FRIGYES RIESZ and BÉLA SZ.-NAGY

FUNCTIONAL ANALYSIS

Translated from the 2nd French edition by

LEO F. BORON
Department of Mathematics, University of Michigan

FREDERICK UNGAR PUBLISHING CO.
NEW YORK

TRANSFORMATIONS WITH SYMMETRIC KERNEL

97. Theorems of Hilbert and Schmidt

Let us apply the theory of completely continuous symmetric transformations to the particular case of a transformation A of the functional space $L^2(a, b)$ which is generated by a *symmetric* or *Hermitian* kernel $A(x, y)$, that is, a kernel such that

$$A(x, y) = \overline{A(y, x)},$$

which belongs to the space $\boldsymbol{L}^2$ of square-summable functions in the plane domain $a \leq x \leq b$, $a \leq y \leq b$. By the remark made in Sec. 69, the transformation A can not be zero for all the elements of L^2 unless the kernel $A(x, y)$ *itself* is zero almost everywhere. Therefore we have the

THEOREM. *If the kernel $A(x, y)$ is not zero almost everywhere, the transformation A has a least one characteristic value different from 0, and each of its characteristic values is of finite multiplicity. There is an orthonormal sequence (finite or infinite) of characteristic functions $\varphi_i(x)$ of A corresponding to the characteristic values $\mu_i \neq 0$, and every function $g(x)$ belonging to L^2 admits the development, convergent in the mean:*

$$g(x) = h(x) + \sum_i (g, \varphi_i)\varphi_i(x) \tag{18}$$

where $h(x)$ is a function (depending on $g(x)$) such that $Ah(x) = 0$; consequently,

$$Ag(x) = \sum_i \mu_i(g, \varphi_i)\varphi_i(x). \tag{19}$$

We shall complete this theorem by several statements which are peculiar to the transformations being considered.

Since the functions

$$\Phi_i(x, y) = \varphi_i(x)\overline{\varphi_i(y)} \qquad (i = 1, 2, \ldots)$$

form an orthonormal system in $\boldsymbol{L}^2$, the series

$$\sum_i (A, \Phi_i)\Phi_i(x, y)$$

converges in the mean to some function $S(x, y)$ belonging to $\boldsymbol{L}^2$. We have

$$(A, \Phi_i) = \int_a^b\int_a^b A(x, y)\varphi_i(y)\overline{\varphi_i(x)}dxdy = (A\varphi_i, \varphi_i) = \mu_i;$$

hence

$$S(x, y) = \sum \mu_i \Phi_i(x, y).$$

For any f and g in L^2, the product $F(x, y) = g(x)\overline{f(y)}$ belongs to $\boldsymbol{L}^2$ and we have

$$(S, F) = \sum_i \mu_i(\Phi_i, F).$$

On the other hand, since $Ah = 0$ implies $(Af, h) = (f, Ah) = 0$, (18) furnishes:

$$(Af, g) = \sum_i (Af, \varphi_i)(\varphi_i, g) = \sum_i (f, A\varphi_i)(\varphi_i, g) = \sum \mu_i (f, \varphi_i)(\varphi_i, g);$$

hence

$$(A, F) = (Af, g) = \sum_i \mu_i (\Phi_i, F),$$

and consequently

$$(A, F) = (S, F).$$

The difference $A(x, y) - S(x, y)$ is therefore orthogonal to all these functions $F(x, y)$ and hence also to all the kernels of finite rank, and since these are everywhere dense in $\boldsymbol{L}^2$, it follows that $A(x, y) - S(x, y) = 0$ almost everywhere. From this we have the

Hilbert-Schmidt theorem

THEOREM.[7] *Every function $A(x, y)$ which is symmetric and square-summable can be developed, in the sense of convergence in the mean, into the series*

$$A(x, y) = \sum_i \mu_i \varphi_i(x) \overline{\varphi_i(y)}, \tag{20}$$

where $\{\varphi_i(x)\}$ denotes the orthonormal sequence of characteristic functions, and $\{\mu_i\}$ the sequence of corresponding characteristic values, of the transformation A generated by the kernel $A(x, y)$.

It follows in particular that

$$\| A(x, y) \|^2 = \sum_i \mu_i^2, \tag{21}$$

and further, denoting by $S_N(x, y)$ the N-th partial sum of the series (20), that

$$\| A(x, y) - S_N(x, y) \|^2 = \sum_{i > N} \mu_i^2.$$

Up to now we have made no hypothesis concerning the arrangement of the characteristic values μ_i. If they are arranged in such a way that their absolute values form a nondecreasing sequence, that is,

$$|\mu_1| \geq |\mu_2| \geq |\mu_3| \geq \ldots,$$

the partial sum $S_N(x, y)$ which is obviously a kernel of finite rank, possesses a remarkable minimum property:

For every symmetric kernel $A_N(x, y)$ of rank N, we have

$$\| A(x, y) - A_N(x, y) \| \geq \| A(x, y) - S_N(x, y) \|.$$

In fact, since the number of characteristic values of A_N which are different from 0 is at most equal to N, the characteristic value of degree $p + N$ of

[7] E. SCHMIDT [1].

$A = (A - A_N) + A_N$, that is, μ_{p+N}, can not exceed in absolute value the characteristic value of degree p of $A - A_N$, which we shall denote by χ_p (see Sec. 95, inequality (14)). Applying relation (21) to $A - A_N$ instead of A, we obtain

$$|A - A_N|^2 = \sum_{p=1}^{\infty} \chi_p^2 \geq \sum_{p=1}^{\infty} \mu_{p+N}^2 = |A - S_N|^2,$$

which was to be proved.

Let us return to the development (19), which is valid in the sense of convergence in the mean for all functions of the form

$$Ag(x) = \int_a^b A(x, y)g(y)dy.$$

It is important to know if, in certain cases, this development is also convergent in the ordinary sense or even absolutely and uniformly convergent. This will be the case in particular if there exists a constant C such that

$$\int_a^b |A(x, y)|^2 dy < C^2 \tag{22}$$

for all values of x. (Example: $A(x, y) = |x - y|^{-\alpha}$, $\alpha < \frac{1}{2}$.)

In fact, condition (22) implies that if the sequence $\{f_n(x)\}$ converges in the mean to $f(x)$, its transform, $\{Af_n(x)\}$, converges uniformly to $Af(x)$, since

$$|Af(x) - Af_n(x)|^2 = |\int_a^b A(x, y)\,[f(y) - f_n(y)]dy|^2 \leq$$

$$\leq \int_a^b |A(x, y)|^2 dy \cdot \int_a^b |f(y) - f_n(y)|^2 dy \leq C^2\,\|f - f_n\|^2.$$

But since the development (18) is convergent in the mean, development (19), which is derived from it by applying the transformation A to both sides, is uniformly convergent in the interval $a \leq x \leq b$.

Moreover, the convergence of development (19) is also absolute, that is, we can rearrange it in any arbitrary manner. In fact, ordinary convergence is, as we have just established, a consequence of mean convergence; but since mean convergence of a series with orthogonal terms $\Sigma\psi_i$ is equivalent to the convergence of the numerical series with non-negative terms $\Sigma\|\psi_i\|^2$, it does not depend on the arrangement of its terms.

Thus we have obtained the

THEOREM. *If the symmetric kernel* $A(x, y)$ *satisfies condition* (22), *the development* (19) *converges absolutely and uniformly, whatever the function* $g(x)$ *belonging to* L^2.

98. Mercer's Theorem

Development (20) of the kernel $A(x, y)$ is not necessarily uniformly convergent, even for continuous $A(x, y)$. However, for an important class of continuous kernels the convergence is uniform. We have namely the theorem of Mercer [**1**]:

Mercers Theorem

THEOREM. *If the transformation A generated by the continuous symmetric kernel $A(x, y)$ is positive, that is, if $(Af, f) \geq 0$ for all f, or, equivalently, if all the characteristic values $\mu_i \neq 0$ are positive, the development* (20) *is uniformly convergent.*

This theorem extends immediately to the case where all but a finite number of the $\mu_i \neq 0$ are of the same sign, positive or negative.

We observe first that since the kernel $A(x, y)$ is continuous, all the image functions

$$Af(x) = \int_a^b A(x, y)f(y)dy$$

are continuous; therefore, in particular, all the characteristic functions $\varphi_i(x) = \frac{1}{\mu_i} A\varphi_i(x)$ are continuous. Consequently the "remainders"

$$A_n(x, y) = A(x, y) - \sum_{i=1}^{n} \mu_i\varphi_i(x)\overline{\varphi_i(y)} \qquad (n = 1, 2, \ldots),$$

are also continuous functions. Since we have

$$A_n(x, y) = \sum_{i=n+1}^{\infty} \mu_i\varphi_i(x)\overline{\varphi_i(y)}$$

in the sense of mean convergence, it follows that

$$\int_a^b\int_a^b A_n(x, y)f(y)\overline{f(x)}dxdy = \sum_{i=n+1}^{\infty} \mu_i(\varphi_i, f)\,(f, \varphi_i) \geq 0 \tag{23}$$

for every element f of L^2.

From this we deduce that $A_n(x, x) \geq 0$. In fact, if we had $A_n(x_0, x_0) < 0$, we should have by continuity $A_n(x, y) < 0$ in a neighborhood

$$x_0 - \delta < x < x_0 + \delta, \quad x_0 - \delta < y < x_0 + \delta$$

of the point (x_0, x_0). Setting $f(x) = 1$ for $x_0 - \delta < x < x_0 + \delta$ and $f(x) = 0$ elsewhere, integral (23) would become negative, a contradiction.

Hence we have

$$A_n(x, x) = A(x, x) - \sum_{i=1}^{n} \mu_i\varphi_i(x)\overline{\varphi_i(x)} \geq 0$$

for $n = 1, 2, \ldots$. From this we conclude that the series of positive terms

$$\sum_{i=1}^{\infty} \mu_i\varphi_i(x)\overline{\varphi_i(x)}$$

is convergent and that its sum is $\leq A(x, x)$. Denoting by M the maximum of the continuous function $A(x, x)$, we have by Cauchy's inequality:

$$\left|\sum_{i=m}^{n} \mu_i\varphi_i(x)\overline{\varphi_i(y)}\right|^2 \leq \sum_{i=m}^{n} \mu_i\,|\varphi_i(x)|^2 \sum_{i=m}^{n} \mu_i\,|\varphi_i(y)|^2 \leq M \sum_{i=m}^{n} \mu_i\,|\varphi_i(x)|^2. \tag{24}$$

From this it follows that the series

$$\sum_{i=1}^{\infty} \mu_i \varphi_i(x)\overline{\varphi_i(y)} \tag{25}$$

converges, for every fixed value of x, uniformly in y; its sum $B(x, y)$ is therefore a continuous function of y, and for every continuous function $f(y)$ we have

$$\int_a^b B(x, y) f(y) dy = \sum_{i=1}^{\infty} \mu_i \varphi_i(x) \int_a^b \overline{\varphi_i(y)} f(y) dy.$$

Now by one of the theorems proved in the preceding section, the series in the second member converges to $Af(x)$. Hence we have

$$\int_a^b [B(x, y) - A(x, y)] f(y) dy = 0;$$

setting in particular $f(y) = \overline{B(x, y) - A(x, y)}$ (for a fixed value of x), it follows that $B(x, y) - A(x, y) = 0$ for $a \leq y \leq b$, hence

$$A(x, x) = B(x, x) = \sum_{i=1}^{\infty} \mu_i \,|\, \varphi_i(x) \,|^2.$$

Since the terms of this series are positive continuous functions of x and its sum $A(x, x)$ is a continuous function, it follows from a known theorem of Dini that the series converges *uniformly*. Applying Cauchy's inequality (24) again, we deduce from this that series (25) converges uniformly with respect to its two variables x and y simultaneously, which was to be proved.

Whatever be the continuous symmetric kernel $A(x, y)$, its iterate

$$A_2(x, y) = \int_a^b A(x, z) A(z, y) dz$$

is continuous and of positive type. In fact,

$$(A_2 f, f) = (A^2 f, f) = (Af, Af) \geq 0.$$

The characteristic functions $\varphi_i(x)$ of A are also characteristic functions for A^2, but they correspond to the squares of the characteristic values μ_i of A:

$$A^2 \varphi_i = A(A\varphi_i) = A(\mu_i \varphi_i) = \mu_i^2 \varphi_i.$$

The sequence $\mu_1^2, \mu_2^2, \ldots$ contains all the characteristic values of A^2 different from 0, each as many times as its multiplicity indicates. If not, there would be a characteristic function φ corresponding to a characteristic value $\mu \neq 0$ of A^2 and orthogonal to all the φ_i. This would be in contradiction to the fact that

$$\mu\varphi = A^2\varphi = \sum_{i=1}^{\infty} (A^2\varphi, \varphi_i)\varphi_i = \sum_{i=1}^{\infty} (\varphi, A^2\varphi_i)\varphi_i = \sum_{i=1}^{\infty} \mu_i^2 (\varphi, \varphi_i)\varphi_i = 0.$$

By the theorem of Mercer we therefore have, for the iterate of an arbitrary continuous kernel $A(x, y)$, the *uniformly convergent* development:

$$A_2(x, y) = \sum_{i=1}^{\infty} \mu_i^2 \varphi_i(x)\overline{\varphi_i(y)}.$$