \int_0^1 x(s)y(s) ds + \Delta\left(\lambda, \frac{x}{y}\right) - \lambda \int_0^1 \left\{ \Delta\left(\lambda, \frac{x}{\bar{y}}\right) \right\}_{\bar{y}(s) = K(s,t)} \cdot y(t) dt = 0. \quad (26)$$

This formula is an identity in λ and holds whenever $x(s)$ and $y(s)$ are continuous functions of their arguments.

If in (26) we set

$$x(r) = K(r, s) \quad \text{and} \quad y(r) = K(r, t)$$

and introduce the abbreviation

$$\Delta(\lambda; s, t) = \lambda \left\{ \Delta\left(\lambda, \frac{x}{y}\right) \right\}_{\substack{x(r) = K(r, s) \\ y(r) = K(r, t)}} - \delta(\lambda)K(s, t), \quad (27)$$

then (26) becomes

$$\delta(\lambda)K(s, t) + \Delta(\lambda; s, t) - \lambda \int_0^1 \Delta(\lambda; s, r)K(t, r) dr = 0. \quad (28)$$

Finally, setting

$$\mathbf{K}(s, t) = -\frac{\Delta(\lambda; s, t)}{\delta(\lambda)},$$

we get

$$K(s, t) = \mathbb{K}(s, t) - \lambda \int_0^1 \mathbb{K}(s, r)K(t, r) dr. \tag{29}$$

In the above equations, $\Delta(\lambda; s, t)$ and $\mathbb{K}(s, t)$ are symmetric functions of the real variables s and t , and in addition, they contain the parameter λ . The formulas (28) and (29) hold identically in s, t , and λ .

The function $\mathbb{K}(s, t)$ is called the *solution function for the kernel $K(s, t)$* .^{*} Specifically, by means of it the original integral equation of the second kind,

$$f(s) = \varphi(s) - \lambda \int_0^1 K(s, t)\varphi(t) dt$$

can be solved as follows:

$$\varphi(s) = f(s) + \lambda \int_0^1 \mathbb{K}(s, t)f(t) dt.$$

We can see this immediately by substituting the right-hand side of the last formula into the previous integral equation. At the same time we see that the solution of the integral equation of the second kind is unique for any λ that is not a zero of $\delta(\lambda)$.

From the above material, we obtain the following power series for $\Delta(\lambda; s, t)$:

$$\Delta(\lambda; s, t) = -K(s, t) + \Delta_1(s, t)\lambda - \Delta_2(s, t)\lambda^2 + \dots,$$

where

$$\Delta_h(s, t) = \frac{1}{h!} \int_0^1 \dots \int_0^1 \begin{vmatrix} K(s, t) & K(s, s_1) & \dots & K(s, s_h) \\ K(s_1, t) & K(s_1, s_1) & \dots & K(s_1, s_h) \\ \dots & \dots & \dots & \dots \\ K(s_h, t) & K(s_h, s_1) & \dots & K(s_h, s_h) \end{vmatrix} ds_1 \dots ds_h.$$

The identity in λ

$$\delta'(\lambda) = \int_0^1 \delta(\lambda; s, s) ds \tag{30}$$

follows easily from this formula.

The formulas obtained in this way are none other than the formulas of FREDHOLM, which we have already mentioned several times.

III.

The Transcendental Problem that Corresponds to the Transformation of a Quadratic Form into a Sum of Squares.

^{*}Hilbert's solution function is also called the resolvent function. It was first introduced, but not explicitly named, by Fredholm.

Our most important task consists of transferring the algebraic inquiry of Section 1 concerning the orthogonal transformation of the quadratic form Kxx into the transcendental domain by carrying out the passage to the limit for $n = \infty$.

To this end we first will prove the following propositions about the zeros of $\delta(\lambda)$.

PROPOSITION 1. *The function $\delta(\lambda)$ has no complex zeros.*

For the proof we will assume on the contrary that we have such a zero at hand, and we will construct a circle around it in the complex λ -plane, whose periphery and interior contain no other zeros of $\delta(\lambda)$ and moreover such that $\delta'(\lambda)$ is not zero on its periphery.* By Lemma 1, $d\left(\frac{\lambda}{n}\right)$ converges uniformly to $\delta(n)$ for infinitely increasing n and $\frac{1}{n}d'\left(\frac{\lambda}{n}\right)$ converges to $\delta'(\lambda)$. Hence for sufficiently large values of n the quotient

$\frac{d\left(\frac{\lambda}{n}\right)}{\frac{1}{n}d'\left(\frac{\lambda}{n}\right)}$ differs from $\frac{\delta(\lambda)}{\delta'(\lambda)}$ by an arbitrarily small quantity on the entire periphery of the circle. But then the difference of the integrals

$$\int \frac{d\left(\frac{\lambda}{n}\right)}{\frac{1}{n}d'\left(\frac{\lambda}{n}\right)} d\lambda \quad \text{and} \quad \int \frac{\delta(\lambda)}{\delta'(\lambda)} d\lambda$$

taken over the periphery of the circle would lie arbitrarily near zero. But this is impossible. For the first integral has the value zero, since the zeros of $d\left(\frac{\lambda}{n}\right)$ are all real, and on the other hand, the second integral will be equal to the integer that is the multiplicity of the zero of $\delta(\lambda)$ at the center of the circle.

In a similar manner because of the uniform convergence stated in Lemma 1, we have the following fact.

PROPOSITION 2. *Suppose that the n roots of the equations $d(\ell) = 0$*

$$\ell^{(1)}, \dots, \ell^{(n)}$$

are ordered by their absolute values in such a way that when equal roots of opposite signs occur the positive ones come first and, moreover, multiple roots appear as often as their multiplicities require. Let the zeros of $\delta(\lambda)$, such as there may be, be ordered likewise. Then

$$\lim_{n=\infty} n\ell^{(1)} = \lambda^{(1)}, \quad \lim_{n=\infty} n\ell^{(2)} = \lambda^{(2)}, \quad \lim_{n=\infty} n\ell^{(3)} = \lambda^{(3)} \dots$$

*As the proof below shows, the circle must not intersect the real axis.

There is no way to conclude from Proposition 2 that the zeros of $\delta(\lambda)$ exist. For it may well happen that as n increases without bound, the absolute value of $n\ell^{(1)}$ grows beyond any specific bound.

We now introduce the following terminology: The zeros of $\delta(\lambda)$ will be called *eigenvalues belonging to the kernel* $K(s, t)$.

Previously we have understood $K(s, t)$ to be a symmetric function of the real variables s and t . Throughout this third section we now make the assumption that function $\delta(\lambda)$ belonging to $K(s, t)$ has *no multiple roots*, so that for any root of the equation $\delta(\lambda) = 0$ it must happen that $\delta'(\lambda)$ is different from zero.

We must further note that the transformation theory, developed toward the end of Section I, of the quadratic form

$$Kxx = \sum_{p,q} K\left(\frac{p}{n}, \frac{q}{n}\right) x_p x_q, \quad (p, q = 1, 2, \dots, n)$$

presupposed that the determinants formed from $K(s, t)$ had no multiple zeros. Now if for any value of n the determinant $d(\ell)$ belonging to $K(s, t)$ should have a multiple zero, we can proceed as follows. For each such value of n we imagine that $K(s, t)$ is replaced by a modified function $\bar{K}(s, t)$, so that the zeros of the determinant $\bar{d}(\ell)$ constructed correspondingly from $\bar{K}(s, t)$ are all simple. Moreover, the difference between the values of the modified function $\bar{K}(s, t)$ and the values of the original kernel are to be so small that for all values of the variables s and t , for all indices h ($= 1, 2, \dots, n$), and for all pairs of continuous functions $x(s)$ and $y(s)$, the following inequalities are satisfied:

$$\begin{aligned} |K(s, t) - \bar{K}(s, t)| &< \frac{1}{n}, \\ |d_h - \bar{d}_h| &< 1, \\ |\ell^{(h)} - \bar{\ell}^{(h)}| &< \frac{1}{n^2}, \quad (h = 1, 2, \dots, n) \\ \left| D_h \begin{pmatrix} x \\ y \end{pmatrix} - \bar{D}_h \begin{pmatrix} x \\ y \end{pmatrix} \right| &< M(x) \cdot M(y). \end{aligned}$$

Here \bar{d}_h and $\bar{D}_h \begin{pmatrix} x \\ y \end{pmatrix}$ denote the coefficients of the determinants $\bar{d}(\ell)$ and $\bar{D} \left(\ell, \begin{pmatrix} x \\ y \end{pmatrix} \right)$ constructed from $\bar{K}(s, t)$. Moreover, $\bar{\ell}^{(h)}$ denotes the corresponding zeros of $\bar{d}(\ell)$, and $M(x)$ and $M(y)$ are the maxima of the absolute values of the functions $x(s)$ and $y(s)$. Clearly, as n increases without bound the expressions

$$\bar{K}(s, t), \quad \bar{d} \left(\frac{\lambda}{n} \right), \quad \frac{1}{n} \bar{D} \left(\frac{\lambda}{n}, \begin{pmatrix} x \\ y \end{pmatrix} \right)$$

approach the limits

$$K(s, t), \quad \delta(\lambda), \quad \Delta \left(\lambda, \begin{matrix} x \\ y \end{matrix} \right),$$

i.e., the same limits as the expressions constructed by means of the unmodified kernel. We are therefore in the position to use the theory of the quadratic form Kxx for whose validity the absence of multiple zeros on $d(\ell)$ was a necessary condition. In problematical cases we must work with the modified expressions, but to keep the big picture in view during the following exposition we will use the original expressions without the bars.

Let $\lambda^{(h)}$ denote the h th zero of $\delta(\lambda)$ with respect to the ordering defined on page 68. From (26) it follows that

$$\Delta \left(\lambda^{(h)}, \begin{matrix} x \\ y \end{matrix} \right) = \lambda^{(h)} \int_0^1 \left\{ \Delta \left(\lambda^{(h)}, \begin{matrix} x \\ \bar{y} \end{matrix} \right) \right\}_{\bar{y}(r)=K(r,t)} y(t) dt, \quad (31)$$

and because of the symmetry of the expression $\delta \left(\lambda, \begin{matrix} x \\ y \end{matrix} \right)$ with respect to $x(s)$ and $y(s)$ we also have

$$\Delta \left(\lambda^{(h)}, \begin{matrix} x \\ y \end{matrix} \right) = \lambda^{(h)} \int_0^1 \left\{ \Delta \left(\lambda^{(h)}, \begin{matrix} \bar{x} \\ y \end{matrix} \right) \right\}_{\bar{x}(r)=K(r,s)} x(s) ds.$$

If we set $y(r) = K(r, t)$, we get

$$\left\{ \Delta \left(\lambda^{(h)}, \begin{matrix} x \\ y \end{matrix} \right) \right\}_{y(r)=K(r,t)} = \lambda^{(h)} \int_0^1 \left\{ \Delta \left(\lambda^{(h)}, \begin{matrix} \bar{x} \\ y \end{matrix} \right) \right\}_{\substack{\bar{x}(r)=K(r,s) \\ y(r)=K(r,t)}} x(s) ds.$$

In view of (27)

$$\left\{ \Delta \left(\lambda^{(h)}, \begin{matrix} x \\ \bar{y} \end{matrix} \right) \right\}_{\bar{y}(r)=K(r,t)} = \int_0^1 \Delta(\lambda^{(h)}; s, t) x(s) ds. \quad (32)$$

From (31) and (32) we obtain

$$\Delta \left(\lambda^{(h)}, \begin{matrix} x \\ y \end{matrix} \right) = \lambda^{(h)} \int_0^1 \int_0^1 \Delta(\lambda^{(h)}; s, t) x(s) y(t) ds dt. \quad (33)$$

At the same time, if we substitute $x(r) = K(r, s)$ in (32) and recall (27), we get

$$\Delta(\lambda^{(h)}; s, t) = \lambda^{(h)} \int_0^1 \Delta(\lambda^{(h)}; r, t) K(r, s) dr. \quad (34)$$

From now on let $\ell^{(h)}$ denote the h th zero of $d(\ell)$ with respect to the ordering defined above. Because of the formula (7) we have generally

$$D\left(\ell^{(h)}, \begin{matrix} x \\ y \end{matrix}\right) D\left(\ell^{(h)}, \begin{matrix} x^* \\ y^* \end{matrix}\right) = D\left(\ell^{(h)}, \begin{matrix} x \\ x^* \end{matrix}\right) D\left(\ell^{(h)}, \begin{matrix} y \\ y^* \end{matrix}\right),$$

and from this it follows that in the limit for infinitely increasing n

$$\Delta\left(\lambda^{(h)}, \begin{matrix} x \\ y \end{matrix}\right) \Delta\left(\lambda^{(h)}, \begin{matrix} x^* \\ y^* \end{matrix}\right) = \Delta\left(\lambda^{(h)}, \begin{matrix} x \\ x^* \end{matrix}\right) \Delta\left(\lambda^{(h)}, \begin{matrix} y \\ y^* \end{matrix}\right),$$

provided that x^* and y^* , like x and y , represent continuous functions of their arguments. In view of (27), it follows that

$$\Delta(\lambda^{(h)}; s, t) \Delta(\lambda^{(h)}; s^*, t^*) = \Delta(\lambda^{(h)}; s, s^*) \Delta(\lambda^{(h)}; t, t^*). \quad (35)$$

Because of (30)

$$\int_0^1 \Delta(\lambda^{(h)}; s, s) ds = \delta'(\lambda). \quad (36)$$

From our assumption that the zeros of $\delta(\lambda)$ are simple, it follows that $\delta'(\lambda^{(h)})$ is different from zero, and hence $\Delta(\lambda^{(h)}; s, s)$ is not identically zero for all values of s . Let s^* be a particular value for which $\Delta(\lambda^{(h)}; s^*, s^*)$ is nonzero. We now set

$$\varphi^{(h)}(s) = \left| \sqrt{\frac{\lambda^{(h)}}{\Delta(\lambda^{(h)}; s^*, s^*)}} \right| \Delta(\lambda^{(h)}; s, s^*), \quad (37)$$

This defines $\varphi^{(h)}(s)$ as a continuous function of the variable s . It will be called the *eigenfunction corresponding to eigenvalue $\lambda^{(h)}$* . On replacing t^* with s^* we obtain from (35) and (37) the equation

$$\lambda^{(h)} \Delta(\lambda^{(h)}; s, t) = \pm \varphi^{(h)}(s) \varphi^{(h)}(t). \quad (38)$$

With the help of (36) it follows that

$$\int_0^1 (\varphi^{(h)}(s))^2 ds = \pm \lambda^{(h)} \delta'(\lambda^{(h)}).$$

Hence we see that the sign in the last two equations is plus or minus depending on whether $\lambda^{(h)} \delta'(\lambda^{(h)})$ is positive or negative.

In view of (33), we can derive the following formulas:

$$\Delta\left(\lambda^{(h)}, \begin{matrix} x \\ y \end{matrix}\right) = \pm \int_0^1 \varphi^{(h)}(s) x(s) ds \cdot \int_0^1 \varphi^{(h)}(s) y(s) ds$$

and

$$\frac{\Delta \left(\lambda^{(h)}, \begin{matrix} x \\ y \end{matrix} \right)}{\lambda^{(h)} \delta'(\lambda^{(h)})} = \frac{\int_0^1 \varphi^{(h)}(s)x(s) ds \cdot \int_0^1 \varphi^{(h)}(s)y(s) ds}{\int_0^1 (\varphi^{(h)}(s))^2 ds}.$$

Finally, formula (34) combined with (38) gives, after cancellation of the factor $\varphi^{(h)}(t)$,

$$\varphi^{(h)}(s) = \lambda^{(h)} \int_0^1 K(s, t) \varphi^{(h)}(t) dt.$$

From this we immediately derive the equation

$$\int_0^1 \varphi^{(k)}(s) \varphi^{(h)}(s) ds = 0 \quad (h \neq k),$$

where $\varphi^{(k)}(s)$ denotes the eigenfunction belonging to a different eigenvalue $\lambda^{(k)}$.

For the sake of brevity it is often preferable to use the functions

$$\psi^{(h)}(s) = \frac{\varphi^{(h)}(s)}{\left| \sqrt{\int_0^1 (\varphi^{(h)}(s))^2 ds} \right|}$$

instead of the eigenfunctions $\varphi^{(h)}(s)$. These functions may be called *normalized eigenfunctions*, or, when there appears to be no chance of misunderstanding, simply *eigenfunctions*. They satisfy the equations

$$\frac{\Delta \left(\lambda^{(h)}, \begin{matrix} x \\ y \end{matrix} \right)}{\lambda^{(h)} \delta'(\lambda^{(h)})} = \int_0^1 \psi^{(h)}(s)x(s) ds \cdot \int_0^1 \psi^{(h)}(s)y(s) ds \quad (39)$$

$$\int_0^1 (\psi^{(h)}(s))^2 ds = 1,$$

$$\int_0^1 \psi^{(h)}(s) \psi^{(k)}(s) ds = 0 \quad (h \neq k),$$

$$\psi^{(h)}(s) = \lambda^{(h)} \int_0^1 K(s, t) \psi^{(h)}(t) dt. \quad (40)$$

We have now finished the preliminaries needed to treat the question of what we obtain from the algebraic problem of the orthogonal transformation of a quadratic form by passage to the limit as n increases without bound.

At the end of Section I we obtained the formulas

$$[x, x] = \frac{D(\ell^{(1)}, x)}{\ell^{(1)}d'(\ell^{(1)})} + \frac{D(\ell^{(2)}, x)}{\ell^{(2)}d'(\ell^{(2)})} + \cdots + \frac{D(\ell^{(n)}, x)}{\ell^{(n)}d'(\ell^{(n)})}$$

$$\frac{D(\ell^{(h)}, x)}{\ell^{(h)}d'(\ell^{(h)})} = \frac{[\varphi^{(h)}, x]^2}{[\varphi^{(h)}, \varphi^{(h)}]}, \quad (h = 1, 2, \dots, n).$$

The last formula shows that every term in the sum on the right-hand side of the expression for $[x, x]$ is positive. Hence, when m is any integer less than n , we have the inequality

$$\frac{D(\ell^{(m+1)}, x)}{\ell^{(m+1)}d'(\ell^{(m+1)})} + \frac{D(\ell^{(m+2)}, x)}{\ell^{(m+2)}d'(\ell^{(m+2)})} + \cdots + \frac{D(\ell^{(n)}, x)}{\ell^{(n)}d'(\ell^{(n)})} \leq [x, x]. \quad (41)$$

Because

$$|[\varphi^{(h)}, x][\varphi^{(h)}, y]| \leq \frac{1}{2}([\varphi^{(h)}, x]^2 + [\varphi^{(h)}, y]^2)$$

we necessarily have

$$\frac{D(\ell^{(h)}, x)}{\ell^{(h)}d'(\ell^{(h)})} \leq \frac{1}{2} \left(\frac{D(\ell^{(h)}, x)}{\ell^{(h)}d'(\ell^{(h)})} + \frac{D(\ell^{(h)}, y)}{\ell^{(h)}d'(\ell^{(h)})} \right).$$

Hence it follows from (41) that

$$\left| \frac{D(\ell^{(m+1)}, x)}{\ell^{(m+1)}d'(\ell^{(m+1)})} \right| + \left| \frac{D(\ell^{(m+2)}, x)}{\ell^{(m+2)}d'(\ell^{(m+2)})} \right| + \cdots + \left| \frac{D(\ell^{(n)}, x)}{\ell^{(n)}d'(\ell^{(n)})} \right| \leq \frac{1}{2}([x, x] + [y, y]),$$

and further that the absolute value of the sum of the last $n - m$ terms in the right-hand side of the formula (14) is not greater than

$$\frac{1}{2|\ell^{(m+1)}|}([x, x] + [y, y]).$$

Therefore, in view of (14) we also have

$$\left| Kxy - \frac{D(\ell^{(1)}, x)}{(\ell^{(1)})^2d'(\ell^{(1)})} + \frac{D(\ell^{(2)}, x)}{(\ell^{(2)})^2d'(\ell^{(2)})} + \cdots + \frac{D(\ell^{(m)}, x)}{(\ell^{(m)})^2d'(\ell^{(m)})} \right| \quad (42)$$

$$\leq \frac{1}{2|\ell^{(m+1)}|}([x, x] + [y, y]).$$

As we have already done earlier, we wish to insert

$$K_{pq} = K\left(\frac{p}{n}, \frac{q}{n}\right) \quad x_p = x\left(\frac{p}{n}\right), \quad y_p = y\left(\frac{p}{n}\right),$$

into the above formula and then take the limit for $n = \infty$ after dividing by n with m held constant. Let us recall the following limits:

$$\begin{aligned} \lim_{n=\infty} \frac{1}{n^2} Kxy &= \lim_{n=\infty} \frac{1}{n^2} K\left(\frac{p}{n}, \frac{q}{n}\right) x_p y_p \\ &= \int_0^1 \int_0^1 K(s, t) x(s) y(t) ds dt, \\ \lim_{n=\infty} n\ell^{(h)} &= \lambda^{(h)}, \\ \lim_{n=\infty} \frac{[x, x]}{n} &= \int_0^1 (x(s))^2 ds, \quad \lim_{n=\infty} \frac{[y, y]}{n} = \int_0^1 (y(s))^2 ds \end{aligned}$$

and also note that the by Lemmas 1 and 2 the expressions $\frac{1}{n} D\left(\frac{\lambda}{n}, \frac{x}{y}\right)$ and $\frac{1}{n} d'\left(\frac{\lambda}{n}\right)$ converge uniformly to $\Delta\left(\lambda, \frac{x}{y}\right)$ and $\delta'(\lambda)$ for all λ less than a fixed bound. Then the inequality (42) in the limit becomes.

$$\begin{aligned} &\left| \int_0^1 \int_0^1 K(s, t) x(s) y(s) ds dt - \frac{\Delta\left(\lambda^{(1)}, \frac{x}{y}\right)}{(\lambda^{(1)})^2 \delta'(\lambda^{(1)})} - \frac{\Delta\left(\lambda^{(2)}, \frac{x}{y}\right)}{\delta'(\lambda^{(2)})^2 \delta'(\lambda^{(2)})} - \dots \right. \\ &\quad \left. - \frac{\Delta\left(\lambda^{(m)}, \frac{x}{y}\right)}{\delta'(\lambda^{(m)})^2 \delta'(\lambda^{(m)})} \right| \leq \frac{1}{2\lambda^{(m+1)}} \left(\int_0^1 (x(s))^2 ds + \int_0^1 (y(s))^2 ds \right). \end{aligned} \quad (43)$$

We now use the fact that if there are infinitely many of the eigenvalues $\lambda^{(m)}$ then their absolute values increase unboundedly with increasing m . Hence with the help of (39) we obtain the following theorem, in which we have replaced the limits of integration 0, 1 with the more general limits a, b .

THEOREM. *Let the kernel $K(s, t)$ of then integral equation of the second kind*

$$f(s) = \varphi(s) - \lambda \int_a^b K(s, t) \varphi(t) dt$$

be a symmetric continuous function of s and t . Moreover, let $\lambda^{(h)}$ be the the eigenvalues of $K(s, t)$ and $\psi^{(h)}$ be the corresponding normalized eigenfunctions. Finally, let $x(s)$ and $y(s)$ be any continuous functions of s . Then we have the expansion

$$\begin{aligned} \int_a^b \int_a^b K(s, t) x(s) y(t) ds dt &= \frac{1}{\lambda^{(1)}} \int_a^b \psi^{(1)}(s) x(s) ds \cdot \int_a^b \psi^{(1)}(s) y(s) ds \\ &+ \frac{1}{\lambda^{(2)}} \int_a^b \psi^{(2)}(s) x(s) ds \cdot \int_a^b \psi^{(2)}(s) y(s) ds + \dots, \end{aligned} \quad (44)$$

in which the right-hand side converges absolutely and uniformly for all functions $x(s)$ and $y(s)$ for which the integrals

$$\int_a^b (x(s))^2 ds, \quad \int_a^b (y(s))^2 ds$$

remain below a fixed, finite bound.

For $x(s) = y(s)$, this is the theorem that corresponds to the algebraic theorem, mentioned in I, concerning the transformation of a quadratic form into the sum of squares of linear forms.

The following are some immediate consequences of this theorem.

The same eigenvalues $\lambda^{(h)}$ and eigenfunctions $\psi^{(h)}(s)$ cannot belong to another kernel that is different from $K(s, t)$. *On the contrary, the $\lambda^{(h)}$ and $\psi^{(h)}$ in their entirety completely determine the kernel $K(s, t)$.*

If we replace $y(t)$ in the formula of the theorem by the integral $\int_a^b K(r, t)y(r) dr$ then in view of (40) we get the following formula:

$$\begin{aligned} \int_a^b \int_a^b KK(s, t)x(s)y(s) ds dt &= \frac{1}{(\lambda^{(1)})^2} \int_a^b \psi^{(1)}(s)x(s) ds \cdot \int_a^b \psi^{(1)}(s)y(s) ds \\ &+ \frac{1}{(\lambda^{(2)})^2} \int_a^b \psi^{(2)}(s)x(s) ds \cdot \int_a^b \psi^{(2)}(s)y(s) ds + \dots, \end{aligned}$$

where for short we have written

$$KK(s, t) = \int_a^b K(s, r)K(t, r) dr.$$

This function $KK(s, t)$ may be called the *twofold combination of $K(s, t)$* . From (44) we see that the twofold combination of $K(s, t)$ has the same eigenfunctions as $K(s, t)$, while the eigenvalues are the squares of the eigenvalues of $K(s, t)$.

This is also a suitable place for a generalization of the formula (29). We will use the notation $K(\lambda; s, t)$ to express the dependence of the solution function $K(s, t)$ on the parameter λ and set

$$F(s, t) = K(\lambda; s, t) - K(\mu; s, t) + (\mu - \lambda) \int_a^b K(\lambda; r, s)K(\mu; r, t) dr$$

as a temporary abbreviation. Then by repeated application of (29) we get the identity

$$F(s, t) - \lambda \int_a^b F(r, s)K(r, t) dr = 0.$$

It follows from a remark at the conclusion of II that $F(s, t)$ vanishes for any value of λ that is different from the eigenvalues $\lambda^{(h)}$. Hence $F(s, t)$ is of necessity identically zero for all arguments λ, μ, s , and t ; i.e., the general formula

$$\mathbf{K}(\lambda; s, t) - \mathbf{K}(\mu; s, t) = (\lambda - \mu) \int_a^b \mathbf{K}(\lambda; r, s) \mathbf{K}(\mu; r, t) dr. \quad (45)$$

is valid.

We can also write this formula in the form

$$\mathbf{K}(\mu; s, t) = \mathbf{K}(\lambda + \mu; s, t) - \lambda \int_a^b \mathbf{K}(\lambda + \mu; r, s) \mathbf{K}(\mu; r, t) dr. \quad (46)$$

Hence it follows that if take $\mathbf{K}(\mu; s, t)$ as the kernel of an integral equation of the second kind, the corresponding solution function is necessarily $\mathbf{K}(\lambda + \mu; s, t)$. At the same time we find that

$$\int_a^b \psi^{(h)}(t) \mathbf{K}(\lambda; s, t) dt = \frac{\psi^{(h)}(s)}{\lambda^{(h)} - \lambda}.$$

From this we see that the same eigenfunctions that belong to the the kernel $K(s, t)$ also belong to the kernel $\mathbf{K}(\mu; s, t)$ while the corresponding eigenvalues are the quantities $\lambda^{(h)} - \mu$.

IV.

The Expansion of an arbitrary Function in Eigenfunctions.

The first important application of the theorem proved in Section III is to answer the question of the existence of the eigenvalues $\lambda^{(h)}$. This question is of particular interest because the corresponding special problem in the theory of linear partial differential equations — namely, to establish the existence of certain important values of parameters appearing in the differential equation or the boundary conditions — has previously been fraught with fundamental difficulties. The far more general problem of the existence of eigenvalues belonging to an integral equation is resolved in a simple and complete manner by the application of our theorems.

Specifically, if we we assume that there are no eigenvalues or only a finite number, say m , then the series (44) appearing in our theorem is finite with m terms. Since the formula (44) of this theorem is valid for all continuous functions, it necessarily follows that

$$K(s, t) = \frac{1}{\lambda^{(1)}} \psi^{(1)}(s) \psi^{(1)}(t) + \dots + \frac{1}{\lambda^{(m)}} \psi^{(m)}(s) \psi^{(m)}(t).$$

In other words, if we regard one variable, say t , as a parameter and assign any constant value to it, $K(s, t)$ may be represented by only m independent functions of the other

variable s . Conversely, if $K(s, t)$ has this property, then, as one sees, all coefficients of the power series for $\delta(\lambda)$ multiplying powers of λ greater than m vanish; that is, $\delta(\lambda)$ becomes an entire rational function, and there are, therefore, only m eigenvalues. Hence we can state the following theorem.

THEOREM 3. *The eigenvalues belonging to $K(s, t)$ are always available in infinite number. For suppose that $K(s, t)$ is representable as a finite sum of products, each of whose factors depends only on the variables s and t . In this case, the number of eigenvalues is equal to the number of the terms in that sum and $\delta(\lambda)$ is an entire rational function of degree equal to that number.*

We now turn to the problem of the expansion of an arbitrary function in an infinite series consisting of eigenfunctions. If in the formula (44) of our theorem we set

$$y(t) = K(r, t)$$

and set

$$f(r) = \int_a^b \int_a^b K(s, t)K(r, t)x(s) ds dt,$$

then, considering that according to (40)

$$\int_a^b f(r)\psi^{(m)}(r) dr = \frac{1}{(\lambda^{(m)})^2} \int_a^b x(s)\psi^{(m)}(s) ds,$$

we find that the formula (44) of our theorem becomes

$$f(r) = \int_a^b f(s)\psi^{(1)}(s) ds \cdot \psi^{(1)}(r) + \int_a^b f(s)\psi^{(2)}(s) ds \cdot \psi^{(2)}(r) + \dots .$$

In other words we have the following theorem.

THEOREM 4. *If a function $f(s)$ can be represented in the form*

$$f(s) = \int_a^b \int_a^b K(r, t)K(s, t)h(r) dr dt,$$

where $h(r)$ is a continuous function of r , then it can be expanded in a Fourier-like series of eigenfunctions as follows:

$$f(s) = c_1\psi^{(1)}(s) + c_2\psi^{(2)}(s) + \dots ,$$

$$c_m = \int_a^b f(s)\psi^{(m)}(s) ds.$$

This series converges absolutely and uniformly.

The assumption about $f(s)$ made in this theorem is equivalent to the requirement that there is a continuous function $h(s)$ such that the integral representation

$$f(s) = \int_a^b K(s,t)h(t) dt$$

holds. This is further equivalent to the requirement that there are two continuous functions $g(s)$ and $h(s)$ such that

$$\begin{aligned} f(s) &= \int_a^b K(s,t)g(t) dt, \\ g(s) &= \int_a^b K(s,t)h(t) dt. \end{aligned}$$

When $K(s,t)$ is a symmetric function of s and t for which the equation

$$\int_a^b K(s,t)g(s) ds = 0$$

is never satisfied for all t by any continuous, nonzero function $g(s)$, then $K(s,t)$ is called a *closed* kernel. It is easy to see from Theorem 3 that a closed kernel always has infinitely many eigenvalues. Moreover we can make the following assertion about a closed kernel.

THEOREM 5. *Let $K(s,t)$ be a closed kernel and let $\psi^{(m)}(s)$ be its eigenfunctions. Then if $h(s)$ is a continuous function such that for all m the equation*

$$\int_a^b h(s)\psi^{(m)}(s) ds = 0$$

holds, then $h(s)$ is identically zero.

To prove this theorem, we set

$$\begin{aligned} g(s) &= \int_a^b K(s,t)h(t) dt, \\ f(s) &= \int_a^b K(s,t)g(t) dt. \end{aligned}$$

By Theorem 4, $f(s)$ admits an expansion in the eigenfunctions $\psi^{(m)}(s)$, and in fact one obtains for the coefficients

$$\begin{aligned} c_m &= \int_a^b f(s)\psi^{(m)}(s) ds = \frac{1}{\lambda^{(m)}} \int_a^b g(s)\psi^{(m)}(s) ds = \\ &= \frac{1}{(\lambda^{(m)})^2} \int_a^b h(s)\psi^{(m)}(s) ds = 0. \end{aligned}$$

It follows that $f(s)$ is identically zero. Since $K(s, t)$ is assumed to be a closed kernel, it follows first that $g(s) = 0$ and then that $h(s) = 0$.

THEOREM 6. *Let $K(s, t)$ be a closed kernel and $f(s)$ any continuous function. If it happens that the Fourier-like series*

$$c_1\psi^{(1)}(s) + c_2\psi^{(2)}(s) + \cdots ,$$

$$c_m = \int_a^b f(s)\psi^{(m)}(s) ds,$$

converges uniformly, then it represents the function $f(s)$.

In fact, by Theorem 5 it turns out that the difference between $f(s)$ and the function of s represented by the series is zero.

In Theorems 4 and 6 we have established certain conditions for the expandability of an arbitrary function in terms of eigenfunctions. We can significantly simplify the conditions of Theorem 4, if we make a certain assumption about the kernel $K(s, t)$. Specifically, we will call a symmetric continuous function $K(s, t)$ a *general kernel* if for every continuous function $g(s)$ and every arbitrarily small positive ϵ , it is always possible to determine a continuous function $h(s)$ such that if we set

$$x(s) = g(s) - \int_a^b K(s, t)h(t) dt$$

then the inequality

$$\int_a^b (x(s))^2 < \epsilon$$

holds. In other words, the kernel $K(s, t)$ is said to be *general* provided that the integral $\int_a^b K(s, t)h(t) dt$ can approximate, in the sense given above, any continuous function $g(s)$ by a suitable choice of the continuous function $h(t)$. In this case, we have the following theorem.

THEOREM 7. *If $K(s, t)$ is a general kernel, then any function that can be represented by means of a continuous function $g(s)$ in the form*

$$f(s) = \int_a^b K(s, t)g(t) dt$$

is expandable in a series of eigenfunctions as follows:

$$f(s) = c_1\psi^{(1)}(s) + c_2\psi^{(2)}(s) + \cdots ,$$

$$c_m = \int_a^b f(s)\psi^{(m)}(s) ds.$$

The series converges absolutely and uniformly.

For the proof we denote by ϵ any arbitrarily small positive quantity and by M the maximum of the function

$$\int_a^b (K(s, t))^2 dt$$

for s in the interval from a to b . Since $K(s, t)$ is a general kernel and $g(s)$ is a continuous function, we can find a continuous function $h(s)$ such that if we set

$$x(s) = g(s) - \int_a^b K(s, t)h(t) dt$$

then the inequality

$$\int_a^b (x(s))^2 ds < \left(\frac{2\epsilon}{3(1+M)} \right) \quad (47)$$

is satisfied. We set

$$g^*(s) = \int_a^b K(s, t)h(t) dt,$$

$$f^*(s) = \int_a^b K(s, t)g^*(t) dt.$$

By Theorem 4 the function $f^*(s)$ has the following series expansion in eigenfunction:

$$f^*(s) = c_1^* \psi^{(1)}(s) + c_2^* \psi^{(2)}(s) + c_3^* \psi^{(3)}(s) + \dots$$

Because of the uniform and absolute convergence of this series it is certainly possible to find an integer m such that for all s

$$|f^*(s) - c_1^* \psi^{(1)}(s) - c_2^* \psi^{(2)}(s) - \dots - c_m^* \psi^{(m)}(s)| < \frac{\epsilon}{3}, \quad (48)$$

and moreover the inequalities that result from replacing m with a larger integer also hold.

Now

$$\left| \int_a^b K(s, t)x(t) dt \right| \leq \sqrt{\int_a^b (K(s, t))^2 dt \cdot \int_a^b (x(t))^2 dt}$$

and in view of (47) the right-hand quantity above is

$$\leq \sqrt{M} \frac{2\epsilon}{3(1+M)} \leq \frac{\epsilon}{3}.$$

Because

$$f(s) = f^*(s) + \int_a^b K(s, t)x(t) dt, \quad (49)$$

we have the inequality

$$|f(s) - f^*(s)| \leq \frac{\epsilon}{3} \quad (50)$$

On the other hand, because of (49)

$$c_j - c_j^* = \int_a^b \int_a^b K(s, t) \psi^{(j)}(s) x(t) ds dt = \frac{1}{\lambda^{(j)}} \int_a^b \psi^{(j)}(t) x(t) dt,$$

and therefore

$$(c_j - c_j^*) \psi^{(j)}(s) = \int_a^b \psi^{(j)}(s) x(s) ds \cdot \int_a^b K(s, t) \psi^{(j)}(t) dt. \quad (51)$$

If we set

$$A = \frac{\int_a^b \psi^{(j)}(s) x(s) ds}{\sqrt{\int_a^b (x(s))^2 ds}}, \quad B = \sqrt{\int_a^b (x(s))^2 ds} \cdot \int_a^b K(s, t) \psi^{(j)}(t) dt,$$

then because

$$|AB| \leq \frac{1}{2}(A^2 + B^2)$$

the inequality

$$|(c_j - c_j^*) \psi^{(j)}(s)| \leq \frac{1}{2} \left\{ \frac{\left(\int_a^b \psi^{(j)}(s) x(s) ds \right)^2}{\int_a^b (x(s))^2 ds} + \sqrt{\int_a^b (x(s))^2 ds} \cdot \left(\int_a^b K(s, t) \psi^{(j)}(t) dt \right)^2 \right\} \quad (52)$$

follows from (51).

We now return to the the formula (16). Since every term on the right hand side of (16) is ≥ 0 , we have the inequality

$$\frac{D(\ell^{(1)}, x)}{\ell^{(1)} d'(\ell^{(1)})} + \frac{D(\ell^{(2)}, x)}{\ell^{(2)} d'(\ell^{(2)})} + \dots + \frac{D(\ell^{(m)}, x)}{\ell^{(m)} d'(\ell^{(m)})} \leq [x, x].$$

If we now suppose, as before, that that we have substituted

$$K_{pq} = K\left(\frac{p}{n}, \frac{q}{n}\right) \quad x_p = x\left(\frac{p}{n}\right),$$

in this formula and then after division by n we have taken the limit for $n = \infty$ while m is held constant, then we get the inequality

$$\begin{aligned} & \left\{ \int_0^1 \psi^{(1)}(s) x(s) \right\}^2 + \left\{ \int_0^1 \psi^{(2)}(s) x(s) \right\}^2 + \dots \\ & + \left\{ \int_0^1 \psi^{(m)}(s) x(s) \right\}^2 \leq \int_0^1 (x(s))^2 ds. \end{aligned} \quad (53)$$

If we sum (52) for $i = 1, 2, \dots, m$, then an application of the last inequality, in which we assume that the integration limits from a to b have been restored, we get

$$\sum_{j=1, \dots, m} |(c_j - c_j^*)\psi^{(j)}(s)| \leq \frac{1}{2} \sqrt{\int_a^b (x(s))^2 ds} \cdot (1 + M).$$

In view of (47) this last expression is

$$\leq \frac{1}{2} \frac{2\epsilon}{3(1 + M)} (1 + M) = \frac{1}{3}\epsilon;$$

i.e.,

$$|c_1\psi^{(1)}(s) + \dots + c_m\psi^{(m)}(s) - c_1^*\psi^{(1)}(s) - \dots - c_m^*\psi^{(m)}(s)| \leq \frac{\epsilon}{3}. \quad (54)$$

From (48), (50), and (54) it follows that for all s

$$|f(s) - c_1\psi^{(1)}(s) - c_2\psi^{(2)}(s) - \dots - c_m\psi^{(m)}(s)| < \epsilon,$$

and it is readily seen that this inequality continues to hold when one chooses a larger value of m in the left hand side. This completes the proof of our theorem.

On the basis of Theorem 7, which was just proved, we can also show that the infinite series

$$\left(\int_a^b \psi^{(1)}(s)x(s) ds \right)^2 + \left(\int_a^b \psi^{(2)}(s)x(s) ds \right)^2 + \dots$$

converges and has the value

$$\int_a^b (x(s))^2 ds.$$

Here $K(s, t)$ is assumed to be a general kernel and $x(s)$ denotes an arbitrary continuous function.

V.

The Variational Problem that Corresponds to the Algebraic Question of the Minima and Maxima of a Quadratic Form

The theory developed in Sections III–IV has special significance for the calculus of variations. Here I would like to treat just the transcendental problem that corresponds to the algebraic question of the relative maxima and minima of a quadratic form when a second different form is held constant. Specifically, the problem is to find the function $x(s)$ for which the double integral

$$J(x) = \int_a^b \int_a^b K(s, t)x(s)x(t) ds dt$$

has a minimal or maximal value, assuming that the side condition

$$\int_a^b (x(s))^2 ds = 1. \quad (55)$$

is satisfied.

If the kernel $K(s, t)$ has the property that the integral $J(x)$ has only positive values for continuous $x(s)$ and is zero only when $x(s) = 0$, we will say that the kernel is *definite*. In what follows we make the assumption that $K(s, t)$ is a definite kernel.

If for some continuous function $x(s)$ we have

$$\int_a^b K(s, t)x(t) dt = 0$$

identically in s , then obviously $J(x) = 0$ and hence $x(s)$ is also zero; that is, *a definite kernel is also a closed kernel*. It must therefore have infinitely many eigenvalues and eigenfunctions.

The eigenvalues of a definite kernel are always positive. For if on the contrary some eigenvalue $\lambda^{(h)}$ were negative, it follows from

$$J(x) = \frac{1}{\lambda} \left\{ \int_a^b \psi^{(1)}(s)x(s) dx \right\}^2 + \frac{1}{\lambda} \left\{ \int_a^b \psi^{(2)}(s)x(s) dx \right\}^2 = \dots \quad (56)$$

that for $x(s) = \psi^{(h)}(s)$ the value of the double integral $J(x)$ would be negative.

The following theorems concern the minima and maxima of $J(x)$.

THEOREM 8. *There is no continuous function satisfying the side condition (55) that causes $J(x)$ to assume a minimum.*

In fact, the eigenfunctions $\psi^{(1)}(s), \psi^{(2)}(s), \dots$ all satisfy the side condition (55). Because

$$J(\psi^{(1)}) = \frac{1}{\lambda^{(1)}}, J(\psi^{(2)}) = \frac{1}{\lambda^{(2)}}, \dots,$$

the minimum we seek can only be equal to zero. But $J(x)$ obtains this value only for $x(s) = 0$.

THEOREM 9. *The largest value the double integral $J(x)$ assumes when $x(s)$ is a continuous function satisfying the side condition (55) is $\frac{1}{\lambda^{(1)}}$. The double integral assumes this value for $x(s) = \psi^{(1)}(s)$.*

If, on the contrary, there were a function $x(s)$ which satisfied the side condition (55) and for which

$$J(x) > \frac{1}{\lambda^{(1)}},$$

then there we would have to be able to choose an integer m such that the sum $S(x)$ of the first m terms in the right-hand side of (56) is greater than $\frac{1}{\lambda(1)}$. We now set

$$x(s) = c_1\psi^{(1)}(s) + c_2\psi^{(2)}(s) + \cdots + c_m\psi^{(m)}(s) + y(s),$$

where for short we have written

$$c_h = \int_a^b \psi^{(h)}(s)x(s) ds \quad (h = 1, 2, \dots, m).$$

Hence

$$\int_a^b \psi^{(h)}(s)y(s) ds = 0 \quad (h = 1, 2, \dots, m).$$

We then easily see that

$$\int_a^b (x(s))^2 ds = c_1^2 + c_2^2 + \cdots + c_m^2 + \int_a^b (y(s))^2 ds \quad (57)$$

and

$$S(x) = \frac{c_1^2}{\lambda(1)} + \frac{c_2^2}{\lambda(2)} + \cdots + \frac{c_m^2}{\lambda(m)}. \quad (58)$$

In view of (55) it follows from (57) that

$$c_1^2 + c_2^2 + \cdots + c_m^2 \leq 1,$$

and perforce we have

$$\frac{c_1^2}{\lambda(1)} + \frac{c_2^2}{\lambda(2)} + \cdots + \frac{c_m^2}{\lambda(m)} \leq \frac{1}{\lambda(1)}.$$

This equation contradicts (58), since $S(x)$ was to be greater than $\frac{1}{\lambda(1)}$; the original assumption therefore does not hold.

In an analogous manner we see that the following more general theorem is true.

THEOREM 10. *The largest value that the double integral $J(x)$ attains when $x(s)$ is a continuous function satisfying the side conditions*

$$\begin{aligned} \int_a^b (x(s))^2 ds &= 1, \\ \int_a^b \psi^{(h)}(s)x(s) ds &= 0, \quad (h = 1, 2, \dots, m-1) \end{aligned}$$

is $\frac{1}{\lambda(m)}$. *The double integral attains this value for $x(s) = \psi^{(m)}(s)$.*

By similar arguments we can also obtain the solutions of more general maximal problems. For example, one can find with no essential difficulty the function $x(s)$ that maximizes $J(x)$ when, in addition to the side condition (55), the side condition

$$\int_a^b f(s)x(s) ds = 0 \quad (59)$$

is satisfied, where $f(s)$ denotes a given function.

The kernel $K(s, t)$ is said to be *relatively definite*, if it has the property that it [i.e., $J(x)$] takes on only positive values when $x(s)$ is a continuous function satisfying the side condition (59).

At most one of the eigenvalues of a relative definite kernel is negative. For if, say, $\lambda^{(1)}$ and $\lambda^{(2)}$ were negative, one could determine constants c_1 and c_2 such that the function

$$x(s) = c_1\psi^{(1)}(s) + c_2\psi^{(2)}(s)$$

satisfies the side condition (59) and in addition $c_1^2 + c_2^2 = 1$. Then by (56), $J(x)$ must be negative.

VI.

Supplementing and Extending the Theory.

Up to now in Sections I–V, we have consistently stipulated that $K(s, t)$ be a *continuous* function of the variables s and t . Our next problem is to ascertain to what extent this assumption can be relaxed.

We will say that $K(s, t)$ has *singularities of order less than $\frac{1}{2}$* whenever there are a finite number of analytical lines L of the forms

$$s = F(t) \quad \text{or} \quad t = G(s)$$

in the s, t -plane for which there is a positive exponent α less than $\frac{1}{2}$ such that the product

$$(s - F(t))^\alpha K(s, t) \quad \text{or} \quad (t - G(s))^\alpha K(s, t)$$

is continuous. We also assume that $K(s, t)$ is continuous off the lines L . We can now make the following assertion.

All the results proved in Sections III–V also hold provided the kernel $K(s, t)$ of our basic integral equation has singularities of order less than $\frac{1}{2}$. At the same time the functions $x(s)$ and $y(s)$ appearing in our theory may also have singularities of order less than $\frac{1}{2}$ at a finite number of points, provided that they are otherwise continuous.

The following is the method by which we see the truth of this assertion. We cover the lines L by a set of strips in the s, t -plane of arbitrarily small width ϵ . We then

construct a function $K_\epsilon(s, t)$ that is zero in the domain of strips and outside is identical to $K(s, t)$. The function $K_\epsilon(s, t)$ is everywhere continuous with the exception of the boundary lines of the set of strips, where clearly jump discontinuities are encountered. Our earlier proofs are valid without any changes for a kernel like $K_\epsilon(s, t)$, whose values are everywhere less than a finite bound, and are discontinuous in certain lines. To see the validity for the kernel $K(s, t)$ requires application of a passage to the limit $\epsilon = 0$. In what follows we will show how this can be effected.

To this end, we first turn to the power series $\delta(\lambda)$ (p. 64) and $\Delta\left(\lambda, \begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ (p. 65). The coefficients δ_h and $\Delta_h\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ cannot be formed if $K(s, s)$ regarded as a function of s has no meaning; i.e. whenever the line $s = t$ or a part of it belongs to the singular lines of the kernel. We will take care of this difficulty by introducing the the formulas

$$\delta_h = \frac{1}{h!} \int_0^1 \cdots \int_0^1 \begin{vmatrix} 0 & K(s_1, s_2) & \cdots & K(s_1, s_h) \\ K(s_2, s_1) & 0 & \cdots & K(s_2, s_h) \\ \cdots & \cdots & \cdots & \cdots \\ K(s_h, s_1) & K(s_h, s_2) & \cdots & 0 \end{vmatrix} ds_1 \cdots ds_h$$

and

$$\Delta_h\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) = \frac{1}{h!} \int_0^1 \cdots \int_0^1 \begin{vmatrix} 0 & x(s_1) & x(s_2) & \cdots & x(s_h) \\ y(s_1) & 0 & K(s_1, s_2) & \cdots & K(s_1, s_h) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y(s_h) & K(s_h, s_1) & K(s_h, s_2) & \cdots & 0 \end{vmatrix} ds_1 \cdots ds_h$$

in place of the formulas used earlier for δ_h (p. 64) and $\Delta_h\left(\begin{smallmatrix} s \\ t \end{smallmatrix}\right)$ (p. 65). It is seen that the new expressions for δ_h and $\Delta_h\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ differ from the originals in that the diagonal elements of the determinants are everywhere zero. The power series $\delta(\lambda)$ and $\Delta\left(\lambda, \begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ formed from the new coefficients agree with the originals up to an negligible exponential factor. This factor is the same for $\delta(\lambda)$ and $\Delta\left(\lambda, \begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ and drops out when the quotient $\frac{\Delta\left(\lambda, \begin{smallmatrix} x \\ y \end{smallmatrix}\right)}{\delta(\lambda)}$ is formed.¹²

LEMMA 3. *For our kernel $K(x, y)$ the new expressions δ_h and $\Delta_h\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ are uniquely defined and the power series $\delta(\lambda)$ and $\Delta\left(\lambda, \begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ formed from them are everywhere convergent.*

For simplicity we will prove this lemma only for the case that $s = t$ is the sole singular line of $K(s, t)$. In the h -fold integral the variables of integration s_1, \dots, s_h are to run through all values between 0 and 1. We first consider the $1/h!$ th part of the

¹²See Kellogg, Zur Theorie der Integralgleichungen, §5. Göttinger Nachr. 1902.

h -dimensional domain of integration characterized by the inequalities

$$s_1 > s_2 > \cdots > s_h.$$

In the determinant with h rows that appears in the expression for δ_h we suppose that the elements of the

first	row are multiplied by	$ s_1 - s_2 ^\alpha$
second	row are multiplied by	$\{ s_1 - s_2 ^{-\alpha} + s_2 - s_3 ^{-\alpha}\}^{-1}$
third	horizontal row are multiplied by	$\{ s_2 - s_3 ^{-\alpha} + s_3 - s_4 ^{-\alpha}\}^{-1}$
.....		
h th	row are multiplied by	$ s_{h-1} - s_h ^\alpha$

As can easily be seen, we obtain a determinant the absolute values of whose elements for all values of the variables are less than a finite positive quantity K . The value of this determinant is certainly less than or equal to $\sqrt{h^h}K^h$. Hence, we have the following upper bound for the h -fold integral taken over T :

$$\sqrt{h^h}K^h \int \cdots \int (s_1 - s_2)^{-\alpha} \{|s_1 - s_2|^{-\alpha} + |s_2 - s_3|^{-\alpha}\} \{|s_3 - s_2|^{-\alpha} + |s_2 - s_3|^{-\alpha}\} \cdots (s_{h-1} - s_h)^{-\alpha} ds_1 ds_2 \cdots ds_h, \tag{60}$$

$$1 > s_1 > s_2 > \cdots > s_h > 0.$$

If in this h -fold integral we introduce the new variables

$$s_1 - s_2 = \sigma_1, s_2 - s_3 = \sigma_3, \dots, s_{h-1} - s_h = \sigma_h, s_h = \sigma_h$$

and multiply out the product under the integral sign, we find that this integral is composed of 2^{h-2} h -fold integrals of the form

$$\int \cdots \int \sigma_1^{\alpha_1} \sigma_2^{\alpha_2} \cdots \sigma_h^{\alpha_h} d\sigma_1 d\sigma_2 \cdots d\sigma_h \left(\begin{array}{l} \sigma_1 > 0, \sigma_2 > 0, \dots, \sigma_h > 0 \\ \sigma_1 + \sigma_2 + \cdots + \sigma_h < 1 \end{array} \right) \tag{61}$$

Here the exponents $\alpha_1, \alpha_2, \dots, \alpha_h$ have the values 0, $-\alpha$ or -2α while their sum $\alpha_1 + \alpha_2 + \cdots + \alpha_h$ is always equal to $-h\alpha$. A calculation involving the integral (61) gives the following expression for an upper bound for the integral:

$$\frac{A^h}{\Gamma(1 + h - \alpha h)} < \frac{B^h}{h^{h(1-\alpha)}}.$$

Here A, B denote certain positive quantities that are independent of h . From this we get an upper bound

$$\frac{C^h}{h^{h(\frac{1}{2}-\alpha)}} \tag{62}$$

for (60). Here again C is a positive quantity that is independent of h . The expression (62) is also an upper bound for the part of the h -fold integral that appears in δ_h taken over T . Since all the remaining $h! - 1$ parts of this h -fold integral have the same values (as can be seen by interchanging the integration variables), it follows that the complete h -fold integral that appears in δ_h has the product of (62) by $h!$ as an upper bound; i.e.,

$$|\delta_h| \leq \frac{C^h}{h^{h(\frac{1}{2}-\alpha)}}. \quad (63)$$

From this follows the correctness of Lemma 3 with respect to $\delta(\lambda)$ since $\alpha < \frac{1}{2}$.¹³

The same proof strategy works for the power series $\Delta\left(\lambda, \frac{x}{y}\right)$.

We now return to the function $K_\epsilon(s, t)$. If we recall how $K_\epsilon(s, t)$ was formed from $K(s, t)$ by the elimination of singularities, we see that $K_\epsilon(s, t)$ is to be considered dependent on the strip width ϵ . Since for a fixed ϵ the absolute value $K_\epsilon(s, t)$ always stays below a finite bound, our earlier theory is valid without alteration for $K_\epsilon(s, t)$. We denote the power series in λ belonging to $K_\epsilon(s, t)$ by $\delta_\epsilon(\lambda)$ and $\Delta_\epsilon\left(\lambda, \frac{x}{y}\right)$. The inequality (63) and the corresponding inequality for $\Delta_h\left(\lambda, \frac{x}{y}\right)$, obviously hold for the coefficients of the power series $\delta_\epsilon(\lambda)$ and $\Delta_\epsilon\left(\lambda, \frac{x}{y}\right)$. Hence by the same techniques used in the proof of Lemma 1 we see the correctness of the following facts.

LEMMA 4. *The functions $\delta_\epsilon(\lambda)$ and $\Delta_\epsilon\left(\lambda, \frac{x}{y}\right)$ converge for $\epsilon = 0$ to $\delta(\lambda)$ and $\Delta\left(\lambda, \frac{x}{y}\right)$. Moreover, the convergence is uniform for all values of λ whose absolute value lies below an arbitrarily chosen positive bound Λ .*

After these preliminaries, it is not difficult to extend the validity of our basic theorem (p. 74) to the case where the kernel $K(s, t)$ has singularities of order less than $\frac{1}{2}$.

We already know that our theorem is valid for the kernel $K_\epsilon(s, t)$ provided that the zeros of the associated function $\delta_\epsilon(\lambda)$ are all simple. Should this assumption not hold for a kernel $K_\epsilon(s, t)$, we imagine the kernel as being slightly modified — as was done at the beginning of Section III — so that the assumption is satisfied and so that the modified kernel still converges uniformly to the same limits $K(s, t)$, $\delta(\lambda)$, and $\Delta\left(\lambda, \frac{x}{y}\right)$.

Now let Λ be any positive quantity. From Lemma 4 we can conclude that the zeros $\lambda_\epsilon^{(h)}$ of $\delta_\epsilon(\lambda)$ whose absolute values remain less than Λ in the limit for $\epsilon = 0$ go over in the limit for $\epsilon = 0$ to the zeros $\lambda^{(h)}$ of $\delta(\lambda)$ whose absolute values lie below Λ . Moreover, the value of $\Delta_\epsilon\left(\lambda_\epsilon^{(h)}, \frac{x}{y}\right)$ belonging to the zero $\lambda_\epsilon^{(h)}$ goes over to the corresponding value of $\Delta\left(\lambda^{(h)}, \frac{x}{y}\right)$.

¹³The presentation of this proof in the Dissertations of Kellogg and Andrae cited earlier is incorrect.

We denote by $\psi_\epsilon^{(1)}(s), \psi_\epsilon^{(2)}(s), \dots$ the eigenfunctions belonging to the kernel $K_\epsilon(s, t)$. By (53) (p. 81) for every value of m however large we have the inequality

$$\left\{ \int_a^b \psi_\epsilon^{(1)} x(s) ds \right\}^2 + \left\{ \int_a^b \psi_\epsilon^{(2)} x(s) ds \right\}^2 + \dots \\ + \left\{ \int_a^b \psi_\epsilon^{(m)} x(s) ds \right\}^2 \leq \int_a^b (x(s))^2 ds.$$

It then follows that

$$\sum_{(\lambda_\epsilon^{(h)} \geq \Lambda)} \left\{ \int_a^b \psi_\epsilon^{(h)} x(s) ds \right\}^2 \leq \int_a^b (x(s))^2 ds. \quad (64)$$

We now replace $y(s)$ by $x(s)$ in formula (44) of our theorem and write the resulting formula in the form

$$\int_a^b \int_a^b K_\epsilon(s, t) x(s) x(t) ds dt = \sum_{(\lambda_\epsilon^{(h)} < \Lambda)} \frac{1}{\lambda_\epsilon^{(h)}} \left\{ \int_a^b \psi_\epsilon^{(h)} x(s) ds \right\}^2 \\ + \sum_{(\lambda_\epsilon^{(h)} \geq \Lambda)} \frac{1}{\lambda_\epsilon^{(h)}} \left\{ \int_a^b \psi_\epsilon^{(h)} x(s) ds \right\}^2 \quad (65)$$

Here the first sum on the right-hand side is to be taken over all eigenfunctions whose corresponding eigenvalues are less in absolute value than Λ , while the second sum on the right hand side contains all remaining terms, just like the sum on the left-hand side of (64). Because of (64) it follows from (65) that

$$\int_a^b \int_a^b K_\epsilon(s, t) x(s) x(t) ds dt = \sum_{(\lambda_\epsilon^{(h)} < \Lambda)} \frac{1}{\lambda_\epsilon^{(h)}} \left\{ \int_a^b \psi_\epsilon^{(h)} x(s) ds \right\}^2 \\ \pm \frac{\vartheta}{\Lambda} \int_a^b (x(s))^2 ds, \quad (0 \leq \vartheta \leq 1).$$

From this equation we get by passage to the limit for $\epsilon = 0$

$$\int_a^b \int_a^b K(s, t) x(s) x(t) ds dt = \sum_{(\lambda^{(h)} < \Lambda)} \frac{1}{\lambda^{(h)}} \left\{ \int_a^b \psi^{(h)} x(s) ds \right\}^2 \\ \pm \frac{\vartheta}{\Lambda} \int_a^b (x(s))^2 ds, \quad (0 \leq \vartheta \leq 1).$$

If Λ is now allowed to grow without bound, we get the formula (44) in our theorem for the case that $x(s) = y(s)$. This last limitation can easily be disposed of.

We can see without difficulty that all the consequences of our theorem are valid — in particular, Theorems 4 and 7 on the expansion of arbitrary functions in the eigenfunctions of $K(s, t)$.

Should the kernel $K(s, t)$ have singular lines of order greater than $\frac{1}{2}$ but less than one, our theorems require certain modifications. This is easily seen if one forms the two-fold and higher-fold combinations of $K(s, t)$. Considering that among these kernels there must always exist kernels for which the theory presented above is valid, we obtain the desired conclusions for the kernel $K(s, t)$.

Previously — even in the development of this section IV — we have consistently made the assumption that for our basic kernel $K(s, t)$ the power series $\delta(\lambda)$ has only *simple* zeros. We must now determine the modifications our theory must undergo when we drop this hypothesis.

To this end, let $K(s, t)$ be a kernel for which $\lambda^{(h)}$ is an n_h -fold eigenvalue, i. e., $\lambda^{(h)}$ is a n_h -fold zero of $\delta(\lambda)$. Then there are no substantial, fundamental difficulties in finding a kernel $K_\mu(s, t)$ with the following properties. The kernel $K_\mu(s, t)$ is a power series in μ , whose coefficients are continuous functions of s and t . The power series converges for sufficiently small values of μ , so that for $\mu = 0$ becomes $K(s, t)$. Let $\delta_\mu(\lambda)$ be the usual power series belonging to $K_\mu(s, t)$, so that for $\mu = 0$, $\delta_\mu(\lambda)$ converges to the power series $\delta(\lambda)$ belonging to $K(s, t)$. As can be easily seen from the earlier proof, $\delta_\mu(\lambda)$ is a power series in λ that converges for all λ and sufficiently small μ . The convergence is uniform for all λ whose absolute value lies below a finite bound Λ and for all sufficiently small μ . Thus $\delta_\mu(\lambda)$ is represented by power series in λ and μ . Finally, let the parameter μ so restricted that the in a neighborhood of $\lambda = \lambda^{(h)}$, the equation

$$\delta_\mu(\lambda) = 0$$

has the following n_h solutions:

$$\begin{aligned} \lambda_\mu^{(h)} &= \beta(\mu) \\ \lambda_\mu^{(h+1)} &= \beta_1(\mu) \\ &\dots \\ \lambda_\mu^{h+n_h-1} &= \beta_{n_h-1}(\mu) \end{aligned} \tag{66}$$

Here $\beta(\mu), \beta_1(\mu), \dots, \beta_{n_h-1}(\mu)$ denote power series in μ and *among these no two are identically equal in μ* . This last stipulation gives substance to a property of the function $K_\mu(s, t)$ that will be essential for the following development: namely, for all sufficiently small, nonzero value of the parameter μ , the function $K_\mu(s, t)$ represents a kernel that has only simple eigenvalues.

In the manner of (27), we now construct for $K_\mu(s, t)$ the power series $\Delta_\mu(\lambda; s, t)$, which for $\mu = 0$ becomes the power series $\Delta(\lambda; s, t)$ belonging to $K(s, t)$. Just like $\delta_\mu(\lambda)$, the power series $\Delta_\mu(\lambda; s, t)$ is uniformly convergent for all λ and sufficiently small μ and, moreover, can be represented as a power series in λ and μ . Finally we construct for $K_\mu(s, t)$ the normalized eigenfunctions

$$\psi_\mu^{(h)}(s), \psi_\mu^{(h+1)}(s), \dots, \psi_\mu^{(h+n_h)}(s)$$

corresponding to

$$\lambda_\mu^{(h)}, \lambda_\mu^{(h+1)}, \dots, \lambda_\mu^{(h+n_h+1)},$$

by replacing in turn λ in $\delta_\mu(\lambda)$ and $\Delta_\mu(\lambda; s, t)$ by the values of $\lambda_\mu^{(h)}, \lambda_\mu^{(h+1)}, \dots, \lambda_\mu^{(h+n_h)-1}$ from (66) regarded as power series in μ . First of all we get

$$\begin{aligned} \psi_\mu^{(h)}(s)\psi_\mu^{(h)}(t) &= \frac{\Delta_\mu(\lambda^{(h)}; s, t)}{\delta'_\mu(\lambda^{(h)})} = \mu^{\pm e} \{ \Psi^{(h)}(s, t) + \Psi_1^{(h)}(s, t)\mu + \dots \}, \\ \psi_\mu^{(h+1)}(s)\psi_\mu^{(h+1)}(t) &= \frac{\Delta_\mu(\lambda^{(h+1)}; s, t)}{\delta'_\mu(\lambda^{(h+1)})} = \mu^{\pm e_1} \{ \Psi^{(h+1)}(s, t) + \Psi_1^{(h+1)}(s, t)\mu + \dots \}, \\ &\dots\dots\dots \\ \psi_\mu^{(h+n_h-1)}(s)\psi_\mu^{(h+n_h-1)}(t) &= \frac{\Delta_\mu(\lambda^{(h+n_h-1)}; s, t)}{\delta'_\mu(\lambda^{(h+n_h-1)})} = \mu^{\pm e_{n_h-1}} \{ \Psi^{(h+n_h-1)}(s, t) \\ &\quad + \Psi_1^{(h+n_h-1)}(s, t)\mu + \dots \}. \end{aligned}$$

Here e, e_1, \dots, e_{n_h-1} are rational exponents that are greater than or equal to zero and the expressions $\Psi(s, t)$ on the right-hand side denote continuous functions of s and t . From this it is not difficult to derive formulas of the following kind for the eigenfunction we seek:

$$\begin{aligned} \psi_\mu^{(h)}(s) &= \mu^{\pm f} \{ \psi^{(h)}(s) + \psi_1^{(h+1)}(s)\mu + \dots \}, \\ &\dots\dots\dots \\ \psi_\mu^{(h+n_h-1)}(s) &= \mu^{\pm f_{n_h-1}} \{ \psi^{(h+n_h-1)}(s) + \psi_1^{(h+n_h-1)}(s)\mu + \dots \}. \end{aligned} \tag{67}$$

Here, as before, f, f_1, \dots, f_{n_h-1} are certain rational exponents that are greater than or equal to zero. Moreover, the expressions $\psi^{(j)}(s)$ on the right-hand side denote continuous functions of s , and in particular we may assume that among the functions

$$\psi^{(h)}(s), \psi^{(h+1)}(s), \dots, \psi^{(h+n_h-1)}(s) \tag{68}$$

none are identically zero in s . On the other hand, since for all sufficiently small, nonzero μ , the equations

$$\int_a^b (\psi_\mu^{(h)})^2 ds = 1, \dots, \int_a^b (\psi_\mu^{(h+n_h-1)})^2 ds = 1,$$

must be satisfied, it follows that the exponents f, f_1, \dots, f_{n_h-1} must all be zero. Then the formula (67) shows that for $\mu = 0$ the functions

$$\psi_\mu^h(s), \psi_\mu^{h+1}(s), \dots, \psi_\mu^{h+n_h-1}(s),$$

become the functions in (68). These functions are called *the eigenfunctions corresponding to the n_h -fold eigenvalue $\lambda^{(h)}$* . By passing to the limit for $\mu = 0$ in the formula for the eigenfunctions $\psi^{(h)}, \psi^{(h+1)}, \dots$ we see that the eigenfunctions corresponding to $\lambda^{(h)}$ satisfy the following equations:

$$\begin{aligned} \int_a^b (\psi^{(k)})^2 ds &= 1, \\ \int_a^b \psi^{(k)} \psi^{(k')} ds &= 0, \\ (k, k' &= h, h+1, \dots, h+n_h-1; k' \neq k'). \end{aligned}$$

We will now apply the formula (43) to the kernel $K_\mu(s, t)$, where μ denotes a sufficiently small nonzero value. Considering the formula (39), we get

$$\begin{aligned} \left| \int_0^1 \int_0^1 K_\mu(s, t) x(s) y(t) ds dt - \frac{1}{\lambda_\mu^{(1)}} \int_0^1 \psi_\mu^{(1)} x(s) ds \int_0^1 \psi_\mu^{(1)} y(s) ds \right. \\ \left. - \frac{1}{\lambda_\mu^{(2)}} \int_0^1 \psi_\mu^{(2)} x(s) ds \int_0^1 \psi_\mu^{(2)} y(s) ds - \dots \right. \\ \left. - \frac{1}{\lambda_\mu^{(m)}} \int_0^1 \psi_\mu^{(m)} x(s) ds \int_0^1 \psi_\mu^{(m)} y(s) ds \right| \\ \leq \frac{1}{2|\lambda_\mu^{(m+1)}|} \left(\int_0^1 (x(s))^2 ds + \int_0^1 (y(s))^2 ds \right). \end{aligned}$$

If we take the limit for $\mu = 0$ and then let m grow unboundedly, we see that for the kernel $K(s, t)$, formula (44) of our fundamental theorem remains valid without change. *For the case of an n_h -fold eigenvalue we have only to take into account, one after the other, the n_h different eigenfunctions corresponding to λ^h , so that in each of these n_h terms the reciprocal of the same eigenvalue $\lambda^{(h)}$ appears as a factor.*

We can find a simple method for calculating the eigenfunctions (68) by proceeding from the formula

$$\int_a^b \int_a^b K(\lambda : s, t) x(t) y(t) = \sum_{(h=1,2,\dots)} \frac{1}{\lambda^{(h)} - \lambda} \int_a^b \psi^{(h)}(s) x(s) ds \int_a^b \psi^{(h)}(s) y(s) ds.$$

If we set

$$K(s, t) = -\frac{\delta(\lambda; s, t)}{\delta(\lambda)},$$

multiply the formula by $\lambda - \lambda^h$, and pass to the limit $\lambda = \lambda^{(h)}$, we conclude that

$$\left[\frac{\frac{\partial^{n_h-1}}{\partial \lambda^{n_h-1}} \Delta(\lambda; s, t)}{\frac{\partial^{n_h-1}}{\partial \lambda^{n_h-1}} \delta(\lambda)} \right]_{\lambda=\lambda_k} = \psi^{(h)}(s)\psi^{(h)}(t) + \psi^{(h+1)}(s)\psi^{(h+1)}(t) + \dots + \psi^{(h+n_h-1)}(s)\psi^{(h+n_h-1)}(t).$$

The eigenfunctions (68) corresponding to the n_h -fold eigenvalues are uniquely determined by this equation, provided we ignore trivial orthogonal combinations with constant coefficients of the eigenvectors.

By means of the generalization of our basic theorem that we have just proved we are in a position to easily handle the remaining theorems in the case of multiple eigenvalues.

In a second communication I will treat some applications of the above theory of integral equations of the second kind to the theory of linear ordinary and partial differential equations.

On the Theory of Linear and Nonlinear Integral Equations.

Part I: The Expansion of Arbitrary Functions by Prescribed Systems.¹

by

ERHARD SCHMIDT at Bonn.

Introduction.

Fredholm² discovered a formula for the solution of the inhomogeneous linear integral equation

$$f(s) = \varphi(s) - \lambda \int_a^b K(s, t)\varphi(t) dt.$$

This formula has the consequence that if λ is not a zero of a certain entire transcendental function $\delta(\lambda)$ then the equation can always be solved for $\varphi(s)$. As Fredholm went on to show, for these and *only* these values of λ —in Hilbert's nomenclature the so-called *eigenvalues of the kernel* $K(s, t)$ —the homogeneous equation

$$0 = \varphi(s) - \lambda \int_a^b K(s, t)\varphi(t) dt$$

admits a solution, which, following Hilbert, we will call the *eigenfunction corresponding to the eigenvalue* λ *of the kernel* $K(s, t)$. In the theory of partial and ordinary differential equations, Hilbert³ used Green's function to reduce the important question of the existence of so-called normal functions and the expandability of arbitrary functions in terms of them to a far more general problem: namely that of establishing the existence of eigenfunctions for a *symmetric* kernel $K(s, t)$ and setting down the conditions for the

¹Except for the newly added Chapter IV, § 13 and some minor alterations in the remaining chapters, this part is a reprint of my Götting Inaugural Dissertation, which appeared in July, 1905.

²Acta Mathematica, V. 27.

³Nachrichten der K. Gesellschaft der Wissenschaften zu Göttingen. Mathem.-Phys. Cl. 1904 Number 3.

expandability of arbitrary functions in terms of them. The Green's function itself is determined by an integral equation with an unsymmetric kernel, which is how Fredholm's formula comes into play. Now by considering the double integral

$$\int_a^b \int_a^b K(s, t)x(s)y(t) ds dt,$$

where $x(s)$ and $y(t)$ are arbitrary continuous functions, as a quadratic form in infinitely many variables, Hilbert⁴ obtained by passage to the limit the decomposition

$$\int_a^b \int_a^b K(s, t)x(s)y(t) ds dt = \sum_{\nu} \frac{1}{\lambda_{\nu}} \int_a^b x(s)\varphi_{\nu}(s) ds \cdot \int_a^b y(t)\varphi_{\nu}(t) dt,$$

a formula that corresponds to the canonical orthogonal decomposition of a quadratic form. Here the $\varphi_{\nu}(s)$ range over all eigenfunctions of the kernel—each scaled so that the integral of their squares is one—and the λ_{ν} range over the corresponding eigenvalues. An immediate consequence of this result is that any symmetric kernel has eigenfunctions. Under the assumption that the kernel is *general*—that is, for any continuous function $\alpha(s)$ and any arbitrarily small positive number ϵ , there is a function $\beta(s)$ such that

$$\int_a^b \left\{ \alpha(s) - \int_a^b K(s, t)\beta(t) dt \right\}^2 ds < \epsilon$$

holds—under this assumption Hilbert then derived the fundamental expansion theorem that any function $g(s)$ that can be represented in terms of a continuous function $h(t)$ by the integral

$$g(s) = \int_a^b K(s, t)h(t) dt$$

can be expanded in a absolutely and uniformly convergent series of the eigenfunctions of the kernel $K(s, t)$. In addition, these theorems imply their analogues for the integral equation

$$0 = \psi(s) - \lambda \int_a^b G(s, t)p(t)\psi(t) dt,$$

where $G(s, t)$ is symmetric and $p(t) > 0$. For this equation can be reduced to the one above, namely

$$0 = \varphi(s) - \lambda \int_a^b K(s, t)\varphi(t) dt,$$

⁴Nachrichten der K. Gesellschaft der Wissenschaften zu Göttingen. Mathem.-Phys. Cl. 1904 Number 1.

by the substitutions

$$\sqrt{p(s)} \cdot \psi(s) = \varphi(s), \quad \sqrt{p(s)} \cdot G(s, t) \cdot \sqrt{p(t)} = K(s, t).$$

Stekloff⁵ has obtained a series of related and partially equivalent theorems by means of the Schwarz-Poincaré method, which he greatly extended.

In the first chapter of this work we establish some supporting lemmas. In the second we give very simple proofs of Hilbert's theorems that avoid passing to the limit from algebraic theorems. First, the existence of eigenvalues is established by a method based on a famous proof of H. A. Schwarz,⁶ which in the language of Fredholm's formula amounts to solving the equation

$$\delta(\lambda) = 0$$

by Bernoulli's method. The expansion theorem then follows from the existence theorem in a manner analogous to the way the expansion of an entire function in a product of linear factors follows from the fundamental theorem of algebra. In this connection, it turns out that the validity of Hilbert's theorem is unconditional; in particular, it does not require the "generality" of the kernel postulated by Hilbert. The decomposition theorem of Hilbert, mentioned above, that corresponds to the canonical orthogonal decomposition of a quadratic form is obtained immediately from the expansion theorem by integration. The complications caused by multiple zeros of the function $\delta(\lambda)$ do not arise in the method of proof given here. Fredholm's formula is not used, and more important the unrestricted validity of the expansion theorem for the case of symmetric kernels gives a new representation of the solution of the inhomogeneous linear integral equation.⁷ Also any *unsymmetric* linear integral equation can be reduced to a *symmetric* equation by a simple substitution, as is shown in § 13.

In the third chapter we drop the assumption that the kernel is symmetric and define the functions $\varphi_\nu(s)$ and $\psi_\nu(s)$ to be a pair of adjoint eigenfunctions of the kernel $K(s, t)$ corresponding to the eigenvalue λ if the equations

$$\varphi_\nu(s) = \lambda_\nu \int_a^b K(s, t) \psi_\nu(t) dt$$

$$\psi_\nu(s) = \lambda_\nu \int_a^b K(t, s) \varphi_\nu(t) dt$$

⁵Mémoires de l'Académie des Sciences de Saint-Pétersbourg 1904, p.7 ff. Annales del la Fac. de Toulouse 2^e S., VI 1905.

⁶H. A. Schwarz, Gesammelte Abhandlungen V. 1, pp. 241-262.

⁷cf. Hilbert's new, comprehensive development—published while present work was in press—of the theory of integral equations based on his theory of quadratic forms in infinitely many variables. Göttinger Nachrichten 1906, fourth and fifth communications.

are satisfied. The function $\varphi_\nu(s)$ may be called an eigenfunction of the first kind; $\psi_\nu(s)$, an eigenfunction of the second kind. The expansion theorem then takes the following form. If the continuous function $g(s)$ can be represented in terms of the continuous function $h(s)$ by the integral

$$g(s) = \int_a^b K(s, t)h(t) dt,$$

then $g(s)$ can be expanded in an absolutely and uniformly convergent series consisting of eigenfunctions of the first kind. If

$$g(s) = \int_a^b K(t, s)h(t) dt,$$

then $g(s)$ can be expanded in the same way in a series of eigenfunctions of the second kind. From this theorem we get by integration the decomposition theorem corresponding to the orthogonal decomposition of bilinear forms:

$$\int_a^b \int_a^b K(s, t)x(s)y(t) ds dt = \sum_\nu \frac{1}{\lambda_\nu} \int_a^b x(s)\varphi_\nu(s) ds \cdot \int_a^b y(t)\psi_\nu(t) dt.$$

To the best of my knowledge, the theorems of this third chapter were unknown before now.

The expansion of functions of two variables in powers, in trigonometric functions, in Kugel functions, and in many other functions can be written in the form of a series whose terms are products of a function of one variable with a function of the other variable.

A question in the calculus of variations that arises in connection with this last observation is the object of the fourth chapter. Let $K(s, t)$ be a continuous function of two variables s and t . We want to find a system of at most m pairs of continuous functions, one in s and one in t , such that the sum of their products approximates the given function $K(s, t)$ as well as possible. As usual we will define the measure of the approximation, whose minimum is required by the statement of the problem, to be the double integral of the square of the error. We will show that the solution of the problem is formed from the first m pairs of adjoint eigenfunction of the kernel $K(s, t)$. In consequence, the measure M_m of the best approximation is given by the formula

$$M_m = \int_a^b \int_a^b (K(s, t))^2 ds dt - \sum_{\nu=1}^{\nu=m} \frac{1}{\lambda_\nu^2},$$

where the λ_ν are the first m eigenvalues of the the kernel $K(s, t)$. It follows that the measure of the best approximation vanishes with increasing m .

All theorems and proofs in the first four chapters remain valid when s and t are points from a bounded n -dimensional domain consisting of a finite number of analytic pieces in an $(n + m)$ -dimensional space. Here ds and dt are the corresponding [differential] elements.

The fifth chapter assumes nothing from the second, third, and fourth chapters—only the lemmas established in the first chapter. The theory of expansion of functions in powers and polynomials, in Fourier series and infinite series of finite trigonometric series, in Kugel functions, and in normal functions of partial and ordinary differential equations suggests the following question. We are given an infinite series $\varphi_1(x), \varphi_2(x), \dots, \varphi_\nu(x), \dots$ of continuous, real valued functions defined in the interval $a \leq x \leq b$. What are the conditions under which any continuous function defined in the interval $a \leq x \leq b$ can be expanded in a uniformly convergent series of the functions $\varphi_\nu(x)$ or of finite linear combinations of them? In other words, under what conditions can any continuous function defined in the interval $a \leq x \leq b$ be uniformly approximated by functions from the system of functions that are formed from the sequence of the $\varphi_\nu(x)$ by the operations of multiplying by a constant and addition? And when such is the case, how can we determine the coefficients of such an expansion?

Let us call the system of function φ_ν in question *closed* if there is no continuous nonzero function $f(x)$ such that for every ν

$$\int_a^b f(x)\varphi_\nu(x) dx = 0.$$

Then it is clear that the *closure* of the system of the ϕ_ν represents a necessary condition for the solution of our problem.⁸ For otherwise all expandable functions would have to satisfy the condition that their products with $f(x)$ integrate to zero from a to b , and this condition would not be satisfied by, e.g., $f(x)$ itself and all sufficiently small perturbations of $f(x)$.

We will then show that *just as the closure of the given system of functions is a necessary condition for the solution of our problem, the closure of the system of its second derivatives is sufficient*, provided, if need be, the functions 1 and x are adjoined. *If the function to be represented is also continuously differentiable, then the coefficients of the representation have simple formulas that are valid in complete generality.*

In a paper that is a direct continuation of the present paper, we will treat a new and very simple method for solving unsymmetric linear integral equations. The underlying principles of this method also permit the treatment of nonlinear integral equations, which is the subject of the second part of this paper. By a nonlinear equation I understand a functional equation that defines the unknown function by requiring that a given function be equal to a convergent infinite series whose terms are formed from

⁸J. P. Gram has already made this observation in Crelles Journal v.94, p.94.

the unknown function and other given functions by the operations of integration and multiplication, and hence the operation of taking positive integer powers. Thus, for example,

$$f(s) - \iint K(s, t, r) (\varphi(t))^m (\varphi(r))^n dt dr = 0$$

is a nonlinear integral equation. Here $\varphi(s)$ is unknown and $f(s)$ and $K(s, t, r)$ are given. Now the ordinary nonlinear equation

$$y = f(x)$$

admits one and only one solution in the neighborhood of a solution provided $f'(x)$ does not vanish, but otherwise branching [into two or more solutions] occurs. Analogously, the nature of the solution of a nonlinear integral equation in the neighborhood of a solution depends on a derived linear integral equation. If the Fredholm denominator $\delta(\lambda)$ of this latter equation does not vanish, then the nonlinear equation in the neighborhood has one and only one solution. But if $\delta(\lambda)$ vanishes, *functional* branching occurs, for which it suffices to invoke the equivalent of Priseux's theorems.

In the case of nonlinear elliptical partial differential equation of the second order, for example, these theorems make it possible to track the dependence of the solution surfaces on the boundary values—and with complete knowledge of branching; that is, of solutions in whose neighborhood even for arbitrarily small perturbation of the boundary values there is no longer a single solution but several. The nature of the branching depends on whether or not the Jacobian linearization of the differential equation has nonzero solutions for zero boundary values. This can be determined by considering a linear integral equation.

In particular, the bifurcation, discovered by Poincaré, in the theory of rotating balanced figures is this kind of branching of a nonlinear integral equation. In a third paper I will give a detailed exposition of this and several other applications.

Chapter I

Preliminary Results on Orthogonal Functions.

§ 1.

The Bessel and Schwarz Inequalities.

Let $\psi_1(x), \psi_2(x), \dots, \psi_n(x)$ be continuous real functions defined in the interval $a \leq x \leq b$ that are pairwise *orthogonal*; i.e., for each pair of distinct indices μ and ν the equation

$$\int_a^b \psi_\mu(x) \psi_\nu(x) dx = 0$$

is satisfied. Furthermore, let all the functions be *normalized*; i.e., for each ν the equation

$$\int_a^b (\psi_\nu(x))^2 dx = 1$$

is satisfied. Then Bessel's identity

$$\int_a^b \left(f(x) - \sum_{\nu=1}^{\nu=n} \psi_\nu(x) \int_a^b f(y) \psi_\nu(y) dy \right)^2 dx = \int_a^b (f(x))^2 dx - \sum_{\nu=1}^{\nu=n} \left(\int_a^b f(y) \psi_\nu(y) dy \right)^2,$$

$$\sum_{\nu=1}^{\nu=n} \left(\int_a^b f(y) \psi_\nu(y) dy \right)^2 \leq \int_a^b (f(x))^2 dx.$$

holds for any real continuous function $f(x)$. If the sequence of pairwise orthogonal, normalized functions is infinite, then the last inequality implies the convergence of the sum

$$\sum_{\nu=1}^{\nu=\infty} \left(\int_a^b f(y) \psi_\nu(y) dy \right)^2,$$

since all the terms in the sum are positive.

Now let $f(x)$ and $\varphi(x)$ be two real continuous functions. If we set $\psi_1(x) = \frac{\varphi(x)}{\sqrt{\int_a^b (\varphi(y))^2 dy}}$, then $\psi_1(x)$ is normalized, and Bessel's inequality for the case $n = 1$ gives

$$\left(\int_a^b f(x) \psi_1(x) dx \right)^2 \leq \int_a^b (f(x))^2 dx$$

$$\left(\int_a^b f(x) \varphi(x) dx \right)^2 \leq \int_a^b (f(x))^2 dx \cdot \int_a^b (\varphi(x))^2 dx.$$

This is the well-known *Schwarz inequality*. Bessel's identity and all results derived from it in this section remain valid when $f(x)$ is a real integrable function whose square when integrated from a to b gives a finite value. In this case, the finiteness and existence of

$$\int_a^b f(x) \psi_\nu(x) dx$$

follows from the inequality

$$f(x) \cdot \psi_\nu(x) \leq (f(x))^2 + (\psi_\nu(x))^2.$$

§ 2

A Convergence Theorem

Let $Q(z, x)$ be a real function defined in the domain $a \leq x \leq b$, $a \leq z \leq b$ that is integrable with respect to x , and suppose that for all $a \leq z \leq b$

$$\int_a^b (Q(z, x))^2 dx \leq A,$$

where A denotes a constant. Further let $\psi_1(x), \psi_2(x), \dots, \psi_\nu(x), \dots$ be an infinite sequence of real, continuous functions that, in the nomenclature of the last section, are normalized and are pairwise orthogonal. If $f(x)$ denotes an arbitrary real integrable function whose square integrated from a to b is finite, then the series

$$\sum_{\nu=1}^{\nu=\infty} \int_a^b f(y)\psi_\nu(y) dy \cdot \int_a^b Q(z, x)\psi_\nu(x) dx = \sum_{\nu=1}^{\nu=\infty} U_\nu(z)$$

converges absolutely and uniformly for $a \leq z \leq b$. In fact,

$$\sum_{\nu=n}^{\nu=\infty} |U_\nu(z)| \leq 2\sqrt{A} \sqrt{\sum_{\nu=n}^{\nu=\infty} \left(\int_a^b f(y)\psi_\nu(y) dy \right)^2},$$

and because of the convergence of the series $\sum_{\nu=n}^{\nu=\infty} \left(\int_a^b f(y)\psi_\nu(y) dy \right)^2$, which was established in the last section, the right hand side of the above expression vanishes with increasing n .

PROOF. Let

$$\sum_{\nu=n}^{\nu=n+m} |U_\nu(z)| = \sum_k U_k(z) - \sum_\rho U_\rho(z),$$

where k runs through the the indices $n, n+1, \dots, n+m$ for which the terms of the sum are positive at the particular value of z , and ρ runs through the indices corresponding to negative terms.

If the sum and the integral on the left hand side are interchanged, it follows from the inequality of Schwarz, which was given in the last section, that

$$\sum_k U_k(z) \leq \sqrt{\int_a^b (Q(z, x))^2 dx} \cdot \sqrt{\int_a^b \left(\sum_k \psi_k(x) \int_a^b f(y)\psi_k(y) dy \right)^2 dx}.$$

Since the functions are pairwise orthogonal and normalized,

$$\int_a^b \left(\sum_k \psi_k(x) \int_a^b f(y) \psi_k(y) dy \right)^2 dx = \sum_\nu \left(\int_a^b f(y) \psi_\nu(y) dy \right)^2$$

$$\leq \sum_{\nu=n}^{\nu=\infty} \left(\int_a^b f(y) \psi_\nu(y) dy \right)^2,$$

and hence

$$\sum_k U_k(z) \leq \sqrt{A} \sqrt{\sum_{k=n}^{k=\infty} \left(\int_a^b f(y) \psi_k(y) dy \right)^2}.$$

We get the same inequality for

$$- \sum_\rho U_\rho(z),$$

and the inequality to be established follows upon adding these two inequalities.

COROLLARY. If we take $Q(z, x)$ to be the discontinuous function that is +1 for $x \leq z$ and 0 for $x > z$, it follows that the series

$$\sum_{\nu=1}^{\nu=\infty} \int_a^b f(y) \psi_\nu(y) dy \cdot \int_a^z \psi_\nu(x) dx$$

is absolutely and uniformly convergent for $a \leq z \leq b$.

§ 3.

The Replacement of Linearly Independent Systems of Functions by Orthogonal Systems

Let $\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x)$ be n continuous, real functions defined for $a \leq x \leq b$

that are assumed to be linearly independent. Then we construct the functions^{*})

$$\begin{aligned}\psi_1(x) &= \frac{\varphi_1(x)}{\sqrt{\int_a^b (\varphi_1(y))^2 dy}} \\ \psi_2(x) &= \frac{\varphi_2(x) - \psi_1(x) \int_a^b \varphi_2(z) \psi_1(z) dz}{\sqrt{\int_a^b (\varphi_2(y) - \psi_1(y) \int_a^b \varphi_2(z) \psi_1(z) dz)^2 dy}} \\ &\vdots \\ \psi_n(x) &= \frac{\varphi_n(x) - \sum_{\rho=1}^{\rho=n-1} \psi_\rho(x) \int_a^b \varphi_n(z) \psi_\rho(z) dz}{\sqrt{\int_a^b (\varphi_n(y) - \sum_{\rho=1}^{\rho=n-1} \psi_\rho(y) \int_a^b \varphi_n(z) \psi_\rho(z) dz)^2 dy}}.\end{aligned}$$

By means of these formulas, for each ν , $\psi_\nu(x)$ is represented recursively as a linear homogeneous combination of $\varphi_1(x)$, $\varphi_2(x)$, \dots , $\varphi_\nu(x)$ with constant coefficients, and conversely, $\varphi_\nu(x)$ is similarly represented by $\psi_1(x)$, $\psi_2(x)$, \dots , $\psi_\nu(x)$. To wit, no denominator in any of the formulas can vanish. For if ν were the first index for which it happened, we would have to have

$$\varphi_\nu(x) - \sum_{\rho=1}^{\rho=\nu-1} \psi_\rho(x) \int_a^b \varphi_\nu(z) \psi_\rho(z) dz = 0.$$

Since the $\psi_\rho(x)$ can be written as linear, homogeneous combinations of $\varphi_1(x)$, $\varphi_2(x)$, \dots , $\varphi_{\nu-1}(x)$ with constant coefficients, we would then obtain a contradiction of the assumed linear independence of the functions $\varphi_1(x)$, $\varphi_2(x)$, \dots , $\varphi_\nu(x)$. Moreover, *the functions $\psi_1(x)$, $\psi_2(x)$, \dots , $\psi_n(x)$ form a system of orthogonal, normalized functions*; that is, they satisfy the equation

$$\int_a^b \psi_\mu(x) \psi_\nu(x) dx = 0 \quad \text{or} \quad 1,$$

according as μ and ν are distinct or equal. To begin with, this is clear for the functions $\psi_1(x)$ and $\psi_2(x)$. If we now assume the normality and the orthogonality of the system

^{8*)} In essence, J. P. GRAM presented the same formulas in the paper "Ueber die Entwicklung reeler Functionen in Reihen mittelst der Methode der kleinsten Quadrate," Crelles Journal V. 94.

$\psi_1(x), \psi_2(x), \dots, \psi_{\nu-1}(x)$, then the same follows for the system $\psi_1(x), \psi_2(x), \dots, \psi_{\nu-1}(x), \psi_\nu(x)$; for

$$\int_a^b (\psi_\nu(x))^2 dx = 1 \quad \text{and} \quad \int_a^b \psi_\nu(x)\psi_\rho(x) dx = 0 \quad \text{for} \quad \rho \leq \nu - 1.$$

If the functions $\varphi_1(x), \varphi_2(x), \dots, \varphi_{n-1}(x)$ form a linearly independent system but the functions $\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x)$ do not, then the linear dependence of the last sequence is given by the equation

$$\varphi_n(x) - \sum_{\rho=1}^{\rho=n-1} \psi_\rho(x) \int_a^b \varphi_n(z)\psi_\rho(z) dz = 0.$$

For if this expression did not vanish identically, then the functions $\psi_1(x), \psi_2(x), \dots, \psi_n(x)$ could be represented as a linear homogeneous combination with constant coefficients of the functions $\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x)$, and hence—because of the linear dependence of the latter functions—of the functions $\varphi_1(x), \varphi_2(x), \dots, \varphi_{n-1}(x)$. Thus the functions $\psi_1(x), \psi_2(x), \dots, \psi_n(x)$ would have to be linearly dependent. But that is impossible; for in a orthogonal system there can be no equation of the form

$$\sum_{\nu} c_\nu \psi_\nu(x) = 0,$$

unless all the c_ν are zero, a fact that follows on multiplying this equation by $\psi_\nu(x) dx$ and integrating from a to b .

Thus in the numerators of the expressions under consideration we have a sequence of homogeneous linear forms of the functions $\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x)$ with the following property. If the functions are linearly dependent, the vanishing of one of the forms is not just a necessary and sufficient condition indicating the dependence but exhibits the dependence itself.

CONCLUDING REMARK. All the formulas and results of this chapter, with the exception of the corollary in § 2, remain valid when x, y , and z are points from a bounded n -dimensional domain consisting of a finite number of analytic pieces in an $(n+m)$ -dimensional space. In this case dx, dy , and dz are the corresponding [differential] elements.

Chapter II

On the Linear Symmetric Integral Equation.

§ 4.

The Concept of Eigenfunction.

Let $K(s, t)$ be a real continuous function defined for $a \leq s \leq b$, $a \leq t \leq b$ that is symmetric in s and t . If $\varphi(s)$ is any continuous function, real or complex, that does not vanish identically and that satisfies

$$\varphi(s) = \lambda \int_a^b K(s, t)\varphi(t) dt$$

identically in s (here λ is a constant), then $\varphi(s)$ is called an *eigenfunction of the kernel $K(s, t)$* corresponding to the *eigenvalue λ* .

Two eigenfunctions $\varphi_\mu(s)$ and $\varphi_\nu(s)$ corresponding to distinct eigenvalues are mutually orthogonal; i.e., they satisfy the equation

$$\int_a^b \varphi_\mu(s)\varphi_\nu(s) ds = 0.$$

For we have

$$\varphi_\mu(s) = \lambda_\mu \int_a^b K(s, t)\varphi_\mu(t) dt$$

and

$$\varphi_\nu(s) = \lambda_\nu \int_a^b K(s, t)\varphi_\nu(t) dt.$$

If we multiply the first of these equations by $\lambda_\nu\varphi_\nu(s) ds$, the second by $\lambda_\mu\varphi_\mu(s) ds$, integrate from a to b , and subtract, then by the symmetry of $K(s, t)$ we get

$$(\lambda_\nu - \lambda_\mu) \int_a^b \varphi_\mu(s)\varphi_\nu(s) ds = 0,$$

from which the equation to be established follows.

If $\varphi_\nu(s)$ were an eigenfunction of $K(s, t)$ corresponding to a complex eigenvalue, then the function conjugate to $\varphi_\nu(s)$ would correspond to the conjugate eigenvalue. Because these two eigenvalues are distinct, $\varphi_\nu(s)$ and the function conjugate to $\varphi_\nu(s)$ would have to be orthogonal, which is impossible since the integral of the product of two conjugate functions is always greater than 0. Thus, all eigenvalues of the kernel $K(s, t)$ are real.

If $\psi(s)$ is a complex eigenfunction, it follows from the fact that the corresponding eigenvalue must be real that $\psi(s) = \varphi(s) + i\bar{\varphi}(s)$, where $\varphi(s)$ and $\bar{\varphi}(s)$ are *real* eigenfunctions corresponding to the same eigenvalue. Because of this, only *real eigenfunctions* will be treated in the remaining theorems of this chapter, and the term “*eigenfunction*” will refer only to real eigenfunctions.

§ 5.

The Full Normalized Orthogonal System.

The number of linearly independent eigenfunctions corresponding to a particular eigenvalue is finite.

PROOF: Any linear, homogeneous combination of eigenfunctions corresponding to the same eigenvalue yields an eigenfunction that also corresponds to that eigenvalue. Hence if the construction of § 3 is applied to any system of linearly independent eigenfunctions corresponding to the eigenvalue λ , the result is a *normalized and mutually orthogonal* system of just as many eigenfunctions of the same eigenvalue. Denote these by $\varphi_1(s)$, $\varphi_2(s)$, \dots , $\varphi_n(s)$. By Bessel's inequality from § 1, for each s

$$\int_a^b (K(s, t))^2 dt \geq \sum_{\nu=1}^{\nu=n} \left(\int_a^b K(s, t) \varphi_\nu(t) dt \right)^2 = \frac{1}{\lambda^2} \sum_{\nu=1}^{\nu=n} (\varphi_\nu(s))^2.$$

If we multiply this inequality by ds and integrate from a to b , then taking into account the fact that

$$\int_a^b (\varphi_\nu(s))^2 ds = 1$$

we get the relation

$$n \leq \lambda^2 \int_a^b \int_a^b (K(s, t))^2 ds dt,$$

which establishes the assertion.

If the number of linearly independent eigenfunctions corresponding to an eigenvalue is equal to m , the eigenvalue in question is said to be m -fold.

We will call a system of normalized and mutually orthogonal eigenfunctions of the kernel $K(s, t)$ a *full* normalized orthogonal system of the kernel* if each eigenfunction of the kernel can be represented as a homogeneous linear combination with constant coefficients of a finite number of functions of the system.

The functions appearing in the representation of an eigenfunction by functions of the system along with the function itself must all correspond to the same eigenvalue.

*“Complete” is a more natural translation of the German *Vollständig*. However, the term complete has come to be used in a somewhat different sense in functional analysis.

For let

$$\psi(s) = \sum_{\nu} c_{\nu} \varphi_{\nu}(s)$$

be such a representation of the eigenfunction $\psi(s)$ in terms of functions of the system. Then because of the orthogonality of the $\varphi_{\nu}(s)$, we have

$$c_{\nu} = \int_a^b \psi(s) \varphi_{\nu}(s) ds.$$

But as we showed earlier, this expression vanishes when $\psi(s)$ and $\varphi_{\nu}(s)$ corresponding to different eigenvalues.

A full normalized orthogonal system for the kernel $K(s, t)$ may be obtained by applying the construction of § 3 to each eigenvalue λ to form as many normalized and mutually orthogonal eigenfunctions as the multiplicity of λ .

If $\varphi_{\rho}(t)$ runs through an arbitrary finite number of functions from a full normalized orthogonal system of the kernel $K(s, t)$, then the inequality

$$\int_a^b (K(s, t))^2 dt \geq \sum_{\rho} \left(\int_a^b K(s, t) \varphi_{\rho}(t) dt \right)^2 = \sum_{\rho} \frac{1}{\lambda_{\rho}^2} (\varphi_{\rho}(s))^2,$$

or

$$\int_a^b \int_a^b (K(s, t))^2 ds dt \geq \sum_{\rho} \frac{1}{\lambda_{\rho}^2}$$

follows from Bessel's inequality. *From this it follows that the eigenvalues of a kernel $K(s, t)$, numbered according to their multiplicities, cannot have a finite point of accumulation. Therefore, if they are arranged in a sequence according to their absolute values and if there are infinitely many of them, then their absolute values grow without bound.*

§ 6

The Iterated Kernels.*)

We define

$$\begin{aligned} K^1(s, t) &= K(s, t), \\ K^2(s, t) &= \int_a^b K(s, r) K^1(r, t) dr, \\ K^{\nu}(s, t) &= \int_a^b K(s, r) K^{\nu-1}(r, t) dr. \end{aligned}$$

^{8*)} Cf. H. A. Schwartz l.c. Fredholm l.c. p.384. Hilbert l.c. pp.244-247.

If one regards $K^{n+1}(s, t)$ as an n -fold integral of the explicit product of $n + 1$ kernels, then it is obvious that

$$K^{\mu+\nu}(s, t) = \int_a^b K^\mu(s, r)K^\nu(r, t) dr \quad (1)$$

and

$$K^\nu(s, t) = K^\nu(t, s).$$

Moreover, none of the functions $K^n(s, t)$ can vanish identically in s and t . For if it were true that

$$K^n(s, t) = 0,$$

then

$$K^{n+1}(s, t) = 0.$$

According to (1) we would then have

$$\int_a^b K^{n_1}(s, r)K^{n_1}(r, t) dr = 0,$$

where n_1 denotes the one of the two numbers $\frac{n}{2}$ and $\frac{n+1}{2}$ that is an integer. Hence we would have

$$0 = \int_a^b K^{n_1}(s, r)K^{n_1}(r, t) dr = \int_a^b (K^{n_1}(s, r))^2 dr,$$

from which it would follow that K^{n_1} vanishes identically in s and t . By repeating this reduction sufficiently many times, we would find that $K(s, t)$ vanishes identically, contrary to hypothesis.

Let

$$\varphi(s) = \lambda \int_a^b K(s, t)\varphi(t) dt.$$

Then

$$\varphi(s) = \lambda^n \int_a^b K^n(s, t)\varphi(t) dt.$$

Thus any eigenfunction of the kernel $K(s, t)$ is also an eigenfunction of the kernel $K^n(s, t)$.

Conversely, suppose

$$\psi(s) = c \int_a^b K^n(s, t)\psi(t) dt.$$

Let the function $\chi_\nu(s)$ be defined by the equation

$$\begin{aligned} n\chi_\nu(s) = & \psi(s) + h_\nu \int_a^b K(s,t)\psi(t) dt + h_\nu^2 \int_a^b K^2(s,t)\psi(t) dt + \cdots \\ & + h_\nu^{n-1} \int_a^b K^{n-1}(s,t)\psi(t) dt \quad (\nu = 1, 2, 3, \dots, n), \end{aligned}$$

where h_ν runs through the roots of the equation $h^n = c$. Then because $\sum_{\nu=1}^{\nu=n} h_\nu^k$ is nonzero if and only if k is divisible by n , we have

$$\psi(s) = \sum_{\nu=1}^{\nu=n} \chi_\nu(s). \quad (2)$$

Moreover, we have

$$\chi_\nu(s) = h_\nu \int_a^b K(s,t)\chi_\nu(t) dt.$$

Therefore, as long as $\chi_\nu(s)$ does not vanish identically, which according to (2) cannot be the case for all ν , it is an eigenfunction of the kernel $K(s,t)$. Since by § 4 the kernel has only real eigenvalues, $\chi_\nu(s)$ must vanish identically for all nonreal h_ν . Hence if n is odd and one writes $h_1 = \sqrt[n]{c}$ for the real root of the equation $h^n = c$, then

$$\psi(s) = \sqrt[n]{c} \int_a^b K(s,t)\psi(t) dt.$$

On the other hand, if n is even, c must be positive. Thus if we write $h_1 = +\sqrt[n]{c}$ and $h_2 = -\sqrt[n]{c}$ for the two real roots of the equation $h^n = c$, then

$$\begin{aligned} \psi(s) &= \chi_1(s) + \chi_2(s), \\ \chi_1(s) &= +\sqrt[n]{c} \int_a^b K(s,t)\chi_1(t) dt, \\ \chi_2(s) &= -\sqrt[n]{c} \int_a^b K(s,t)\chi_2(t) dt, \end{aligned}$$

which implies that at most one of the two functions $\chi_1(s)$ and $\chi_2(s)$ can vanish identically. *Therefore, if n is odd, each eigenfunction of $K^n(s,t)$ is an eigenfunction of $K(s,t)$. On the other hand, if n is even, each eigenfunction of $K^n(s,t)$ is either an eigenfunction of $K(s,t)$ or is a sum of two such eigenfunctions.*

Any full normalized orthogonal system of the kernel $K(s,t)$ is also a full normalized orthogonal system for the kernel $K^n(s,t)$.

§ 7.

The Fundamental Theorem.

Each kernel $K(s, t)$ that does not vanish identically has at least one eigenfunction.
In order not to disturb the flow of ideas, I will leave the proof for this fundamental theorem for § 11.

§ 8.

Expansion of the Kernel and Its Iterates.

Let the functions $\varphi_1(s), \varphi_2(s), \dots, \varphi_\nu(s), \dots$ form a full normalized orthogonal system of the kernel $K(s, t)$, and let their eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_\nu, \dots$ be ordered according to absolute value. *If the series*

$$\sum_{\nu} \frac{\varphi_\nu(s)\varphi_\nu(t)}{\lambda_\nu}$$

converges uniformly for $a \leq s \leq b, a \leq t \leq b$, then

$$K(s, t) = \sum_{\nu} \frac{\varphi_\nu(s)\varphi_\nu(t)}{\lambda_\nu}. \quad (3)$$

In particular, it follows that this equation is always true if the number of eigenfunctions of the full normalized orthogonal system is finite.

PROOF: We set

$$K(s, t) - \sum_{\nu} \frac{\varphi_\nu(s)\varphi_\nu(t)}{\lambda_\nu} = Q(s, t).$$

Then $Q(s, t)$ is also a continuous symmetric function of s and t , and

$$\int_a^b Q(s, t)\varphi_\nu(t) dt = 0 \quad (4)$$

for all values of ν . Now if $Q(s, t)$ were not identically zero, by the fundamental theorem of the last section there would have to be a continuous function $\psi(s)$ such that

$$\psi(s) = c \int_a^b Q(s, t)\psi(t) dt.$$

From (4) it follows that for all values of ν

$$\int_a^b \psi(s)\varphi_\nu(s) ds = 0. \quad (5)$$

Hence

$$\int_a^b Q(s, t)\psi(t) dt = \int_a^b K(s, t)\psi(t) dt,$$

and

$$\psi(s) = c \int_a^b K(s, t)\psi(t) dt.$$

Thus $\psi(s)$ would be an eigenfunction for the kernel $K(s, t)$ which, by equation (5), is orthogonal to all the functions φ_ν and which cannot be represented by them as a homogeneous linear combination with constant coefficients, contradicting the hypothesis that the functions $\varphi_1(s), \varphi_2(s), \dots, \varphi_\nu(s), \dots$ form a *full* normalized orthogonal system of the kernel $K(s, t)$. Therefore, $Q(s, t)$ is identically zero, which is what was to be shown.

From this we conclude: *It is always true that*

$$K^4(s, t) = \sum_{\nu} \frac{\varphi_{\nu}(s)\varphi_{\nu}(t)}{\lambda_{\nu}^4}, \quad (6)$$

and the series on the right converges absolutely and uniformly. For by § 6 the functions $\varphi_{\nu}(s)$ form a full normalized orthogonal system of the kernel $K^4(s, t)$ with corresponding eigenvalues λ_{ν}^4 . Hence our assertion follows from the one just proved, provided only that the absolute and uniform convergence of the series can be established. But

$$\sum_{\nu=n}^{\nu=n+m} \left| \frac{\varphi_{\nu}(s)\varphi_{\nu}(t)}{\lambda_{\nu}^4} \right| \leq \frac{1}{2\lambda_n^2} \left(\sum_{\nu=n}^{\nu=n+m} \frac{(\varphi_{\nu}(s))^2}{\lambda_{\nu}^2} + \sum_{\nu=n}^{\nu=n+m} \frac{(\varphi_{\nu}(t))^2}{\lambda_{\nu}^2} \right).$$

Since by Bessel's inequality

$$\int_a^b (K(s, t))^2 dt \geq \sum_{\nu=n}^{\nu=n+m} \left(\int_a^b K(s, t)\varphi_{\nu}(t) dt \right)^2 = \sum_{\nu=n}^{\nu=n+m} \frac{(\varphi_{\nu}(s))^2}{\lambda_{\nu}^2},$$

it follows that

$$\sum_{\nu=n}^{\nu=n+m} \left| \frac{\varphi_{\nu}(s)\varphi_{\nu}(t)}{\lambda_{\nu}^4} \right| \leq \frac{1}{2\lambda_n^2} \int_a^b (K(s, t))^2 dt,$$

from which follows the absolute and uniform convergence that was to be established.

§ 9

Expansion of Arbitrary Functions.⁹

As in the previous sections, let the functions $\varphi_1(s), \varphi_2(s), \dots, \varphi_\nu(s), \dots$ form a full normalized orthogonal system for the kernel $K(s, t)$, and let their eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_\nu, \dots$ be ordered according to absolute value. If $h(s)$ is a continuous function such that

$$\int_a^b K(s, t)h(t) dt = 0,$$

then by multiplying this equation by $\varphi_\nu(s)$ and integrating from a to b we get the equation

$$\int_a^b h(s)\varphi_\nu(s) ds = 0,$$

which is valid for all ν . Conversely if this equation holds for all ν , then

$$\int_a^b K(s, t)h(t) dt = 0.$$

PROOF. If the equation

$$K^4(s, t) = \sum_{\nu} \frac{\varphi_\nu(s)\varphi_\nu(t)}{\lambda_\nu^4},$$

which was established in the last section, is multiplied by $h(s)h(t) ds dt$ and integrated with respect to s and t from a to b , it follows that

$$\begin{aligned} 0 &= \int_a^b \int_a^b K^4(s, t)h(s)h(t) ds dt \\ &= \int_a^b dr \int_a^b K^2(s, r)h(s) ds \int_a^b K^2(t, r)h(t) dt \\ &= \int_a^b dr \left(\int_a^b K^2(s, r)h(s) ds \right)^2. \end{aligned}$$

Hence

$$\int_a^b K^2(s, r)h(s) ds = 0$$

⁹Cf. Hilbert l.c. pp.72–78, Stekloff l.c. pp.404–425.

identically in r , and consequently

$$0 = \int_a^b \int_a^b K^2(s, t) h(s) h(t) ds dt.$$

By repeating the reduction we have just used we get

$$\int_a^b K(r, s) h(s) ds = 0$$

identically in r , which is what was to be proved.

Let the continuous function $g(s)$ be represented by the equation

$$g(s) = \int_a^b K(s, t) p(t) dt,$$

where $p(t)$ is a continuous function. Then

$$\begin{aligned} g(s) &= \sum_{\nu} \varphi_{\nu}(s) \int_a^b g(t) \varphi_{\nu}(t) dt = \sum_{\nu} \frac{\varphi_{\nu}(s)}{\lambda_{\nu}} \int_a^b p(t) \varphi_{\nu}(t) dt \\ &= \sum_{\nu} \int_a^b K(s, t) \varphi_{\nu}(t) dt \int_a^b p(t) \varphi_{\nu}(t) dt, \end{aligned}$$

and the series on the right converges absolutely and uniformly.

PROOF. We can deduce the absolute and uniform convergence of the series from the third form of its general term and the convergence theorem proved in § 2.

If we set

$$g(s) - \sum_{\nu} \varphi_{\nu}(s) \int_a^b g(t) \varphi_{\nu}(t) dt = h(s),$$

then for all ν

$$\int_a^b \int_a^b h(s) \varphi_{\nu}(s) ds = 0, \quad (7)$$

and hence by the theorem just established

$$\int_a^b K(s, t) h(t) dt = 0. \quad (8)$$

Now

$$\int_a^b (h(s))^2 ds = \int_a^b h(s) g(s) ds - \sum_{\nu} \int_a^b h(s) \varphi_{\nu}(s) ds \int_a^b g(t) \varphi_{\nu}(t) dt.$$

Since by (7) the sum on the right vanishes, we have by (8)

$$\int_a^b (h(s))^2 = \int_a^b h(s)g(s) ds = \int_a^b p(r)dr \int_a^b K(s,r)h(s) ds = 0.$$

Hence $h(s)$ is identically zero, which is what was to be established.

Let $p(s)$ and $q(s)$ be two continuous functions. If the equation just established is multiplied by $q(s) ds$ and integrated from a to b , the result is

$$\int_a^b \int_a^b K(s,t)q(s)p(t) ds dt = \sum_{\nu} \frac{1}{\lambda_{\nu}} \int_a^b q(s)\varphi_{\nu}(s) ds \int_a^b p(t)\varphi_{\nu}(t) dt.$$

This is Hilbert's fundamental formula, which he obtained from the canonical decomposition of a quadratic form by passage to the limit. From it he then derived the results of § 8 and the first theorem of § 9 for all kernels and the expansion theorem for "general" kernels.

§ 10.

The Inhomogeneous Linear Integral Equation.

Given a continuous function $f(s)$, we wish to determine a continuous function $\varphi(s)$ such that

$$f(s) = \varphi(s) - \lambda \int_a^b K(s,t)\varphi(t) dt. \quad (9)$$

Set

$$\varphi(s) = f(s) + g(s).$$

Then

$$g(s) = \lambda \int_a^b K(s,t) (f(t) + g(t)) dt. \quad (10)$$

Hence from the expansion theorem of the last section,

$$g(s) = \sum_{\nu} \varphi_{\nu}(s) \int_a^b g(t)\varphi_{\nu}(t) dt, \quad (11)$$

where the $\varphi_{\nu}(s)$ range over a full normalized orthogonal system for the kernel $K(s,t)$ and the series is absolutely and uniformly convergent. Multiplying (10) by $\varphi_{\nu}(s) ds$ and

integrating, we get

$$\begin{aligned} \int_a^b g(s)\varphi_\nu(s) ds &= \lambda \int_a^b (f(t) + g(t)) dt \int_a^b K(s,t)\varphi_\nu(s) ds, \\ \int_a^b g(t)\varphi_\nu(t) dt &= \frac{\lambda}{\lambda_\nu} \int_a^b f(t)\varphi_\nu(t) dt + \frac{\lambda}{\lambda_\nu} \int_a^b g(t)\varphi_\nu(t) dt, \\ \int_a^b g(t)\varphi_\nu(t) dt &= \frac{\lambda}{\lambda_\nu - \lambda} \int_a^b f(t)\varphi_\nu(t) dt. \end{aligned} \quad (12)$$

Hence by (11)

$$\varphi(s) = f(s) + \lambda \sum_\nu \frac{\varphi_\nu(s)}{\lambda_\nu - \lambda} \int_a^b f(t)\varphi_\nu(t) dt. \quad (13)$$

Conversely, when λ is distinct from all λ_ν , the above series converges absolutely and uniformly by the results of § 2 since

$$\frac{\varphi_\nu(s)}{\lambda_\nu - \lambda} \int_a^b f(t)\varphi_\nu(t) dt = \frac{1}{1 - \frac{\lambda}{\lambda_\nu}} \int_a^b K(s,t)\varphi_\nu(t) dt \int_a^b f(t)\varphi_\nu(t) dt.$$

Moreover, if we substitute equation (13) into equation (9) and take into account the equation

$$\int_a^b K(s,t)f(t) dt = \sum_\nu \frac{\varphi_\nu(s)}{\lambda_\nu} \int_a^b f(t)\varphi_\nu(t) dt,$$

which follows from the expansion theorem of the previous section, we find that (13) represents a solution of equation (9). *We therefore see that if λ is not an eigenvalue of the kernel $K(s,t)$ then equation (9) has one and only solution given by equation (13). However, if λ is a k -fold eigenvalue, it follows from (12) that for equation (9) to have a solution the k equations*

$$\int_a^b f(t)\varphi_{n+\nu}(t) dt = 0$$

must be satisfied, where $n+1, n+2, \dots, n+k$ are the indices of the eigenfunction of the full normalized orthogonal system that correspond to the k -fold eigenvalue. In this case, as substitution in equation (9) shows,

$$\begin{aligned} \varphi(s) &= f(s) + a_1\varphi_{n+1}(s) + a_2\varphi_{n+2}(s) + \dots + a_k\varphi_{n+k}(s) \\ &\quad + \lambda \sum_\nu \frac{\varphi_\nu(s)}{\lambda_\nu - \lambda} \int_a^b f(t)\varphi_\nu(t) dt, \end{aligned}$$

where ν ranges through all indices of the orthogonal system with the exception of $n+1, n+2, \dots, n+k$ and a_1, a_2, \dots, a_k , are arbitrary constants. These results for symmetric kernels are essentially the same as the [general] theorems that Fredholm proved by means of his series.

§ 11.

Proof of the Fundamental Theorem

We now turn to the proof of the fundamental theorem — every kernel has at least one eigenfunction — which we postponed in § 7.

Set

$$U_1 = \int_a^b K^1(s, s) ds, \quad U_2 = \int_a^b K^2(s, s) ds, \dots, \quad U_n = \int_a^b K^n(s, s) ds, \dots$$

Then it follows from (1) in § 6 that

$$U_{\mu+\nu} = \int_a^b \int_a^b K^\mu(s, r) K^\nu(s, r) dr ds, \quad (14)$$

$$U_{2\nu} = \int_a^b (K^\nu(s, r))^2 dr ds. \quad (15)$$

Now since $K^\nu(s, t)$ cannot vanish identically, as we showed in § 6, it follows that all $U_{2\nu}$ are different from zero and positive. Let $n \geq 2$. If we substitute $n+1$ for ν and $n-1$ for μ in (14) and apply Schwarz's inequality from § 1 (which according to the concluding remark in the first chapter is also valid for multiple integrals), we get

$$U_{2n}^2 \leq U_{2n-2} U_{2n+2},$$

or

$$\frac{U_{2n}}{U_{2n-2}} \leq \frac{U_{2n+2}}{U_{2n}}.$$

If we now set

$$\frac{U_{2n+2}}{U_{2n}} = c_n, \quad (16)$$

then

$$c_{n-1} \leq c_n. \quad (17)$$

Now by equation (1) of § 6,

$$\begin{aligned} K^{\mu+\nu}(s, t) &= \int_a^b K^\mu(s, r) K^\nu(r, t) dr, \\ (K^{\mu+\nu}(s, t))^2 &\leq \int_a^b (K^\mu(s, r))^2 dr \int_a^b (K^\nu(t, r))^2 dr, \\ \int_a^b \int_a^b (K^{\mu+\nu}(s, t))^2 ds dt &\leq \int_a^b \int_a^b (K^\mu(s, r))^2 dr ds \cdot \int_a^b \int_a^b (K^\nu(t, r))^2 dr dt. \end{aligned}$$

Therefore by (15)

$$\begin{aligned} U_{2\mu+2\nu} &\leq U_{2\mu} \cdot U_{2\nu}, \\ U_{2\nu} &\geq \frac{U_{2\mu+2\nu}}{U_{2\mu}}, \end{aligned}$$

and by (16) and (17)

$$U_{2\nu} \geq c_\mu^\nu. \quad (18)$$

In view of (17),

$$\lim_{\mu=\infty} c_\mu = c,$$

where c is a finite, positive number. Moreover

$$\frac{U_{2\nu}}{c^\nu} \geq 1. \quad (19)$$

Because $c_n \leq c$, it follows from (16) that

$$\frac{U_{2n+2}}{c^{n+1}} \leq \frac{U_{2n}}{c^n}. \quad (20)$$

From (19) and (20) it follows that

$$\lim_{n=\infty} \frac{U_{2n}}{c^n} = U, \quad (21)$$

where $U \geq 1$ is finite.

Now

$$\begin{aligned} &\frac{K^{2n+2m}(s,t)}{c^{n+m}} - \frac{K^{2n}(s,t)}{c^n} \\ &= \frac{1}{c} \int_a^b \int_a^b K(s, r_1) \left\{ \frac{K^{2n+2m-2}(r_1, r_2)}{c^{n+m-1}} - \frac{K^{2n-2}(r_1, r_2)}{c^{n-1}} \right\} K(r_2, t) dr_1 dr_2. \end{aligned}$$

Hence

$$\begin{aligned} &\left(\frac{K^{2n+2m}(s,t)}{c^{n+m}} - \frac{K^{2n}(s,t)}{c^n} \right)^2 \leq \frac{1}{c^2} \int_a^b \int_a^b (K(s, r_1) K(r_2, t))^2 dr_1 dr_2 \times \\ &\times \int_a^b \int_a^b dr_1 dr_2 \left\{ \left(\frac{K^{2n+2m-2}(r_1, r_2)}{c^{n+m-1}} \right)^2 - 2 \frac{K^{2n+2m-2}(r_1, r_2) K^{2n-2}(r_1, r_2)}{c^{2n+m-2}} + \left(\frac{K^{2n-2}(r_1, r_2)}{c^{n-1}} \right)^2 \right\}. \end{aligned}$$

From this, in view of (14) and (15), it follows that

$$\begin{aligned} &\left(\frac{K^{2n+2m}(s,t)}{c^{n+m}} - \frac{K^{2n}(s,t)}{c^n} \right)^2 \leq \frac{1}{c^2} \int_a^b (K(s, r_1))^2 dr_1 \int_a^b (K(t, r_2))^2 dr_2 \times \\ &\times \left\{ \frac{U_{4n+4m-4}}{c^{2n+2m-2}} - 2 \frac{U_{4n+2m-4}}{c^{2n+m-2}} + \frac{U_{4n-4}}{c^{2n-2}} \right\} \end{aligned}$$

Now as n increases, the expression on the right side becomes infinitely small, independently of m , s , and t . From this it follows that with increasing n , $\frac{K^{2n}(s,t)}{c^n}$ converges uniformly to a necessarily continuous function $u(s,t)$, which, since

$$\int_a^b u(s,s) ds = \lim_{n=\infty} \int_a^b \frac{K^{2n}(s,s)}{c^n} ds = \lim_{n=\infty} \frac{U_{2n}}{c_n} = U \geq 1,$$

cannot vanish identically in s and t . Moreover, from

$$\frac{K^{2n+2}(s,t)}{c^{n+1}} = \frac{1}{c} \int_a^b K^2(s,r) \frac{K^{2n}(r,t)}{c^n} dr,$$

it follows that

$$u(s,t) = \frac{1}{c} \int_a^b K^2(s,r) u(r,t) dr.$$

If we now choose a value t_1 for which $u(s,t_1)$ does not vanish identically in s , then by the last equation $u(s,t_1)$ is an eigenfunction of the kernel $K^2(s,t)$. From this it follows according to § 6 that $K(s,t)$ must also have an eigenfunction, which is what was to be proved.

§ 12.

Generalization of the Hypotheses

We can allow the symmetric kernel to be discontinuous in cases that satisfy the following conditions.

I. The point set in the s, t -plane consisting of the points of discontinuity of $K(s,t)$ (and which is therefore closed) has outer content zero on any line $s = \text{const}$.

II. $\int_a^b (K(s,t))^2 dt$ is defined and finite for $a \leq s \leq b$ and represents a continuous function of s .

Let the square domain of definition of $K(s,t)$ be divided into 2^{2n} equal squares by lines parallel to the sides and let Q_n denote the region formed from the union of the squares that have points of discontinuity of $K(s,t)$ in their interiors or on their boundaries. Then the following result can be proved with no difficulty from I and II: for any arbitrarily small positive quantity ϵ there is a number n such that for all lines $s = \text{const}$. the total size of the region covered by Q_n and the value of the integral

$$\int_{Q_n} (K(s,t))^2 dt$$

taken over the totality of this region are both less than ϵ .*)

From this it first of all follows that for all lines $s = \text{const.}$

$$\int_{Q_n} |K(s, t)| dt < \epsilon.$$

For by the Schwarz inequality

$$\left(\int_{Q_n} |K(s, t)| dt \right)^2 \leq \int_{Q_n} (K(s, t))^2 dt \cdot \int_{Q_n} dt \leq \epsilon^2.$$

From the inequality just established it easily follows that the planar content of the set of points of discontinuity of $K(s, t)$ is zero. For any continuous function $m(t)$, $\int_a^b (K(s, t)) m(t) dt$ and $\int_a^b (K(s, t))^2 m(t) dt$ is well determined and finite for $a \leq s \leq b$ and represents a continuous function of s . Similarly one can easily prove by an application of the Schwarz inequality to the integral over the product of two kernels that

$$K^2(s, t) = \int_a^b K(s, r) K(r, t) dr$$

is well determined and finite for $a \leq s \leq b$, $a \leq t \leq b$ and represents a continuous function of s and t , which because

$$K^2(s, s) = \int_a^b (K(s, r))^2 dr$$

can vanish identically only if $K(s, t)$ vanishes identically in its region of continuity.

These consequences of the assumptions I and II easily ensure that all the operations occurring in §§ 4, 5, 6, especially the frequent exchange of the order of integration, are legitimate.*) Thus the theorems and proofs of §§ 4, 5, 6 remain valid as they are. The fundamental theorem stated in § 7 still stands, since K^2 , which is a continuous kernel, must have eigenfunctions, and from this the existence of eigenfunctions of $K(s, t)$ follows as in § 6.

In § 8 the validity of equation (3) is confined to the region of continuity of the kernel, while equation (6) is remains valid as is. In §§ 9 and 10 all theorems and proofs are unchanged.

It is also permissible for the conditions I and II to be violated along a finite number of lines $s = \text{const.}$, where the kernel takes on different values at the two boundaries of each line. However, I and II must be assumed to hold in each of the rectangles,

^{9*)}Instead of the assumptions I and II, one can require that I and this inequality be satisfied. From these assumptions II easily follows.

⁹See, e.g., Jordan, Cours d'Analyse, V. II, Ch. II, II [sic].

including their boundaries, into which the square of definition of the kernel is divided by the lines. Moreover, both the eigenfunctions and the solutions of the inhomogeneous integral equation must be allowed to jump at the values of s in question.

Further, the range of validity of the development theorem in § 9 can be extended by replacing the assumption of the continuity of $p(t)$ by the integrability of $p(t)$ and its square.

Likewise, nothing changes in the theorems and proofs of this chapter if s, t, r, \dots are points from a bounded n -dimensional domain consisting of a finite number of analytic pieces in an $(n + m)$ -dimensional space. In this case ds, dt, dr, \dots are the corresponding [differential] elements.

In this case too the conditions I and II determine a permissible range of discontinuity.

Chapter III

On the linear unsymmetric integral Equation.

§ 13.

The inhomogeneous integral equation.

Let the kernel $K(s, t)$, which is no longer assumed to be symmetric, and the function $f(s)$ be real continuous functions defined for $a \leq s \leq b$, $a \leq t \leq b$. We seek a real continuous function $\varphi(s)$ that satisfies the integral equation

$$f(s) = \varphi(s) - \int_a^b K(s, t)\varphi(t) dt. \quad (22)$$

If we set

$$g(t) = \chi(t) - \int_a^b K(s, t)\chi(s) ds, \quad (23)$$

we get the identities

$$g(s) - \int_a^b K(s, t)g(t) dt = \chi(s) - \int_a^b Q(s, t)\chi(t) dt, \quad (24)$$

$$\int_a^b (g(s))^2 ds = \int_a^b \chi(s)(\chi(s) - \int_a^b Q(s, t)\chi(t) dt) ds, \quad (25)$$

where

$$Q(s, t) = K(s, t) + K(t, s) - \int_a^b K(s, r)K(t, r) dr$$

is necessarily symmetric.

Any real, continuous, nonzero function that causes the right-hand side of equation (22) to vanish will be called a *null solution in s of the kernel*. Any real, continuous, nonzero function that causes the right hand side of equation (23) to vanish will be called a *null solution in t* . According to the first theorem of § 5 for $\lambda = 1$, whose proof does not use the assumption of the symmetry of the kernel, the number of linearly independent null solutions in s , as well as in t , is *finite*. If $\chi(t)$ is a null solution in t , it follows from (23) and (24) that $\chi(t)$ is an eigenfunction of the symmetric kernel $Q(s, t)$ corresponding to the eigenvalue $\lambda = 1$. The converse follows from (25) and (23). Hence we obtain the first set of functions as we form the second set, since they have been shown to be *identical*.

Now a necessary and sufficient condition for the solution of equation (22) is the orthogonality of $f(s)$ to all possible null solutions in t . Moreover, all solutions may be obtained from a single solution by the additive combination with all null solutions in s .)*

For the necessity of this condition is easily seen by multiplying equation (22) by a null solution in t and integrating. The straightforward application of the existence theorem of § 10 to the symmetric integral equation

$$f(s) = \chi(s) - \int_a^b Q(s, t)\chi(t) dt$$

shows that the condition is sufficient. For by (24) the equation (22) can be reduced to this equation by the substitution

$$\varphi(t) = \chi(t) - \int_a^b K(s, t)\chi(s) ds.$$

§ 14.

The Concept of Eigenfunction

Let $K(s, t)$, defined for $a \leq s \leq b$, $a \leq t \leq b$, be a real continuous function that will not be assumed symmetric. If $\varphi(s)$ and $\psi(s)$ are real or complex functions that are not identically zero and that together satisfy the equations

$$\varphi(s) = \lambda \int_a^b K(s, t)\psi(t) dt \tag{26}$$

and

$$\psi(s) = \lambda \int_a^b K(t, s)\varphi(t) dt, \tag{27}$$

^{9*)}This theorem was first proved by by Fredholm, *loc. cit.*

then they will be said to be a *pair of adjoint eigenfunctions of the kernel* $K(s, t)$ *corresponding to the eigenvalue* λ .

We now define

$$\overline{K}(s, t) = \int_a^b K(s, r)K(t, r) dr \quad (28)$$

and

$$\underline{K}(s, t) = \int_a^b K(r, s)K(r, t) dr. \quad (29)$$

Then $\overline{K}(s, t)$ and $\underline{K}(s, t)$ are symmetric.

If we substitute (27) into (26) and (26) into (27), we get the equations

$$\varphi(s) = \lambda^2 \int_a^b \overline{K}(s, t)\varphi(t) dt \quad (30)$$

and

$$\psi(s) = \lambda^2 \int_a^b \underline{K}(s, t)\psi(t) dt. \quad (31)$$

Now if we were to have

$$\varphi(s) = \varphi_1(s) + i\varphi_2(s) \quad \text{and} \quad \psi(s) = \psi_1(s) + i\psi_2(s),$$

then, since λ^2 is the eigenvalue of the symmetric kernel $\overline{K}(s, t)$ and hence, as was shown in § 4, is real, we would have from (30) the following equations

$$\begin{aligned} \varphi_1(s) &= \lambda^2 \int_a^b \overline{K}(s, t)\varphi_1(t) dt \\ \int_a^b (\varphi_1(s))^2 ds &= \lambda^2 \int_a^b \int_a^b \varphi_1(s)\overline{K}(s, t)\varphi_1(t) ds dt \\ &= \lambda^2 \int_a^b dr \int_a^b K(s, r)\varphi_1(s) ds \int_a^b K(t, r)\varphi_1(t) dt \\ \int_a^b (\varphi_1(s))^2 ds &= \lambda^2 \int_a^b dr \left(\int_a^b K(s, r)\varphi_1(s) ds \right)^2. \end{aligned}$$

In the same way it follows that

$$\int_a^b (\varphi_2(s))^2 ds = \lambda^2 \int_a^b dr \left(\int_a^b K(s, r)\varphi_2(s) ds \right)^2.$$

Now since for at least one of these two equations both sides cannot be identically zero, it follows that λ^2 is positive and hence that λ is real. Therefore, at least one of the pairs $\varphi_1(s)$ and $\psi_1(s)$ and $\varphi_2(s)$ and $\psi_2(s)$ must be adjoint eigenfunctions of the kernel

$K(s, t)$. For this reason we will treat only *real pairs* of adjoint eigenfunctions in what follows and the term eigenfunction will refer only to real pairs. By a suitable choice of the sign of $\psi(s)$ we can assume that the eigenvalues of an unsymmetric kernel are all positive.

Equation (26) follows from (30) if $\psi(s)$ is defined by (27), and (31) follows from the substitution of (26) in (27). Likewise, (27) follows from (31) if $\varphi(s)$ is defined by (26), and (30) follows by the substitution of (27) in (26). *Therefore, to each eigenfunction of the symmetric kernel $\overline{K}(s, t)$ there corresponds an eigenfunction of the symmetric Kernel $\underline{K}(s, t)$ and conversely — and in such a way that this pair of functions forms a pair of eigenfunctions of the unsymmetric kernel $K(s, t)$.*

§ 15

The full normalized orthogonal system of an unsymmetric kernel.

The adjoint functions of a full normalized orthogonal system of the kernel $\overline{K}(s, t)$ form a full normalized orthogonal system of the kernel $\underline{K}(s, t)$, and conversely.

PROOF.

$$\begin{aligned} \int_a^b \psi_\mu(r) \psi_\nu(r) dr &= \int_a^b dr \lambda_\mu \int_a^b K(t, r) \varphi_\mu(t) dt \lambda_\nu \int_a^b K(s, r) \varphi_\nu(s) ds \\ &= \lambda_\mu \lambda_\nu \int_a^b \overline{K}(s, t) \varphi_\mu(t) \varphi_\nu(s) ds dt. \end{aligned}$$

It follows from (30) that

$$\int_a^b \psi_\mu(s) \psi_\nu(s) ds = \frac{\lambda_\nu}{\lambda_\mu} \int_a^b \varphi_\mu(s) \varphi_\nu(s) ds.$$

It therefore follows from this equation that if the functions $\varphi_1, \varphi_2, \dots, \varphi_n, \dots$ form a full normalized orthogonal system of the kernel $\overline{K}(s, t)$, then the adjoint functions $\psi_1, \psi_2, \dots, \psi_n, \dots$ are all normalized and pairwise orthogonal. Now let $\psi(s)$ be an eigenfunction of $\underline{K}(s, t)$ and let $\varphi(s)$ be its adjoint, which is therefore an eigenfunction of $\overline{K}(s, t)$. Then according to our assumption [of the fullness of the system of the φ_ν] we have

$$\varphi(s) = \sum_{\rho} c_{\rho} \varphi_{\rho}(s)$$

where ρ runs through a finite number of indices. By § 5 all the $\varphi_{\rho}(s)$ correspond to the same eigenvalue as $\varphi(s)$. Then from the equations

$$\begin{aligned} \psi_{\rho}(s) &= \lambda \int_a^b K(s, t) \varphi_{\rho}(t) dt, \\ \psi(s) &= \lambda \int_a^b K(t, s) \varphi(t) dt, \end{aligned}$$

it follows that

$$\psi(s) = \sum_{\rho} c_{\rho} \psi_{\rho}(s).$$

Hence the functions $\psi_1, \psi_2, \dots, \psi_n, \dots$ form a full normalized orthogonal system for the kernel $\underline{K}(s, t)$, which is what was to be proved. The converse is established similarly.

By a *full normalized orthogonal system of the unsymmetric kernel $K(s, t)$* we will understand the above pair of adjoint normalized orthogonal systems of the kernels $\overline{K}(s, t)$ and $\underline{K}(s, t)$.

§ 16

Expansion of Arbitrary Functions

Let the functions

$$\begin{aligned} &\varphi_1(s), \varphi_2(s), \dots, \varphi_n(s), \dots \\ &\psi_1(s), \psi_2(s), \dots, \psi_n(s), \dots \end{aligned}$$

corresponding to the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n, \dots$ (which are ordered by magnitude) form a full normalized orthogonal system of the unsymmetric kernel $K(s, t)$ as defined in the previous section. Then we have the following theorem:

If $h(s)$ is a continuous function and

$$\int_a^b K(t, s)h(t) dt = 0$$

identically in s , then for all ν

$$\int_a^b h(s)\varphi_{\nu}(s) ds = 0.$$

This can be shown by multiplying equation (26) by $h(s) ds$ and integrating from a to b . Similarly, if

$$\int_a^b K(s, t)h(t) dt = 0$$

identically in s , then for all ν

$$\int_a^b h(s)\psi_{\nu}(s) ds = 0.$$

Conversely, if for all ν

$$\int_a^b h(s)\varphi_{\nu}(s) ds = 0.$$

then the equation

$$\int_a^b K(t, s)h(t) dt = 0$$

holds; and if for all ν

$$\int_a^b h(s)\psi_\nu(s) ds = 0,$$

then the equation

$$\int_a^b K(s, t)h(t) dt = 0$$

holds.

PROOF. We will prove only the first assertion, since the proof of the second is the same. Since by hypothesis $h(s)$ is orthogonal to all functions of a full normalized orthogonal system for the symmetric kernel $\overline{K}(s, t)$, it follows from § 9 that

$$\int_a^b \overline{K}(s, t)h(t) dt = 0$$

and

$$\begin{aligned} 0 &= \int_a^b \int_a^b \overline{K}(s, t)h(s)h(t) ds dt \\ &= \int_a^b dr \int_a^b K(s, r)h(s) ds \int_a^b K(t, r)h(t) dt \\ &= \int_a^b dr \left(\int_a^b K(s, r)h(s) ds \right)^2. \end{aligned}$$

It follows that

$$\int_a^b K(s, r)h(s) ds = 0$$

identically in r — which is what was to be proven.

If

$$g(s) = \int_a^b K(s, t)h(t) dt,$$

where $h(t)$ is a continuous function, then

$$\begin{aligned} g(s) &= \sum_\nu \varphi_\nu(s) \int_a^b g(t)\varphi_\nu(t) dt \\ &= \sum_\nu \frac{\varphi_\nu(s)}{\lambda_\nu} \int_a^b h(t)\psi_\nu(t) dt \\ &= \sum_\nu \int_a^b K(s, t)\phi_\nu(t) dt \int_a^b h(t)\psi_\nu(t) dt. \end{aligned}$$

If

$$g(s) = \int_a^b K(t, s)h(t) dt,$$

then

$$\begin{aligned} g(s) &= \sum_{\nu} \psi_{\nu}(s) \int_a^b g(t)\psi_{\nu}(t) dt \\ &= \sum_{\nu} \frac{\psi_{\nu}(s)}{\lambda_{\nu}} \int_a^b h(t)\varphi_{\nu}(t) dt \\ &= \sum_{\nu} \int_a^b K(t, s)\varphi_{\nu}(t) dt \int_a^b h(t)\varphi_{\nu}(t) dt. \end{aligned}$$

The series in the right-hand sides of both equations converge absolutely and uniformly.

PROOF. We will only prove the first assertion, since the proof of the second is the same. From the third representation of the general term in the series, the convergence theorem in § 2 allows us to conclude that the series converges absolutely and uniformly. If we now set

$$g(s) - \sum_{\nu} \varphi_{\nu}(s) \int_a^b g(t)\varphi_{\nu}(t) dt = f(s),$$

then it follows that

$$\int_a^b f(s)\varphi_{\nu}(s) ds = 0. \quad (32)$$

From this and the theorem we have just proved it follows that

$$\int_a^b K(t, s)f(t) dt = 0. \quad (33)$$

Now by (33)

$$\int_a^b (f(s))^2 ds = \int_a^b f(s)g(s) ds = \int_a^b h(t) dt \int_a^b K(s, t)f(s) ds = 0.$$

Hence $f(s) = 0$, which was what was to be proven.

Let $p(s)$ and $q(s)$ be two continuous functions. Then the theorem we just proved gives

$$\int_a^b K(s, t)q(t) dt = \sum \frac{\varphi_{\nu}(s)}{\lambda_{\nu}} \int_a^b q(t)\psi_{\nu}(t) dt.$$

Multiplying this equations by $p(s) ds$ and integrating from a to b , we get

$$\int_a^b \int_a^b K(s, t)p(s)q(s) ds dt = \sum_{\nu} \frac{1}{\lambda_{\nu}} \int_a^b p(s)\varphi_{\nu}(s) ds \int_a^b q(t)\psi_{\nu}(t) dt.$$

This theorem corresponds to the canonical decomposition of a bilinear form.

From the theorem just proven it follows that if $\sum_{\nu} \frac{\varphi_{\nu}(s)\psi_{\nu}(t)}{\lambda_{\nu}}$ converges uniformly, then

$$K(s, t) = \sum_{\nu} \frac{\varphi_{\nu}(s)\psi_{\nu}(t)}{\lambda_{\nu}}. \quad (34)$$

In particular, *this equation always holds if the full normalized orthogonal system of the kernel $K(s, t)$ consists of only a finite number of pairs of functions.*

§ 17.

Generalization of the Hypotheses

As an argument fully analogous to the one in § 12 shows, we can permit the discontinuity of the unsymmetric kernel in cases that satisfy the following conditions.

I. The point set in the s, t -plane consisting of the points of discontinuity of $K(s, t)$ must have zero outer content on any line $s = \text{const.}$, $t = \text{const.}$.

II. $\int_a^b (K(s, t))^2 dt$ and $\int_a^b (K(t, s))^2 dt$ must be finite and well defined for $a \leq s \leq b$ and represent continuous functions of s that are not identically zero.

Then all the theorems and proofs of this chapter continue to hold. Only the validity of equation (34) depends on the continuity of $K(s, t)$.

Likewise, nothing is changed in the theorems and proofs in this chapter, when s , t , and r are points from a bounded n -dimensional domain consisting of a finite number of analytic pieces in an $(n + m)$ -dimensional space. Here ds , dt , and dr are the corresponding [differential] elements. In this case too, the conditions I and II determine a realm in which discontinuity is permitted.

Chapter IV

On the Best Approximation of Functions of Two Variables by Sums of Products of Functions of One Variable

§ 18.

The approximation theorem.

Let $K(s, t)$ be a real continuous function defined on $a \leq s \leq b$, $a \leq t \leq b$. We wish to approximate $K(s, t)$ as well as possible by a sum of at most m products of a continuous function of s and a continuous function of t . Here, as usual, we will take as the measure [of the quality] of the approximation the double integral of the square error ranging over the domain of definition of the given function.

Let the functions

$$\begin{aligned} &\varphi_1(s), \varphi_2(s), \dots, \varphi_\nu(s), \dots, \\ &\psi_1(s), \psi_2(s), \dots, \psi_\nu(s), \dots, \end{aligned}$$

corresponding to the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_\nu \dots$ arranged in ascending order form a *full normalized orthogonal system of the unsymmetric kernel* $K(s, t)$ as defined in § 15. In the case where the number of adjoint eigenfunctions traversed by the index ν is less than or equal to m , then equation (34) gives an immediate and trivial solution of our problem. If, however, if this number is infinite or finite and greater than or equal to m , then the solution is given by the sum of products

$$\sum_{\nu=1}^{\nu=m} \frac{\varphi_\nu(s)\psi_\nu(t)}{\lambda_\nu}.$$

PROOF. The measure of the approximation M_m , whose minimum is required by the statement of the problem, by hypothesis is defined as

$$M_m = \int_a^b \int_a^b \left(K(s, t) - \sum_{\nu=1}^{\nu=m} \frac{\varphi_\nu(s)\psi_\nu(t)}{\lambda_\nu} \right)^2 ds dt.$$

From the definitions (26) and (27) and in view of the of the orthogonality and normalization of the system of functions $\varphi_\nu(s)$ and the system of functions $\psi_\nu(s)$, the above expression is easily reduced to the formula

$$M_m = \int_a^b \int_a^b (K(s, t))^2 ds dt - \sum_{\nu=1}^{\nu=m} \frac{1}{\lambda_\nu^2}. \quad (35)$$

We must therefore show that

$$\int_a^b \int_a^b (K(s, t) - \sum_{\nu=1}^{\nu=m} \alpha_\nu \beta_\nu)^2 ds dt \geq \int_a^b \int_a^b (K(s, t))^2 ds dt - \sum_{\nu=1}^{\nu=m} \frac{1}{\lambda_\nu^2}. \quad (36)$$

holds for all systems of n continuous function pairs

$$\begin{aligned} &\alpha_1(s), \alpha_2(s), \dots, \alpha_n(s), \\ &\beta_1(s), \beta_2(s), \dots, \beta_n(s), \end{aligned}$$

where $n \leq m$. Here we have written α_ν for $\alpha_\nu(s)$ and β_ν for $\beta_\nu(s)$. We may assume that the functions $\beta_1, \beta_2, \dots, \beta_n$ are normalized and pairwise orthogonal. For if that were not so, we could, as in § 3, represent them by a homogeneous combination of at most n

such functions with linear coefficients and then order the sum of products according to the latter. It then follows that

$$\begin{aligned}
\int_a^b \int_a^b (K(s,t) - \sum_{\nu=1}^{\nu=m} \alpha_\nu \beta_\nu)^2 ds dt &= \int_a^b \int_a^b (K(s,t))^2 ds dt \\
&\quad + \sum_{\nu=1}^{\nu=n} \int_a^b (\alpha_\nu^2 - 2\alpha_\nu \int_a^b K(s,t) \beta_\nu dt) ds \\
&= \int_a^b \int_a^b (K(s,t))^2 ds dt + \sum_{\nu=1}^{\nu=n} \int_a^b (\alpha_\nu - 2\alpha_\nu \int_a^b K(s,t) \beta_\nu dt)^2 ds \\
&\quad - \sum_{\nu=1}^{\nu=n} \int_a^b (\int_a^b K(s,t) \beta_\nu dt)^2 ds.
\end{aligned} \tag{37}$$

The inequality (36), which we are to prove, therefore follows immediately from the inequality

$$0 \leq \sum_{\nu=1}^{\nu=m} \frac{1}{\lambda_\nu^2} - \sum_{\nu=1}^{\nu=n} \int_a^b (\int_a^b K(s,t) \beta_\nu dt)^2 ds.$$

Moreover, since $m \leq n$, this inequality follows immediately from

$$0 \leq \sum_{\nu=1}^{\nu=n} \frac{1}{\lambda_\nu^2} - \sum_{\nu=1}^{\nu=n} \int_a^b (\int_a^b K(s,t) \beta_\nu dt)^2 ds, \tag{38}$$

which we will now establish.

According to the expansion theorem given in § 16, we have

$$\int_a^b K(s,t) \beta_\nu dt = \sum_{\rho} \frac{\varphi_\rho(s)}{\lambda_\rho} \int_a^b \beta_\nu \psi_\rho(t) dt, \tag{39}$$

where the sum ranges over pairs of adjunct functions $\varphi_\rho(s)$, $\psi_\rho(t)$ of the full normalized orthogonal system. In view of the orthonormality of the system of functions $\varphi(s)$, it follows easily from equation (39) that

$$\int_a^b (\int_a^b K(s,t) \beta_\nu dt)^2 ds = \sum_{\rho} \frac{1}{\lambda_\rho^2} \int_a^b \beta_\nu \psi_\rho(t) dt. \tag{40}$$

Now according to Bessel's inequality in § 1

$$1 = \int_a^b \beta_\nu^2 dt \geq \sum_{\rho} (\int_a^b \beta_\nu \psi_\rho(t) dt)^2, \tag{41}$$

and hence the above sum converges. It then follows from a simple transformation of the right-hand side of equation (40) that

$$\begin{aligned} \int_a^b \left(\int_a^b K(s, t) \beta_\nu dt \right)^2 ds &= \frac{1}{\lambda_n^2} + \sum_{\mu=1}^{\mu=n} \left(\frac{1}{\lambda_\mu^2} - \frac{1}{\lambda_n^2} \right) \left(\int_a^b \beta_\nu \psi_\mu(t) dt \right)^2 \\ &- \sum_k \left(\frac{1}{\lambda_n^2} - \frac{1}{\lambda_k^2} \right) \left(\int_a^b \beta_\nu \psi_k(t) dt \right)^2 - \frac{1}{\lambda_n^2} \left[1 - \sum_\rho \left(\int_a^b \beta_\nu \psi_\rho(t) dt \right)^2 \right], \end{aligned} \quad (42)$$

where k runs through all indices of the full orthogonal system that are greater than n . By hypothesis the inequality

$$\lambda_k \geq \lambda_n$$

is satisfied for all values of k . This and the inequality (41) show that

$$\sum_k \left(\frac{1}{\lambda_n^2} - \frac{1}{\lambda_k^2} \right) \left(\int_a^b \beta_\nu \psi_k(t) dt \right)^2 \geq 0$$

and

$$\frac{1}{\lambda_n^2} \left[1 - \sum_\rho \left(\int_a^b \beta_\nu \psi_\rho(t) dt \right)^2 \right] \geq 0.$$

Hence the inequality (38) to be established follows directly from the inequality

$$\begin{aligned} 0 &\leq \sum_{\nu=1}^{\nu=n} \frac{1}{\lambda_\nu^2} - \frac{n}{\lambda_n^2} + \sum_{\nu=1}^{\nu=n} \sum_{\mu=1}^{\mu=n} \left(\frac{1}{\lambda_\mu^2} - \frac{1}{\lambda_n^2} \right) \left(\int_a^b \beta_\nu \psi_\mu(t) dt \right)^2 \\ &= \sum_{\mu=1}^{\mu=n} \left(\frac{1}{\lambda_\mu^2} - \frac{1}{\lambda_n^2} \right) \left[1 - \sum_{\nu=1}^{\nu=n} \left(\int_a^b \beta_\nu \psi_\rho(t) dt \right)^2 \right], \end{aligned}$$

But the fact that the last expression is nonnegative follows from the inequality

$$\lambda_\mu \leq \lambda_n,$$

which is true by hypothesis, and from Bessel's inequality

$$1 = \int_a^b (\psi_\mu(t))^2 dt \geq \sum_{\nu=1}^{\nu=n} \left(\int_a^b \psi_\mu(t) \beta_\nu(t) dt \right)^2,$$

which true because the system of functions β_ν is orthonormal.

§ 19.

The measure of the best approximation.

The measure M_m , defined in the previous section, of the best approximation to a function $K(s, t)$ by a sum of at most m products of a function of s and a function of t vanishes as m grows unboundedly.

PROOF. According to equation (35) the assertion to be proved can be written in the form

$$\sum_{\rho} \frac{1}{\lambda_{\rho}^2} = \int_a^b \int_a^b (K(s, t))^2 ds dt, \quad (43)$$

where λ_{ρ} runs through all the eigenvalues of the unsymmetric kernel $K(s, t)$, each counted according to its multiplicity.

By the expansion theorem of § 16, we have

$$\int_a^b K(s, t)K(r, t) dt = \sum_{\rho} \frac{\varphi_{\rho}(s)}{\lambda_{\rho}} \int_a^b K(r, t)\psi_{\rho}(t) dt = \sum_{\rho} \frac{\varphi(s)\varphi(r)}{\lambda_{\rho}^2}, \quad (44)$$

in which the sum converges uniformly in r for fixed s and uniformly in s for fixed r . If we set $r = s$, we get

$$\int_a^b (K(s, t))^2 dt = \sum_{\rho} \frac{(\varphi_{\rho}(s))^2}{\lambda_{\rho}^2}. \quad (45)$$

Equation (43), which is to be established, follows from (45) by integrating in s from a to b , provided we can integrate the the right-hand side *termwise* — in particular, provided the series on the right-hand side of (45) converges *uniformly*. A theorem of Dini* is sufficient to establish the uniform convergence of the series (45) that is required to conclude the proof, and moreover to prove the uniform convergence of (44) in s and t because

$$\frac{\varphi(s)\varphi(r)}{\lambda_{\rho}^2} \leq \frac{1}{2} \left(\frac{(\varphi_{\rho}(s))^2}{\lambda_{\rho}^2} + \frac{(\varphi_{\rho}(r))^2}{\lambda_{\rho}^2} \right).$$

Dini's theorem states that if a series of positive, continuous functions of the variable s defined for $a \leq s \leq b$ converges so that the sum represents a continuous function then the convergence is also uniform.

PROOF. Let

$$v(s) = \sum_{\nu=1}^{\nu=\infty} u_{\nu}(s), \quad (46)$$

where $v(s)$ and the $u_{\nu}(s)$ are continuous and nonnegative for $a \leq s \leq b$.

*Dini, "Fondamenti per la teoria funzioni di variabili reali," Pisa 1878, §99.

Denote by P_n the point set consisting of all points for which the continuous function

$$R_n(s) = v(s) - \sum_{\nu=1}^{\nu=n} u_\nu(s)$$

attains its maximum, which we will denote by $\text{Max}(R_n)$. Then let a single point α_n be chosen from each of the point sets P_n , and let α be an accumulation point of the point set consisting of the points $\alpha_1, \alpha_2, \dots, \alpha_n, \dots$. Now let ϵ be an arbitrarily small, positive, nonzero quantity. By the assumed convergence of the series (46), there is an index p such that

$$R_p(\alpha) < \frac{\epsilon}{2}. \quad (47)$$

By the continuity of $R_p(s)$ and because α is an accumulation point of the point set $\alpha_1, \alpha_2, \dots, \alpha_n, \dots$, we can find an index $q > p$ such that

$$|R_p(\alpha_q) - R_p(\alpha)| < \frac{\epsilon}{2}. \quad (48)$$

It follows from (47) and (48) that

$$R_p(\alpha_q) < \epsilon.$$

Now because the functions $u_\nu(s)$ are assumed to be positive, $R_n(s)$ cannot be negative and for fixed x cannot grow with increasing n . Therefore, for $m > q > p$

$$0 \leq \text{Max}(R_m) = R_m(\alpha_m) \leq R_q(\alpha_m) \leq \text{Max}(R_q) = R_q(\alpha_q) \leq R_p(\alpha_q) < \epsilon.$$

Therefore

$$\lim_{n=\infty} \text{Max}(R_n) = 0,$$

which is what was to be proved.

CONCLUDING REMARKS.

Again, under the conditions stated in § 17, discontinuity in the kernel is permitted.

Likewise, nothing is changed in the theorems and proofs in this chapter, when s, t, r, \dots are points from a bounded n -dimensional domain consisting of a finite number of analytic pieces in an $(n + m)$ -dimensional space. Here ds, dt, dr, \dots are the corresponding elements.

Chapter V

On the Expansion of Arbitrary Functions by Prescribe Systems

§ 20.

The prescribed system of functions vanishes at the endpoints of the interval of definition

Let $\varphi_1(x), \varphi_2(x), \dots, \varphi_\nu(x), \dots$ be an infinite sequence of real, continuous, and twice continuously differentiable functions on the interval $a \leq x \leq b$ that furthermore vanish for both $x = a$ and $x = b$. In addition, let the system

$$\varphi_1''(x), \varphi_2''(x), \dots, \varphi_\nu''(x), \dots,$$

where we have written $\varphi_\nu''(x)$ for $\frac{s^2 \varphi_\nu(x)}{dx^2}$, be a *closed system*; that is (as we have stated in the introduction) a system for which there is no continuous function $f(x)$ other than the zero function for which the equation

$$\int_a^b f(x) \varphi_\nu''(x) dx = 0$$

is satisfied for all ν . We then form

$$\begin{aligned} \psi_1(x) &= \frac{\varphi_1(x)}{\sqrt{\int_a^b (\varphi_1'(y))^2 dy}} \\ \psi_2(x) &= \frac{\varphi_2(x) - \psi_1(x) \int_a^b \varphi_2'(z) \psi_1'(z) dz}{\sqrt{\int_a^b (\varphi_2'(y) - \psi_1'(y) \int_a^b \varphi_2'(z) \psi_1'(z) dz)^2 dy}} \\ &\vdots \\ \psi_\nu(x) &= \frac{\varphi_\nu(x) - \sum_{\rho=1}^{\rho=\nu-1} \psi_\rho(x) \int_a^b \varphi_\nu'(z) \psi_\rho'(z) dz}{\sqrt{\int_a^b (\varphi_\nu'(y) - \sum_{\rho=1}^{\rho=\nu-1} \psi_\rho'(y) \int_a^b \varphi_\nu'(z) \psi_\rho'(z) dz)^2 dy}} \\ &\vdots \end{aligned}$$

where we have written $\varphi_\nu'(x)$ and $\psi_\nu'(x)$ for $\frac{d\varphi_\nu(x)}{dx}$ and $\frac{d\psi_\nu(x)}{dx}$.

We also note the following. As was shown in § 3, a denominator in the above expression can vanish if and only if the corresponding function $\varphi'(x)$ can be represented as a homogeneous, linear combination with constant coefficients of its predecessors. But because

$$\varphi_\nu(a) = 0,$$

any such linear, homogeneous relation between the $\varphi_\nu'(x)$ remains valid for the $\varphi_\nu(x)$ and conversely. It therefore follows that a denominator vanishes if and only if the

corresponding function $\varphi_n(x)$ is linearly dependent on its predecessors. In this case we ignore the corresponding function $\varphi_n(x)$ and proceed with the formation of the functions $\psi_\nu(x)$ as if the function $\varphi_n(x)$ had never appeared in the first place. *Then by the above formulas, all $\psi_\nu(x)$ are represented homogeneous linear combinations of the $\varphi_\nu(x)$, and vice versa.*

Now let $g(x)$ be a continuously differentiable function in the interval $a \leq x \leq b$ that vanishes for $x = a$ and $x = b$. Then

$$g(x) = \sum_{\nu=1}^{\nu=\infty} \psi_\nu(x) \int_a^b g'(y) \psi'_\nu(y) dy,$$

and the sum on the right-hand side converges absolutely and uniformly.

PROOF. By § 3 we have the equations

$$\int_a^b \psi_\mu(x) \psi_\nu(x) dx = 1 \quad \text{or} \quad 0,$$

depending on whether μ and ν are equal or unequal. The absolute and uniform convergence of the series on the right-hand side of the equation to be proved follows from the the corollary of § 2. We now set

$$g(x) - \sum_{\nu=1}^{\nu=\infty} \psi_\nu(x) \int_a^b g'(y) \psi'_\nu(y) dy = f(x).$$

Then because

$$\int_a^b g(x) \psi''_\rho(x) dx = - \int_a^b g'(x) \psi'_\rho(x) dx$$

and

$$\int_a^b \psi_\nu(x) \psi''_\rho(x) dx = - \int_a^b \psi'_\nu(x) \psi'_\rho(x) dx = -1 \quad \text{or} \quad 0$$

depending on whether ν and ρ are equal or unequal, it follows that for each ρ

$$\int_a^b f(x) \psi''_\rho(x) dx = 0.$$

But since each function $\varphi''_\nu(x)$ can be represented as a homogeneous linear combination with constant coefficients of a finite number of the ψ''_ν , it follows that for each ν

$$\int_a^b f(x) \varphi''_\nu(x) dx = 0.$$

Hence the assumed closure of the system of the φ''_ν allows us to conclude that $f(x)$ vanishes identically—which is what was to be proved.

§ 21.

The general case.

Let $\varphi_1(x), \varphi_2(x), \dots, \varphi_\nu(x), \dots$ be an infinite sequence of real, continuous, and twice continuously differentiable functions on the interval $a \leq x \leq b$ but which are subject to no boundary conditions. Moreover let the system

$$\varphi_1''(x), \varphi_2''(x), \dots, \varphi_\nu''(x), \dots$$

be a closed system. If for each index ν we define

$$\bar{\varphi}_\nu(x) = \varphi_\nu(x) - \varphi_\nu(a) - \frac{x-a}{b-a} (\varphi_\nu(b) - \varphi_\nu(a)),$$

then the following equations hold:

$$\bar{\varphi}_\nu(a) = \bar{\varphi}_\nu(b) = 0$$

$$\bar{\varphi}_\nu''(x) = \varphi_\nu''(x).$$

Hence the system $\bar{\varphi}_1''(x), \bar{\varphi}_2''(x), \dots, \bar{\varphi}_\nu''(x), \dots$ is also a closed system. As in the previous section, we now construct the sequence of functions

$$\begin{aligned} \psi_1(x) &= \frac{\bar{\varphi}_1(x)}{\sqrt{\int_a^b (\bar{\varphi}_1'(y))^2 dy}} \\ &\vdots \\ \psi_\nu(x) &= \frac{\bar{\varphi}_\nu(x) - \sum_{\rho=1}^{\nu-1} \psi_\rho(x) \int_a^b \bar{\varphi}_\rho'(z) \psi_\rho'(z) dz}{\sqrt{\int_a^b (\bar{\varphi}_\nu'(y) - \sum_{\rho=1}^{\nu-1} \psi_\rho'(y) \int_a^b \bar{\varphi}_\rho'(z) \psi_\rho'(z) dz)^2 dy}} \\ &\vdots \end{aligned}$$

Then if $g(x)$ is an arbitrary continuous and continuously differentiable function in the interval $a \leq x \leq b$ and if we set

$$\bar{g}(x) = g(x) - g(a) - \frac{x-a}{b-a} (g(b) - g(a)),$$

$$\bar{g}(a) = \bar{g}(b) = 0$$

then the expansion theorem established in the previous paragraph implies that

$$\bar{g}(x) = \sum_{\nu=1}^{\nu=\infty} \psi_\nu(x) \int_a^b \bar{g}'(y) \psi_\nu'(y) dy,$$

or

$$g(x) = \frac{bg(a) - ag(b)}{b-a} + x \frac{g(b) - g(a)}{b-a} + \sum_{\nu=1}^{\nu=\infty} \psi_\nu(x) \int_a^b g'(y) \psi_\nu'(y) dy,$$

and the series on the right converges absolutely and uniformly.

In the proof of the expansion theorems in this and the previous section, we have assumed that the system of the $\varphi_\nu''(x)$ is a closed system. But it would have been sufficient to have assumed a little less — specifically, that any function that is orthogonal to all the $\varphi_\nu''(x)$ is linear. For the vanishing of $f(x)$, which was required by the proof given in the previous section, results from the fact that $f(x)$ must be linear because of the vanishing of $f(x)$ at the endpoints.

Since any continuous can be uniformly approximated by a continuously differentiable function, from the last theorem we have the following theorem. *Let $\varphi_1(x), \varphi_2(x), \dots, \varphi_\nu(x), \dots$ be an infinite sequence of real, twice continuously differentiable functions defined on $a \leq x \leq b$ whose second derivatives form a closed system. Then any continuous function defined in $a \leq x \leq b$ can be expanded in a uniformly convergent sequence of finite linear combinations of the functions $1, x, \varphi_1(x), \varphi_2(x), \dots, \varphi_\nu(x), \dots$.*