ON THE ZEROS AND CLOSURE OF ORTHOGONAL FUNCTIONS

BY

D. D. KOSAMBI.

[Received 2 September, 1941]

This note deals primarily with infinite sets of continuous functions $\phi_n(x)$ uniformly bounded and orthonormal over (0,1), i.e. the sequence being $\phi_1, \phi_2, \ldots, \phi_n$,

$$\int_0^1 \phi_i \phi_j dx = \delta_{ij}, \ |\phi_n| \leqslant M \text{ for all } n, \ 0 \leqslant x \leqslant 1. \qquad (1)$$

An elementary knowledge of the subject, such as is contained in a standard textbook like vol. I of Courant-Hilbert: *Methoden d. Math. Physik* is assumed for the sake of brevity. Only main details of proofs are given, and results obviously extensible to L-integrable functions are stated only for continuous functions in order to keep the treatment relatively simple.

Without further explicit statement, we use the following well-known relationship between ordinary convergence and convergence in the mean. A boundedly convergent sequence $\{ f_n(x) \}$ also converges in the mean almost everywhere to the same limit function, and for a sequence that c.i.m., a subsequence can be found to converge almost everywhere to a summable function.

1. We can state at once, a simple but not quite trivial theorem about the average behavior of $\{ \phi \}$.

THEOREM 1. *No sequence* $\{ \phi \}$ *can converge (c.i.m.) over* (0,1); *nor to any value except zero over any subinterval of* (0,1). *But every such sequence is summable in the mean to zero by most regular methods of summability over* (0,1) *or any subinter-*

val thereof.

PROOF. Convergence implies $\int_0^1 (\phi_n - \phi_m)^2 dx \to 0$. But for distinct orthonormal functions, the integral always has the value 2. For the second part, let (a, b) be any subinterval of $(0,1)$ and consider $\sigma_n = \frac{\phi_1 + \phi_2 + \cdots + \phi_n}{n}$.

Now $$\int_a^b \sigma_n^2 dx \leqslant \int_0^1 \sigma_n^2 dx = \frac{1}{n} \to 0.$$

This extends to any Toeplitz Matrix where the sum of squares of elements in the ith row tends to 0 as $i \to \infty$. If $\{\phi\}$ converged over a subinterval, it would be summable to the same limit function.

As a corollary, we see that no $\{\phi\}$ can be uniformly continuous in n over the whole of $(0,1)$, nor have a difference quotient uniformly bounded for all n. Or else, a subsequence could be found to converge over $(0,1)$. That is, while every infinite set of orthogonal functions averages out to zero for any permissible method of averages, the randomness increases in general with the index.

From Bessel's inequality, $\sum c_n^2$ converges, $c_n = \int_0^1 f\phi_n dx$. The same is true if $c_n = \int_a^b f\phi_n dx$ for every (a, b) in $(0,1)$, as is seen by taking $f(x)$ as zero outside (a, b). By taking $f > 0$, in (a, b) we see that the necessary condition $c_n \to 0$ cannot be satisfied unless $\phi_n \to 0$ or the ϕ's change sign, for every (a, b) in $(0,1)$.

2. It is easy to show that *for closure, infinitely many of the functions ϕ_n must have at least one change of sign each in every subinterval (a, b) however small.* From the Parseval equality, taking the functions as zero outside the proper interval, it follows that *for a closed $\{\phi\}$ and any integrable f, $\sum c_n c_n' = \int f^2 dx$, where $c_n = \int_a^b f\phi_n dx, c_n' = \int_{a'}^{b'} f\phi_n dx$ and the integral of f^2 above is taken over the region common to the two*

intervals (a, b), (a', b'). Now suppose that only a finite number of the ϕ_n have any change of sign in some subinterval, whence a still smaller interval could be found such that no ϕ changes sign therein. Let this be subdivided into two, (a'', ξ), (ξ, b''). For a positive f, the associated coefficients c_n, c'_n must both have the same sign. But by the above lemma, $\Sigma c_n c'_n = 0$, which establishes a contradiction and proves our theorem.

If $\phi_n \not\to 0$ over some (a, b), then this tendency must not be too strong if $\{\phi\}$ is to be closed. In fact, suppose $|\phi_n| \leqslant \delta_n \to 0$ over (a, b). Taking $f = 1$ in (a, b), $f = 0$ outside, we have $\Sigma c_n^2 = b - a \leqslant (b-a)^2 \Sigma \delta_n^2$. This leads to $1 \leqslant (b-a)\Sigma\delta_n^2$. Decreasing the interval does not increase any δ_n, hence, if $\Sigma\delta_n^2$ converges, the right side can be made arbitrarily small. *For closure of* $\{\phi\}$, $\Sigma\delta_n^2$ *and* $\sum\int_a^b \phi_n^2 dx$ *must both diverge.* The divergence of the series of integrals follows from Schwarz's inequality $\left(\int_a^b f^2 dx\right) \times \left(\int_a^b \phi_n^2 dx\right) \geqslant c_n^2$ and from the closure property $\sum c_n^2 = \int_a^b f^2 dx$. It is a consequence of these that $\sum\int_a^b \phi_n^2 dx \geqslant 1$. Now, if the series of integrals converge for some (a, b) the remainder after a sufficiently large but fixed number of terms may be made arbitrarily small; the initial terms, no matter how many, being fixed in number, can also be made to have an arbitrarily small sum by reducing the interval—which does not increase the remainder. Thus, the left side could be made as small as desired, but would still remain greater than unity, which leads to a cont adiction in case $\sum\int_a^b \phi_n^2 dx$ converges.

It would seem intuitively probable that the number of changes of sign must not be too great, if $\{\phi\}$ is to be closed. Let μ_n be defined as the number of changes of

sign of ϕ_n in (0,1). For the common orthogonal functions of mathematical physics, we usually have $\mu_n = n-1$. This can be proved directly from the differential equations, which have boundary conditions making the end points of the interval conjugate. For the nth function of the set, the previous $n-1$ conjugate points move inside the fundamental interval and furnish precisely $n-1$ changes of sign.

While this is too strong a result for the general type of ϕ, something like it holds for polynomials, at least of a particular kind. For those constructed for interpolation, it is again usually true that $\mu_n = n-1$. For a more general case, we can take, according to a classic theorem of Muntz, any set of powers $\{x^{p_n}\}$, which is closed if (and only if) $\Sigma 1/p_n$ diverges. The set of orthonormal functions constructed from these by the usual process will also be closed. But by Descartes's rule of signs, the nth of these functions (polynomial with n terms) cannot have more than $n-1$ positive roots, hence $\mu_n \leqslant n-1$ even here.

3. Before proceeding to the general case, we consider a set of functions that operate by change of sign alone, without the restriction of orthogonality.

Let $\{x_n\}$ be a set of distinct points everywhere dense in (0,1), with $0<x_n<1$ for all n. Let $\gamma_r(x)$ be defined over (0,1) by $\gamma_r = 1$, $0 \leqslant x \leqslant x_r$ and $\gamma_r = 0$, $x_r < x \leqslant 1$. The set $\{\gamma\}$ is closed over (0,1) as $\int_0^1 f\gamma_n dx = 0$ for all n implies that the integral of f vanishes for a set of points everywhere dense in (0,1), i.e. that the function itself vanishes almost everywhere in the fundamental interval.

We now consider a set $\{\psi\}$ constructed as follows:

$0<x_{i_1}<x_{i_2}<\ldots<x_{i_r}<1$ being distinct points of the set $\{x_n\}$ above, ψ_i assumes alternately the values ± 1 over the intervals (inclusive of the right hand end point for each) marked by 0, $x_{i_1}, x_{i_2}, \ldots, x_{i_r}$, 1. That is,

$$\psi_i = 1+2(\gamma_{i_1}-\gamma_{i_2})+2(\gamma_{i_3}-\gamma_{i_4})+\cdots+\begin{cases} 2(\gamma_{i_{r-1}}-\gamma_{i_r}), & r \text{ even.} \\ 2(\gamma_{i_r}-1), & r \text{ odd.} \end{cases} \tag{2}$$

Finally, let it be required by hypothesis that any finite number of the ψ's are linearly independent as also the γ's and that every point of $\{x_n\}$ is ultimately used for the construction of the ψ's. We can then state

THEOREM 2. *A necessary and sufficient condition for the closure of* $\{\psi\}$ *is that the number of distinct points of the set* $\{x_n\}$ *needed for the construction of* n *functions of* $\{\psi\}$ *should be equal to* $n-1$ *infinitely often.*

Formula (2) shows that each ψ is composed of the function 1 and as many of the γ's as it has changes of sign. Because of linear independence, no set of n of the ψ's can be composed of less than n of the component functions 1, $\{\gamma\}$. From this, the condition of the theorem is seen to be sufficient for closure, as it implies that n of the component functions can be solved for in terms of the ψ's infinitely often, and as every x_n is used up, every γ can ultimately be expressed as a linear combination of the ψ's.

The condition is also necessary, as a contradiction would follow if $\{\psi\}$ were closed without it. The least unfavourable case would be that $\psi_1, \psi_2, \psi_3, \ldots, \psi_n$ would need, for all large n, at least $n+1$ of the component functions. To solve for the component functions would then ultimately need all the ψ's and at least one extra function ψ_0. That is, the set $\{\psi\}$ cannot be closed without the inclusion of at least one more function, ψ_0.

The function 1 is not itself included in the set considered in Theorem 2. If this be included, we have a corollary, that a necessary and sufficient condition for the closure of the enlarged set is that n of the ψ's, with the exclusion of 1, should require just n points of $\{x_n\}$ infini-

tely often. This covers the case of the orthonormal system constructed (Kaczmarz, Walsh, Paley) by multiplication of the proper number of Rademacher functions. There, $\psi_0 = 1$, and $\psi_1, \psi_2, \psi_3, \ldots, \psi_n$ have changes of sign at points all included in the points dividing $(0,1)$ into $n+1$ equal sub-intervals, for $n = 2^k - 1$.

4. The preceding theorem rests on the lemma that if n functions of one set can be expressed as linearly independent linear combinations of n of another set *infinitely often,* the two sets must be closed together. For the general case, we need something similar, with the substitution of approximation for exact equality. This is given by

THEOREM 3. *If* $\{\phi\}$ *is closed, there exists at least one closed set of powers* $\{x^{p_n}\}$ *such that it is possible to approximate simultaneously to r functions of* $\{\phi\}$ *by means of exactly r powers of the set for any given degree of approximation, and infinitely many values of r.*

One such process of simultaneous approximation can be built up as follows. Take a set of positive, distinct, decreasing constants λ_i such that $\Sigma\lambda_i^2$ converges and $\Sigma\lambda_i\phi_i(s)\phi_i(t)$ converges uniformly in the unit square. Then the series defines the symmetric kernel $K(s,t)$ of the integral equation

$$\int_0^1 K(s,t)\phi(t)\,dt = \lambda\phi(s).$$

The kernel is positive definite by construction, without any multiple characteristic values, provided, of course, $\{\phi\}$ is closed as in our present hypothesis. Our desired process of approximation is now derived by approximating to $K(s,t)$ by symmetric polynomials $K_n(s,t)$. The degenerate kernels K_n then have characteristic functions and characteristic values of λ which approximate to func-

tions of $\{\phi\}$ and the corresponding λ_i. Moreover, for any kernel, the (approximating) characteristic polynomials are orthonormal.

An important property of this process is that a subsequence at least of the orthonormal polynomials obtained for K_n always tends to the appropriate limit as $\lambda_i^{(n)} \to \lambda_i$ and $K_n \to K$. There are as many mutually orthogonal polynomials as there are distinct powers of the variable in the degenerate kernel.

If, now, the theorem is assumed false, we can again show a contradiction. Suppose that for all large n, $\phi_1, \phi_2, \phi_3, \ldots, \phi_n$ need at least $n+1$ powers for the purpose of simultaneous approximation. Then there exists at least one extra characteristic function for each of the approximating kernels, and a subsequence of these can be chosen so as to tend to a limiting function. But this extra function cannot tend to any fixed ϕ_i, as there would then be two distinct function sequences tending to the same limit, yet mutually orthogonal. The extra function cannot approximate successively to some ϕ_{n+p} because this extra function (in a sub-sequence) tends to a limit, which is impossible for any sub-sequence of $\{\phi\}$ by Theorem 1. So, there would exist a function in L^2 orthogonal to all the ϕ's but not equivalent to zero, as it has a unit norm. This contradicts the hypothesis of closure, and so, the theorem is established.

This result can be applied to suitably restricted $\{\phi\}$ in order to put a limit on μ_n. We introduce a new definition:

DEFINITION: *A set ϕ is said to be strongly random if* (*a*) *for any* (a,b) *in* $(0,1)$ *lim inf* $\int_a^b \phi_n^2 dx \geqslant k(b-a)$, *with* $k>0$, *and* (*b*) *for all large* n, *max* $|\phi_n| \geqslant K>0$ *between any two consecutive changes of sign of* ϕ_n.

It must be kept in mind that $\{\phi\}$ is a uniformly bounded continuous, closed, orthonormal set. The present definition gives restrictions that make the graph of ϕ_n arch away from the x-axis, preventing convergence to zero. It makes $\mu_n \to \infty$, forbids vacuoles (rectangle with sides parallel to the axes such that no graph of any ϕ passes through it; their absence is a fair substitute for almost-everywhere denseness of the values of $\{\phi\}$) in the rectangle formed in the x, y plane by $0 \leqslant x \leqslant 1$, $-K \leqslant y \leqslant K$. The behaviour of such functions is, then, close to that of the functions ψ of Theorem 2, in a general way. The arching away from the x-axis enables us to represent all the changes of sign in the approximating polynomials. If the approximating polynomial failed to follow one of the "arches", across the axis of x, then by the second condition for strong randomness, it would fail to be a good approximation in the ordinary sense. The first condition would also prevent it from being a good approximation in the mean. This would lead to a contradiction if the set admits simultaneous approximation. We now apply Theorem 3 and Descartes's rule of signs as before, to get our final result:

THEOREM 4. *A necessary condition for the closure of a strongly random set of orthonormal functions is that* $\mu_n \leqslant n-1$ *infinitely often.*

That is, the set cannot be closed if $\mu_n \geqslant n$ for all large n. In particular, as n/μ_n does not tend to zero even if the μ_n be in non-decreasing order, a necessary condition for the closure of $\{\phi\}$ is that $\sum \frac{1}{\mu_n}$ diverge. Also, if no two ϕ's of a closed set have the same number of changes of sign, $\mu_n = n-1$ with a suitable rearrangement of $\{\phi\}$ if necessary. Clearly, no condition of this type can possibly be sufficient.

As the discussion was meant to be elementary, no attempt has been made to obtain "best possible" results. For a strongly random set, μ_n is a fair measure of the "randomness". The content of Theorem 4, is that this, which we might call the entropy of the functions, cannot increase too rapidly if the functions are to enable us to represent all possible states.

[I apologise to the reader if circumstances beyond my control should make it impossible for me to prevent the simultaneous publication abroad of this note.]