Metric Entropy, Widths, and Superpositions of Functions
Author(s): G. G. Lorentz
Source: *The American Mathematical Monthly*, Vol. 69, No. 6 (Jun. - Jul., 1962), pp. 469-485
Published by: Taylor & Francis, Ltd. on behalf of the Mathematical Association of America
Stable URL: https://www.jstor.org/stable/2311185
Accessed: 09-03-2020 22:15 UTC

is sound teaching of each of the distinct, but closely related, aspects of mathematics which I have discussed in this paper. If I have been critical of some of the present trends, I do so as a supporter of the movement and not as its enemy. There is still much more work that needs to be done, and I hope that by these efforts we can bring about the extinction of the genus of "Narrow Mathematicians."

METRIC ENTROPY, WIDTHS, AND SUPERPOSITIONS OF FUNCTIONS*

G. G. LORENTZ, Syracuse University

1. Introduction. The purpose of this paper is to give an introduction to some recent developments connected with properties of compact sets of continuous functions. These developments, because of their importance on one hand and their simple and basic character on the other deserve to be more widely known. It is impossible in this paper to give complete proofs, but we will at least explain their main ideas. Our bibliography is not complete; it is restricted to papers most useful for the first reading and to a few papers of historical interest.

It is convenient to begin with the well-known theorems of D. Jackson and S. Bernstein (compare [1] and [9]). We wish to approximate continuous 2π-periodic functions $f(x)$, defined on the real line, by trigonometric polynomials of degree n

$$T_n(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx);$$

the degree of approximation is measured in the uniform norm

$$\|f - T_n\| = \max_x |f(x) - T_n(x)| .$$

We want this deviation of f from T_n to be as small as possible; an elementary fact is that there is always a polynomial $\tilde{T}_n$ of best approximation, for which $\|f - T_n\|$ attains its minimum. The quantity

$$E_n(f) = \min_{T_n} \|f - T_n\| = \|f - \tilde{T}_n\|$$

is called the degree of approximation of f (by the trigonometric polynomials).

Let $\Lambda(p+\alpha, M)$, where $p=0, 1, \cdots, 0<\alpha\leqq 1, M>0$ be the set of all

* Written with the support of the United States Air Force through the AFOSR, under Contract no. AF 49(638)-619. This paper is based on an invited address delivered at the Annual Meeting of the Mathematical Association of America in Washington, D. C., January 1961.

periodic continuous functions f, p times differentiable, and such that the pth derivative satisfies a Lipschitz condition of order α: $|f^{(p)}(x)-f^{(p)}(x')| \leqq M|x-x'|^{\alpha}$. Then Jackson's theorem asserts that if $f\in\Lambda(p+\alpha, M)$, then, as a reward for this smoothness, f has a small degree of approximation:

$$E_n(f) \leqq \text{Const. } n^{-(p+\alpha)}. \tag{1}$$

The theorem of Bernstein asserts that inversely, if (1) is satisfied for all n, then f is smooth: $f\in\Lambda(p+\alpha, M)$.

We shall look upon these theorems as dealing with classes of functions. For example, Bernstein's theorem can be explained by saying that the class $\Lambda(p+\alpha, M)$ has such a great thickness, width, massivity, that it is not possible to approximate all its functions too well by trigonometric polynomials of a given degree n. One will notice that this formulation carries with it the conjecture that the degree of approximation will not improve substantially if we replace the trigonometric polynomials by other means of approximation. And such is indeed the case.

The dealing with classes of functions rather than with individual functions is also justified, I am sure, from the point of view of the modern computer. He has often to compute many functions at once, functions whose properties he does not know well, or has no time to investigate. He will want to base his approximation methods on properties which are common to all functions of his class, without being interested to know that for some of them a better approximation exists.

2. Entropy and capacity.

2.1. Let A be a compact metric space with a metric ρ. The following definitions are well known. A finite set of points $x_1, \cdots, x_p$ of A is called an *ϵ-net in A* if $\epsilon>0$ and if for each $x\in A$ there is at least one point x_i of the net at a distance from x not exceeding ϵ: $\rho(x, x_i)\leqq\epsilon$. A family $U_1, \cdots, U_n$ of sets is an *ϵ-covering of A* if $A\subset\bigcup U_i$ and if the diameter of each set U_i does not exceed 2ϵ. It is convenient to take here 2ϵ and not ϵ in order to simplify the relation between the two definitions. In fact, if $x_1, \cdots, x_p$ is an ϵ-net for A, then the balls with centers x_i and radius ϵ form an ϵ-covering of A. A standard theorem of topology guarantees that a compact metric space A has a (finite) ϵ-net for each $\epsilon>0$. Hence A has also a finite ϵ-covering for each $\epsilon>0$.

Of course, the number n of sets U_i in a covering family depends on its choice, but the minimal value of n, $N_\epsilon(A)=\min n$ is an invariant of the set A which depends on $\epsilon>0$. The logarithm

$$H_\epsilon(A) = \log N_\epsilon(A) \tag{1}$$

is the entropy of the set A.* It was Kolmogorov's idea [11, 13] to characterize the "massiveness" of a set A by means of this function. We are interested, of

* There is a relation of analogy between $H_\epsilon(A)$ and the *probabilistic entropy* [7]. To underline the distinction, we propose to call $H_\epsilon(A)$ the *metric entropy* of A.

course, in the asymptotic behavior of $H_\epsilon(A)$ for $\epsilon \to 0$; in general, $H_\epsilon(A)$ will increase rapidly to infinity. We take logarithms in (1) partly because $N_\epsilon(A)$ is often unwieldily large.

Points $y_1, \cdots, y_m$ of A are called *ϵ-distinguishable* if the distance between any two of them exceeds ϵ. Again, their number m depends on the choice of the points, but the maximal value of m (which exists, as we shall see in a moment), $M_\epsilon(A) = \max m$ is an invariant of A. We put

$$C_\epsilon(A) = \log M_\epsilon(A); \tag{2}$$

this is the *capacity* of A.

The general theory of entropy and capacity is not rich; the main result is the following simple

THEOREM 1. *For each compact metric set A and each $\epsilon > 0$,*

$$C_{2\epsilon}(A) \leqq H_\epsilon(A) \leqq C_\epsilon(A). \tag{3}$$

The reasons for these inequalities are easy to see. Let $y_1, \cdots, y_m$ be a 2ϵ-distinguishable set of points of A, and let $U_1, \cdots, U_n$ be an ϵ-covering of A. Then $m \leqq n$, for otherwise at least one set U_i would contain two points y and consequently would have a diameter greater than 2ϵ. We have therefore $M_{2\epsilon}(A) \leqq N_\epsilon(A)$. (It follows also that $M_{2\epsilon}(A)$ is finite for each 2ϵ.) On the other hand, if $z_1, \cdots, z_{m'}$ is a maximal set of ϵ-distinguishable points (with $m' = M_\epsilon(A)$), then it is also an ϵ-net in A, for otherwise there would exist a point z in A with $\rho(z, z_i) > \epsilon$ for all i, and this would contradict the maximality of the set of the z_i's. As we know, there is then an ϵ-covering of A which consists of m' sets. Hence $N_\epsilon(A) \leqq M_\epsilon(A)$, and taking logarithms in the established relations, we obtain (3).

The exact determination of the entropy and the capacity of a set A is in most cases very difficult. Often one is satisfied to compute them only up to a strong or a weak equivalence; we write $f(\epsilon) \sim g(\epsilon)$ or $f(\epsilon) \approx g(\epsilon)$ for $\epsilon \to 0$ if $f(\epsilon)/g(\epsilon) \to 1$ or if respectively this quotient remains between two positive constants for $\epsilon \to 0$. In this connection Theorem 1 is used in the following way: one tries to find an upper bound for $H_\epsilon(A)$ (often also for the logarithm of the number of points of an ϵ-net in A) and a lower bound for $C_{2\epsilon}(A)$; if the bounds are close to each other, a good estimation of both entropy and capacity will result.

We shall discuss the most important sets A for which entropies are known (see [**13**, **18**]). Of course, each time not only the set A, but also its metric must be given.

2.2. *s-dimensional spaces.* If A is a bounded closed set with interior points in the s-dimensional euclidean space R^s, then

$$N_\epsilon(A) \approx \epsilon^{-s}, \quad \text{hence} \quad H_\epsilon(A) = s \log \frac{1}{\epsilon} + O(1). \tag{4}$$

If A has some regularity, for instance if it is an s-dimensional parallelopiped with measure $|A|$, then more exactly

$$N_\epsilon(A) \sim \nu_s |A| \epsilon^{-s}, \qquad M_\epsilon(A) \sim \mu_s |A| \epsilon^{-s} \tag{5}$$

with some constants ν_s, μ_s. All this is quite elementary.But the calculation of ν_s, μ_s for $s>2$ is a difficult (unsolved) problem of number theory; it is connected with the determination of the tightest packing of balls into R^s, and of the most economical covering of R^s by balls.

2.3. *Smooth functions.* Let B be an s-dimensional parallelopiped in R^s. A function $f(x)=f(x_1, \cdots, x_s)$ on B is smooth with the degree of smoothness $p+\alpha$, where $p=0, 1, \cdots, 0<\alpha\leqq 1$, if it has on B all partial derivatives of orders less than or equal to p. In addition, each partial derivative of order p, $D^pf(x_1, \cdots, x_s)$ must satisfy a Lipschitz condition with exponent α:

$$| D^pf(x_1, \cdots, x_s) - D^pf(x_1', \cdots, x_s') | \leqq M \max_{i=1,\cdots,s} |x_i - x_i'|^\alpha. \tag{6}$$

For example, if $p=0$, we obtain the class Lip α.

The metric will be determined by the uniform norm on B: $\|f\| = \max_{x\in B} |f(x)|$. But the classes just defined are not compact in this metric, and have no entropy. We consider therefore the smaller sets

$$\Lambda = \Lambda(p + \alpha, s, C_0, \cdots, C_p, M).$$

The set Λ consists of all functions f on B, smooth of degree $p+\alpha$, for which all kth derivatives D^kf, $k=0, \cdots, p$ satisfy $|D^kf| \leqq C_k$, while all pth derivatives satisfy in addition (6).

THEOREM 2 [**13, 18**]. *With the metric given by the uniform norm, one has*

$$H_\epsilon(\Lambda(p + \alpha, s, C_0, \cdots, C_p, M)) \approx \epsilon^{-s/(p+\alpha)} \tag{7}$$

the same is true for the capacity $C_\epsilon(\Lambda)$.

Consider first the case $s=1$, when B is a segment $[a, b]$. Let $\Delta>0$ be a small number. The points $x_1=a+\Delta, \cdots, x_n=a+n\Delta$, where n is so chosen that $a+n\Delta\leqq b<a+(n+1)\Delta$, form not only a Δ-net for B, but also a Δ-*chain*: any two consecutive points x_k, x_{k+1} are at a distance $\leqq\Delta$ from each other. The only computational device available for functions $f\in\Lambda$ is Taylor's formula; it is clear therefore that we must use it. This formula gives for instance the value $f(x_{k+1})$, if the values of the function f and of its derivatives at the preceding point x_k are known:

$$f(x_{k+1}) = f(x_k) + \frac{\Delta}{1} f'(x_k) + \cdots + \frac{\Delta^p}{p!} f^{(p)}(x_k) + \frac{\Delta^p}{p!} [f^{(p)}(\xi) - f^{(p)}(x_k)], \tag{8}$$

$$x_k < \xi < x_{k+1},$$

Let us first try to estimate $H_\epsilon(\Lambda)$ from above; for this purpose we must find a fairly economical covering of Λ by sets U of diameter $\leqq 2\epsilon$. The formula (8) gives us the idea to let each U consist of functions $f \in \Lambda$ for which the derivatives $f^{(i)}(x_k)$, $i=0, 1, \cdots, p$; $k=1, \cdots, n$, are given with an error not exceeding small fixed numbers $\epsilon_i > 0$. More formally, each U will be given by the inequalities

$$m_{ik}\epsilon_i \leqq f^{(i)}(x_k) < (m_{ik}+1)\epsilon_i. \tag{9}$$

Thus each set of integers m_{ik} determines a set U; but some of the U may be empty.

How should we select the ϵ_i? Surely in such a way that if we substitute $m_{ik}\epsilon_i$ for $f^{(i)}(x_k)$ into (8), and zero for the last term, the errors of all terms would be approximately equal: it does not make sense to compute some terms of a sum very carefully if large errors are committed in others. This leads to the formula $\epsilon_i = C\epsilon\Delta^{-i}$. An easy calculation shows that the diameter of each U indeed does not exceed 2ϵ if one selects the constant C properly and puts $\Delta = \epsilon^{1/(p+\alpha)}$.

It remains to estimate the number N of sets U which contain functions from Λ. For each k, for instance for $k=1$, it is easy by means of $|f^{(i)}(x_1)| \leqq C_i$ to estimate the number of the possible integers m_{i1} from above; one obtains that the number of m_{i1} with a given i does not exceed $2C_i/\epsilon_i + 2$. This gives us an upper bound for all possible combinations of the m_{i1}, $i=0, \cdots, p$ the number const. $\prod_{i=0}^{p} \epsilon_i^{-1}$. If done for each k, this would give an unsatisfactorily large upper bound for N. Instead, we do this only for $k=1$; for the following values of k, the smoothness of our function and Taylor's formula give a much more precise information about the number of possible integers m_{ik} once the $m_{i,k-1}$ are known. Thus one obtains $N \leqq C^n$, hence $H_\epsilon(\Lambda) \leqq \text{const.}\, n \leqq \text{const.}\, \Delta^{-1}$ $\leqq \text{const.}\ \epsilon^{-1/(p+\alpha)}$. We do not discuss the estimation of $C_{2\epsilon}(A)$ from below: this is much simpler.

In this way we obtain (7) for $s=1$. For larger values of s, the computation remains the same; the only difference is that now a Δ-chain in B consists of $\approx \Delta^{-s}$ points.

2.4. *Functions of bounded variation.* The set V of all functions f of bounded variation on $[a, b]$, restricted by $|f(x)| \leqq 1$, $\operatorname{Var} f \leqq 1$ is not compact in the uniform norm. But if the distance $\rho(f_1, f_2)$ is defined as the Hausdorff distance [**8**, p. 166] between the curves $y=f_1(x)$, $y=f_2(x)$, G. F. Clements (Syracuse University) obtains $H_\epsilon(V) \approx \epsilon^{-1}$. If V_α is the intersection of V with $\operatorname{Lip}_1 \alpha$, $0<\alpha<1$, then in the uniform norm, $H_\epsilon(V_\alpha) \approx \epsilon^{-1} \log 1/\epsilon$.

2.5. *Analytic functions.* Classes of analytic functions have smaller entropy than the classes Λ. We begin with $A_r(C)$, $r>1$, this consists of all functions $f(z)$ analytic in $|z|<r$ and continuous in $|z| \leqq r$, with the absolute value in $|z| \leqq r$ bounded by C, the metric being given by the uniform norm on $|z| \leqq 1$.

THEOREM [13, 18]. *Both the entropy $H_\epsilon(A_r)$ and the capacity $C_\epsilon(A_r)$ are given by*

$$\text{(10)} \qquad \frac{\log^2 \dfrac{1}{\epsilon}}{\log r} + \mathcal{O}\left(\log \frac{1}{\epsilon} \log \log \frac{1}{\epsilon}\right).$$

To obtain this result, we use the Taylor series representations

$$\text{(11)} \qquad f(z) = \sum_{k=0}^{\infty} c_k z^k \qquad |z| < r$$

of functions $f \in A_r$ and the following facts about their Taylor coefficients c_k:

$$\text{(12)} \qquad \|f\| \leqq \sum_0^\infty |c_k|; \qquad \text{(12a)} \qquad \max_{|z| \leqq r} |f(z)| \leqq \sum_0^\infty |c_k| r^k;$$

$$\text{(13)} \qquad |c_k| \leqq \|f\|; \qquad \text{(13a)} \qquad |c_k| \leqq C r^{-k}.$$

The first two relations are obvious, while the last two follow from the Cauchy estimates of Taylor coefficients applied to circles $|z| \leqq 1$ and $|z| \leqq r$, respectively.

The entropy $H_\epsilon(A_r)$ can be estimated as follows. For an $f \in A_r$ given by (11), we put $r_n(z) = \sum_{k=n+1}^\infty c_k z^k$, and begin by finding by means of (13a) an n for which $\|r_n\| \leqq \frac{1}{2}\epsilon$ for all $f \in A_r$. One gets $n \approx \log 1/\epsilon$. Now consider the Taylor coefficients $c_k = \alpha_k + i\beta_k$ of functions f for $k = 0, 1, \cdots, n$. Let α_k', β_k' denote the closest approximations from below to α_k, β_k by means of integral multiples of $\epsilon/2(n+1)$:

$$\text{(14)} \qquad \alpha_k' = \epsilon l_k/2(n+1), \qquad \beta_k' = \epsilon m_k/2(n+1), \qquad k = 0, \cdots, n.$$

Let U be the set of all $f \in A_r$ for which α_k', β_k' have some given set of values. It is obvious from (14) that the diameter of U does not exceed 2ϵ, and it is clear that the U cover A_r. The number N of the sets can be estimated from above if we note that by (13a), each of the integers l_k, m_k in (14) takes $\leqq$const. $r^{-k} n \epsilon^{-1}$ values. This allows us to show that $\log N$ does not exceed (10).

In an equally simple way we can estimate the capacity $C_{2\epsilon}(A_r)$ from below. One has merely to find constants a_k such that

$$\text{(15)} \qquad |\alpha_k| \leqq a_k, \qquad |\beta_k| \leqq a_k, \qquad k = 0, 1, \cdots$$

will imply $f(z) = \sum (\alpha_k + i\beta_k) z^k \in A_r$. For instance $a_k = C/(4kr^k)$ will do. Then we allow α_k, β_k to take all integral multiples of 3ϵ for some $\epsilon > 0$ which satisfy (15). The number of all possible selections is easily estimated, and (13) insures that the corresponding functions are 2ϵ-distinguishable. In this way (10) proves to be a lower bound for $C_{2\epsilon}(A_r)$.

More general sets A of analytic functions are the following. Let K be a bounded continuum in the plane, and G a bounded open set containing K. Let A consist of all functions analytic in G continuous, and bounded by a constant on $\overline{G}$, with the distance given by the uniform norm on K. Then [5, 13]

(16) $$H_\epsilon(A) \sim \tau(K, G) \log^2 \frac{1}{\epsilon} \cdot$$

The constant $\tau(K, G)$ can be determined in many cases (for instance by means of the "conformal radius" of the pair K, G). The main difference with the proof sketched above is that one uses other representations instead of the Taylor series: expansions in series of Tchebyshev polynomials (if K is a segment and G an ellipse), of Faber polynomials, etc. In the general case, expansions suggested by Erohin [4] can be used. If A consists of functions $f(z)$, $z = x + iy$ analytic in a strip $-l \leqq y \leqq l$, periodic with period 2π, then Fourier series expansions will be used. Very similar computations are possible for analytic functions of several variables. Let for example $A = A_{r_1, \cdots, r_s}(C)$ consist of functions $f(z_1, \cdots, z_s)$ analytic in the region $|z_1| < r_1, \cdots, |z_s| < r_s$, $r_i > 1$, $i = 1, \cdots, s$, and continuous and bounded in absolute value by C on its closure; then for the uniform norm on $|z_1| \leqq 1, \cdots, |z_s| \leqq 1$ one has [18]:

$$H_\epsilon(A) \sim \frac{2}{(s+1)! \log r_1 \cdots \log r_s} \log^{s+1} \frac{1}{\epsilon} \cdot$$

Also the entropies of sets of entire functions with some restrictions on their growth at infinity have been determined [13, 18].

3. Widths.

3.1. We shall now discuss another function associated with a set A, which characterizes its "massivity" in a different way, the n-dimensional width $d_n(A)$ of A. Also this notion is due to Kolmogorov [10], who computed for the first time the exact widths of some sets. Let us first agree on some notations. Let A be a bounded subset of a Banach space X, $G = \{g_n\}$ a sequence of elements of X. Then

(1) $$E_n^G(f) = \inf_{a_i} \left\| f - \sum_{i=1}^{n} a_i g_i \right\|,$$

where the infimum is taken for all selections of real scalars a_i, is the degree of approximation of f by G;

(2) $$E_n^G(A) = \sup_{f \in A} E_n^G(f)$$

is the degree of approximation of the set A by G; finally

(3) $$d_n(A) = \inf_G E_n^G(A),$$

is the optimal degree of approximation, the *n-th width of A*. The infimum in (3) is taken for all possible selections of sequences G in X. (Actually, only the first n elements $g_i \in G$ play a role.) The name width is justified geometrically, because in order to obtain $d_n(A)$, we take the maximal distance of A to an n-dimensional plane in X passing through the origin, and then vary the plane, and take the minimum of the distances, for all possible planes.

The width $d_n(A)$ depends not only on the set A, but also on the norm selected, or more exactly upon the space X; one writes $d_n^X(A)$ in order to emphasize this dependence.

Examples: if A is the closed unit ball in X, then one shows that $d_n(A)=1$ for all n, which are smaller than the dimension of X; and if A is compact, then $d_n(A)\to 0$ for $n\to\infty$.

It is the case of a compact set A which shall occupy us in the following. For a few sets A of functions, the widths are exactly known; in most cases one has only their estimates. It is the asymptotic behavior of $d_n(A)$ for $n\to\infty$ that shall be of primary interest to us. The inequality $d_n(A)\leqq E_n^G(A)$ is often useful as an estimate from above; in fact, the degree of approximation $E_n^G(A)$ is known in cases when the linear combinations $\sum_1^n a_i g_i$ are all polynomials, or all trigonometric polynomials, etc. And it is to be expected that this selection of G will make $E_n^G(A)$ its minimum, or at least close to it.

3.2. *Estimation of $d_n(A)$ from below.* In many simple cases, the following lemma, given by Kolmogorov and the author [**14, 16**] is useful. Let A be a compact subset of the space $C(B)$ of continuous functions which are defined on a compact infinite metric set B. Consider all numbers $\delta\geqq 0$ with the following property. There exist $n+1$ points of B, $x_0, x_1, \cdots, x_n$ such that for each distribution of signs, $\epsilon_k=\pm 1$, $k=0, 1, \cdots, n$, one can find a function $f\in A$ with

$$\operatorname{sign} f(x_k) = \epsilon_k, \qquad |f(x_k)| \geqq \delta, \qquad k = 0, \cdots, n. \tag{4}$$

Then one has:

LEMMA 1. *For each $A\subset C(B)$ and each δ of the described kind,*

$$d_n(A) \geqq \delta. \tag{5}$$

Thus the main problem in application of this lemma is to find the best distribution of points x_i in B.

The proof is simple. Let $G=\{g_n\}$ be arbitrary. The n linear equations $\sum_{k=0}^n c_k g_i(x_k)=0$, $i=1, \cdots, n$, with $n+1$ unknowns c_k have a nonzero solution, which we may assume to be normed by $\sum_{k=0}^n |c_k|=1$. Select a function of $f\in A$ satisfying (4) with $\epsilon_k=\operatorname{sign} c_k$. Then.

$$\begin{aligned}\left\| f - \sum_{i=1}^n a_i g_i \right\| &\geqq \left| \sum_{k=0}^n c_k \left[f(x_k) - \sum_{i=1}^n a_i g_i(x_k) \right] \right| \\ &= \left| \sum_{k=0}^n c_k f(x_k) \right| \geqq \sum |c_k| \delta = \delta,\end{aligned}$$

hence $E_n^G(f)\geqq\delta$, and (5) follows.

Also for other metrics similar lemmas exist. If A is a subset of the space $L^1(B)$ of integrable functions on B with respect to a measure μ, then one has:

LEMMA 2 [14]. *Assume that n, p are positive integers, $\delta \geqq 0$ and that $B = \bigcup_1^{n+p} B_k$ is a partition of B into disjoint measurable sets with the property that for each distribution of signs $\epsilon_k = \pm 1$, $k = 1, \cdots, n+p$, and for each selection of p sets B_{k_i}, $i = 1, \cdots, p$, there is a function $f \in A$ with*

$$\operatorname{sign} \int_{B_k} f d\mu = \epsilon_k, \; k = 1, \cdots, n+p; \qquad \left| \int_{B_{k_i}} f d\mu \right| \geqq \delta, \qquad i = 1, \cdots, p.$$

Then $d_n(A) \geqq p\delta$.

These and similar tools for the estimation of $d_n(A)$ from below, and classical results about approximation of functions by polynomials for the estimation of $d_n(A)$ from above, lead for example to the following estimates [14]:

THEOREM 4.

$$d_n(\Lambda(p + \alpha, s, C_0, \cdots, C_p, M)) \approx n^{-(p+\alpha)/s} \tag{6}$$

$$d_n(A) \approx \rho^{-n}, \qquad \rho > 1. \tag{7}$$

Here A is the set of all functions which are analytic inside the ellipse E_ρ, with foci ± 1 and the sum of half-axes equal to ρ, and bounded by 1 and continuous in the closed region; the metric in (6) or (7) is the C- or the L^1-metric (or any intermediate metric).

Thus (6) and (7) mean roughly that the known degrees of approximation by polynomials of the functions of classes Λ, A cannot be essentially improved if polynomials are replaced by other linear means of approximation.

3.3. *Cases where the widths can be exactly determined.* Several such cases have been found by Tihomirov [16]. To explain his method, let us start with the following remarks. From the definition of the width of a set A in a Banach space X, $d_n^X(A)$ it follows immediately that

$$d_n^X(A) \leqq d_n^{X'}(A) \qquad \text{if } X' \text{ is a subspace of } X. \tag{8}$$

It is more striking to know [16, p. 119] that inequality can occur here even if X and X' are finite dimensional. Tihomirov bases his method on the following

THEOREM 5 [6, p. 47]. *Let A be the closed unit ball of an $(n+1)$-dimensional subspace X' of a Banach space X. Then*

$$d_n^X(A) = 1.$$

Since $d_n^{X'}(A) = 1$, this means, that the nth width of A does not change when A is imbedded into a larger space X.

This theorem can be applied as follows. Let $\tilde{W}_p$, $p \geqq 1$ be the set of all 2π-periodic functions f with the property that $f^{(p)}(x)$ exists almost everywhere and satisfies $|f^{(p)}(x)| \leqq 1$, with $f^{(p-1)}$ being an indefinite integral of $f^{(p)}$. We consider $\tilde{W}_p$ in the uniform norm, as a subspace of the space $\tilde{C}$ of periodic continuous functions.

A classical result of Favard, and Akhiezer and Kreĭn (see [1]), gives exactly the degree of approximation of $\tilde{W}_p$ by trigonometrical polynomials of degree $n-1$:

$$E^{G}_{2n-1}(\tilde{W}_p) = K_p n^{-p} = \rho, \qquad K_p = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^{k(p+1)}}{(2k+1)^{p+1}}; \tag{9}$$

moreover, a function $f_{pn} \in \tilde{W}_p$ is known for which the polynomial of best approximation is zero, and $\|f_{pn}\| = \rho$. This f_{pn} is the pth periodic indefinite integral of sign sin nx (if p is odd) or of sign cos nx (if p is even). Tihomirov considers the $2n$-dimensional subspace X' of C which consists of all pth periodic integrals of the functions of the form $g(x) = c_k$, $x \epsilon \Delta_k$, $k = 0, 1, \cdots, 2n-1$, where the c_k are constants, and Δ_k are the intervals of constancy of sign sin nx (or of sign cos nx). Let U be the subset of X' corresponding to the values $|c_k| \leqq 1$, $k = 0, \cdots, 2n-1$. It is clear that $U \subset \tilde{W}_p$, on the other hand one can prove, using Rolle's theorem and the special form of f_{pn}, that if $f \in X'$ and $\|f\| \leqq \|f_{pn}\| = \rho$ then $f \in U$.

Consider the ball U_p of X' with the center in the origin and the radius ρ. What has been just said implies that $U_\rho \subset \tilde{W}_p$; on the other hand, by Theorem 5, $d_{2n-1}(U_\rho) = \rho$. Hence $d_{2n-1}(\tilde{W}_p) \geqq \rho$, and in view of (9) we obtain

$$d_{2n-1}(\tilde{W}_p) = K_p n^{-p}. \tag{10}$$

Tihomirov obtains further similar results, for example

$$d_{2n-1}(W_p) = K_p n^{-p} + \mathcal{O}(n^{-(p+1)}) \tag{11}$$

in the uniform norm; W_p consists of functions on $[0, 2\pi]$ with the same restrictions as for $\tilde{W}_p$, except that the condition of periodicity is dropped.

3.4. *A result of Kolmogorov.* We consider the subspace Y_p, $p \geqq 1$ of $L^2\ [0, 1]$, which consists of all functions f such that $f^{(p)}(x)$ exists almost everywhere and belongs to L^2, and that $f^{(p-1)}$ is an indefinite integral of $f^{(p)}$. Let V_p be the subset of Y_p with $\|f^{(p)}\| \leqq 1$, in the L^2-norm.

THEOREM 6 [10]. *Let* λ_n, $n = 1, 2, \cdots$ *denote all nonzero eigenvalues of the differential equation*

$$\begin{cases} y^{(2p)} = (-1)^p \lambda y \\ y^{(p)}(0) = y^{(p)}(1) = \cdots = y^{(2p-1)}(0) = y^{(2p-1)}(1) = 0. \end{cases} \tag{12}$$

All λ_n *are positive; if they are arranged in increasing order one has*

$$d_n(V_p) = +\infty \qquad \text{for } n = 0, 1, \cdots, p-1; \tag{13}$$

$$d_n(V_p) = \lambda_{n-p+1}^{-1/2} \qquad \text{for } n = p, p+1, \cdots. \tag{14}$$

To explain the proof, consider the equation

$$\tag{15} \begin{cases} z^{(2p)} = (-1)^p\lambda z, \\ z(0) = z(1) = \cdots = z^{(p-1)}(0) = z^{(p-1)}(1) = 0. \end{cases}$$

It is easy to see that for $\lambda \neq 0$ there is a one-to-one correspondence between the solutions of (12) and (15) given by $y^{(p)} = z$. For $\lambda = 0$, (15) has no solutions, while (12) has as solutions all polynomials of degree $p-1$.

Two sources of information are available for the study of equations (12) and (15). One is the general theory of eigenvalues of differential equations, the other the elementary formula, obtained by partial integration:

$$\tag{16} \int_0^1 z\overline{f^{(p)}}dx = \lambda \int y\bar{f}dx; \qquad z = y^{(p)}.$$

Here y is a solution of (12) corresponding to an eigenvalue λ, and $f \in Y_p$.

From the first source we obtain that the solutions ϕ_n, $n = 1, 2, \cdots$ of (15), corresponding to the eigenvalues λ_n, form an orthogonal system, which is complete in L^2. Then it follows from (16) very easily: all eigenvalues of (12) except $\lambda = 0$ are greater than zero; if ψ_n is the solution of (12) corresponding to λ_n and given by $\psi_n^{(p)} = \phi_n$, then also the ψ_n form an orthogonal system; each ψ_n is orthogonal to each polynomial P_{p-1} of degree $p-1$. We shall assume $\|\psi_n\| = 1$ and obtain then

$$\tag{17} \|\phi_n\| = \lambda_n^{1/2}.$$

It follows also that for each $f \in Y_p$ one has the expansions, convergent in the L^2-norm,

$$\tag{18} f = P_{p-1} + \sum_{n=1}^{\infty} a_n\psi_n, \qquad f^{(p)} = \sum_{n=1}^{\infty} a_n\phi_n,$$

where P_{p-1} is properly chosen. This, together with (17), implies that $f \in V_p$ if and only if $\sum_1^\infty a_n^2\lambda_n \leq 1$.

From this fact the theorem follows easily. On one hand, V_p contains the linear subspace X_p of the P_{p-1} of dimension p (hence (13)) and for each m contains all f in (18) with $\sum_1^m a_k^2 \leq \lambda_m^{-1}$. Thus for $n \geq p$, V_p contains an $n+1$ dimensional ball with the center in the origin and the radius $\lambda_{n-p+1}^{-1/2}$. By means of Theorem 5 we obtain the relation

$$d_n(V_p) \geq \lambda_{n-p+1}^{-1/2}, \qquad n \geq p.$$

On the other hand, the distance from an $f \in V_p$ given by (18) to the n-dimensional subspace of L^2 spanned by X_p and the first $n-p$ vectors ψ_k is equal to

$$\left\{\sum_{k=n-p+1}^{\infty} a_k^2\right\}^{1/2} \leq \lambda_{n-p+1}^{-1/2}\left\{\sum_1^{\infty} \lambda_k a_k^2\right\}^{1/2} \leq \lambda_{n-p+1}^{-1/2},$$

hence the theorem.

The reader will be interested to learn that all eigenvalues λ_n are simple and that

$$\lambda_n = (\pi n)^{2p} + O(n^{2p-1}).$$

If he would ask how one could arrive at the equation (12) before knowing the solution of our problem, the following answer could be offered. In finding the widths of V_p, we try to find more and more functions $y_1, \cdots, y_n$ in V_p, orthogonal to each other. This means that $y=y_n$ is the solution of the variational problem $\int_0^1 y^2 dx = \max.$ with the restraint $\int_0^1 y^{(p)2} dx = 1$. The Euler equation of this problem is exactly (12).

3.5. *Relations between widths and entropy. Nonlinear approximation.* Both the entropy $H_\epsilon(A)$ and the widths $d_n(A)$ of a compact set A measure its "size." The widths measure the amount of deviation of A from linear subspaces, while the entropy depends much less upon the shape of A. Thus it is not astonishing that inequalities connecting these two functions exist, but are not very conclusive. Mitiagin ([**15**]; see also [**13**]) proves the following:

$$\int_0^{1/2\epsilon} \frac{l(t)}{t}\, dt \leqq H_\epsilon(A) \leqq \left[m\left(\frac{2}{\epsilon}\right) + 1\right] \log \frac{8(d_0+\epsilon)}{\epsilon}\,. \tag{19}$$

Here A is a compact convex subset of a Banach space, symmetric with respect to the origin, $d_0 = d_0(A) = \sup_{x\in A} \|x\|$, and the functions $m(t)$, $l(t)$ for $t>0$ are defined as the largest integers m or l satisfying $d_m \geqq t^{-1}$ or respectively $l^{-1}d_{l-1} \geqq t^{-1}$. The left hand side of (19) can be obtained in the following way. From the definition of widths and by induction one can find in A points x_n such that the distance from x_n to the linear space spanned by $x_0, \cdots, x_{n-1}$ is not less than $d_n(A) = d_n$. This implies that A contains for each n the octahedron Ω_n formed by the points $\sum_0^n \xi_k x_k$ with $\sum |\xi_k| \leqq 1$. If we consider the points of the form $y = \sum_0^n \eta_k y_k$ with $y_k = \epsilon x_k / d_k$ and integral η_k belonging to Ω_n, then it is easy to see that different y's are at a distance $\geqq \epsilon$ from each other and to estimate their number in Ω_n. This leads to an estimate from below of $M_{2\epsilon}(A)$ and hence of $H_\epsilon(A)$ (see 2.1 and Theorem 1).

We must make at least a passing mention of the nonlinear approximation. The notion of width is based upon the approximation of a function $f(x)$ by expressions $P(x) = a_1\phi_1(x) + \cdots + a_n\phi_n(x)$, linear in the parameters $a_1, \cdots, a_n$. In the book [18], Vituškin presents a theory which gives estimates from below of the degree of approximation of f by much more general algorithms P. For example, P may be the quotient of two polynomials in $a_1, \cdots, a_n$ with coefficients which are functions of x; also the operations max and min are allowed. This theory uses essentially new means (multidimensional variations of sets, topological properties of level surfaces of polynomials in Euclidean spaces) and cannot be reviewed here.

4. Representation of functions by sums and superpositions of functions of fewer variables.

4.1. Superpositions are functions of functions. For example,

$$F(x, y, z) = f(g(x, y), h(x, k(y, z))) \tag{1}$$

is a superposition of functions of two variables. Also the sum

$$x_1 + x_2 + \cdots + x_n = x_1 + (x_2 + \cdots + (x_{n-1} + x_n) \cdots)$$

is a superposition of functions of two variables. Studying the roots of equations of the seventh degree, Hilbert conjectured that not all continuous functions of three variables are superpositions of continuous functions of two variables. This is the 13th problem of Hilbert of his famous address at the International Congress of Mathematicians in Paris 1900. Recently Kolmogorov [12] and Arnol'd [2] disproved Hilbert's conjecture. The remarkable result of Kolmogorov is that each continuous function of s variables can be represented by sums and superpositions of functions of one variable! For example, if $s=2$, Kolmogorov's result can be formulated as follows:

THEOREM 7. *There exist ten continuous monotone increasing functions $\phi_q(x)$, $\psi_q(x)$, $0 \leqq x \leqq 1$, $q=1, \cdots, 5$ with values in the interval $[0, 1]$ which have the property that each continuous function $f(x, y)$ on the two-dimensional unit square B is representable in the form*

$$f(x, y) = \sum_{q=1}^{5} g(\phi_q(x) + \psi_q(y)), \tag{2}$$

where $g(u)$ is some continuous function defined for $0 \leqq u \leqq 2$.

Thus, the functions ϕ_q, ψ_q are independent of f, while g depends upon it. A similar statement holds for function of s variables, which are representable in the form

$$f(x_1, \cdots, x_s) = \sum_{q=1}^{2s+1} g\left(\sum_{p=1}^{s} \phi_{pq}(x_p)\right).^* \tag{3}$$

We restrict ourselves to $s=2$, because the general case is treated in exactly the same way.

The idea of the proof, roughly speaking, is to construct a pair of functions $\phi(x), \psi(y)$ which have the property that on disjoint subsets B', B'' of B, the sum $\phi(x)+\psi(y)$ takes values belonging to two disjoint subsets of the real line. Of course, this cannot be achieved for all pairs B', B'' of subsets of B, but we want it to be true for "sufficiently many" pairs.

Kolmogorov begins by constructing two increasing continuous functions $\phi_1(x), \psi_1(x)$ defined on $[0, 1]$ and with values in this interval with the following

* Actually, Kolmogorov [12] writes $2s+1$ functions g_q instead of g in (3), five functions g_q in (2). Only apparently weaker, this is in fact equivalent to our formulation.

property. For each $k=1, 2, \cdots$ there are finitely many closed subsets B_{1i}^k, $i=1, 2, \cdots, m_k$ of B and corresponding closed disjoint intervals I_{1i}^k of the real line such that if a point (x, y) belongs to B_{1i}^k, then $\phi_1(x)+\psi_1(y)$ belongs to I_{1i}^k. The diameter of B_{1i}^k tends to zero for $k\to\infty$ (uniformly in i; the sets B_{1i}^k are actually squares). For the proof this is not enough: the sets B_{1i}^k are disjoint for each fixed k, and therefore cannot cover B completely.

Therefore Kolmogorov repeats this construction five times and obtains ten increasing continuous functions $\phi_q(x)$, $\psi_q(x)$, $q=1, \cdots, 5$ and corresponding closed subsets $B_{qi}^k\subset B$ and intervals $I_{qi}^k\subset[0, 2]$ with $k=1, 2, \cdots; q=1, \cdots, 5$; $i=1, \cdots; m_k$ such that $\phi_q(x)+\psi_q(y)\in I_{qi}^k$ if $(x, y)\in B_{qi}^k$. All intervals I_{qi}^k are disjoint for a fixed k; therefore the B_{qi}^k are disjoint for each fixed q and k. It is very essential that the B_{qi}^k cover B very well; precisely, for each k, each point (x, y) in B is covered at least three times by the B_{qi}^k (there are at least "three hits out of five shots"). The following lemma concerning the functions ϕ_q, ψ_q is useful:

LEMMA 3. *Let C be the space of continuous functions on B, and let C' consist of all functions of the form*

$$h(x, y) = \sum_{q=1}^{5} g(\phi_q(x) + \psi_q(y)) \tag{4}$$

with an arbitrary continuous function g. Assume that there is a constant $0<\lambda<1$ such that for each $f\in C$ there is an $h\in C'$ given by (4) with $\|f-h\|\leqq\lambda\|f\|$; $\|g\|\leqq\|f\|$. Then $C'=C$.

The proof is easy: one first selects an h_1 with the corresponding g_1 so that $\|f-h_1\|\leqq\lambda\|f\|$, $\|g_1\|\leqq\|f\|$, then an h_2 with g_2 so that $\|(f-h_1)-h_2\|\leqq\lambda\|f-h_1\|\leqq\lambda^2\|f\|$, $\|g_2\|\leqq\lambda\|f\|$, and so on. Then the series $g=\sum_{n=1}^{\infty} g_n$ converges uniformly, and with h connected with this g by (4) one has $f=h=\sum_1^{\infty} h_n$.

To complete the proof of the theorem, we take $\frac{2}{3}<\lambda<1$. We can fix k so large that the oscillation of the given function f is not more than $\epsilon\|f\|$ on each B_{qi}^k, and that $\epsilon>0$ is so small that $\frac{2}{3}+\epsilon<\lambda$. Let z_{qi}^k denote some point of B_{qi}^k. We determine $g(u)$, $0\leqq u\leqq 2$ in the following way. On each I_{qi}^k we let g be a constant: $g(u)=\frac{1}{3}f(z_{qi}^k)$. Onto the rest of $[0, 2]$ we extend g in an arbitrary fashion, but so that $g(u)$ remains continuous and that $|g(u)|\leqq\frac{1}{3}\|f\|$. Now consider the sum (4) for some $(x, y)\in B$. Three of its terms $g(\phi_q(x)+\psi_q(y))$ have the property that $(x, y)\in B_{qi}^k$ for some i; each of these terms is equal to $\frac{1}{3}f(x, y)$ with an error not exceeding $\frac{1}{3}\epsilon\|f\|$. The remaining two terms are each $\leqq\frac{1}{3}\|f\|$. Hence

$$|f(x, y) - h(x, y)| \leqq \tfrac{2}{3}\|f\| + 3\cdot\tfrac{1}{3}\epsilon\|f\| < \lambda\|f\|,$$

and the proof is completed by applying Lemma 3.

The functions ϕ_q, ψ_q in Kolmogorov's construction seem not to be as wild as one could expect a priori. For example, they prove to belong to the class Lip α for $\alpha=\frac{1}{5}$. Theorem 7 is very important in principle. Will it have useful applications? Theoretically, one could hope to derive by means of it results concerning functions of several variables from corresponding results about functions of one variable. For example, from the Weierstrass theorem about the approximation of functions of one variable by polynomials at once follows the corresponding theorem for functions of s variables. This is not very astonishing, since most proofs of the theorem of Weierstrass generalize immediately to higher dimensions. One wonders whether Kolmogorov's theorem can be used to obtain positive results of greater depth.

4.2. Let us return to Hilbert's conjecture. Despite its disproof, it originated from a sound idea: that "bad" functions in general cannot be represented by superpositions of "good" functions. From this point of view, Theorem 7 makes it clear, that the characteristic of badness $\chi=s$ of $f(x_1, \cdots, x_s)$, equal to the number of variables of f, has failed. The reason is that general continuous functions are so bad, that the number of variables is not particularly important. Are there other characteristics of this kind which will work? That this is the case has been found by Vituškin [**17**] and Kolmogorov [**13**, p. 80]. Let a function of s variables $f(x_1, \cdots, x_s)$ on the s-dimensional unit square have all partial derivatives of orders not exceeding p, with the derivatives of order p satisfying a Lipschitz condition with exponent $\alpha(p=0, 1, \cdots, 0<\alpha\leqq 1)$. We shall then say that f belongs to the characteristic $\chi=s/(p+\alpha)$.

THEOREM 8. *Not all continuous functions with a given characteristic χ_0 can be represented by superpositions of functions of characteristics $\chi=s/(p+\alpha)<\chi_0$ and with $p+\alpha\geqq 1$.*

In order to sketch the proof, we first introduce the notion of a *type* T of superpositions. A type T is a set of functions F given by a formula such as (1) and by an indication of classes of type $\Lambda(p+\alpha, s, C_0, \cdots, C_p, M)$ (see 2.3) to which the functions $f, g, \cdots$ contained in the formula are allowed to belong. We shall assume that all $C_0, \cdots, C_p, M$ are positive integers and that the classes have characteristics $\chi=s/(p+\alpha)<\chi_0$. Then T consists of all functions F which can be obtained by varying the functions $f, g, \cdots$, but leaving the structure of the formula and the given classes fixed. One sees immediately that there are only countably many different types T.

The entropy (in the uniform norm) of a class $\Lambda(p+\alpha, s, C_0, \cdots, C_p, M)$ is of the order $\epsilon^{-\chi}$, $\chi=s/(p+\alpha)$. What is the entropy of a type T? Consider the superpositions

$$F(x_1, \cdots, x_s) = f(g_1(x_1, \cdots, x_s), \cdots, g_r(x_1, \cdots, x_s)), \tag{5}$$

where the f satisfy the Lipschitz condition

$$\left| f(y_1, \cdots, y_r) - f(y_1', \cdots, y_r') \right| \leqq M \max_{i=1,\ldots,r} \left| y_i - y_i' \right|,$$

and let $f, g_1, \cdots, g_r$ run through sets $A, B_1, \cdots, B_r$ with known entropies (with respect to the uniform norm). We shall show that the entropy of the set T of the F's satisfies

$$H_{(M+1)\epsilon}(T) \leqq H_\epsilon(A) + \sum_1^r H_\epsilon(B_i). \tag{6}$$

Let the functions $\bar{f}$, $\bar{g}_i$ run through ϵ-nets in A, B_i respectively, and let N_ϵ, N_ϵ^i be the numbers of elements in these sets. Consider the functions

$$\overline{F} = \bar{f}(\bar{g}_1(x_1, \cdots, x_s), \cdots, \bar{g}_r(x_1, \cdots, x_s)), \tag{7}$$

$N_\epsilon \prod_{i=1}^r N_\epsilon^i$ in number. For each function (5) there exist $\bar{f}$, $\bar{g}_i$ with

$$\|f - \bar{f}\| < \epsilon \qquad \|g_i - \bar{g}_i\| < \epsilon, \qquad i = 1, \cdots, r,$$

and then

$$\begin{aligned} |F - \overline{F}| &\leqq |f(g_1, \cdots, g_r) - f(g_1, \cdots, g_r)| + |f(g_1, \cdots, g_r) - \bar{f}(\bar{g}_1, \cdots, \bar{g}_r)| \\ &\leqq \epsilon + M\epsilon = (M + 1)\epsilon. \end{aligned}$$

Hence the $\overline{F}$ form an $(M+1)\epsilon$-net in T, so that $N_{(M+1)\epsilon}(T) \leqq N_\epsilon \prod_{i=1}^r N_\epsilon^i$. Relation (6) follows from this by taking logarithms. In particular, if A, B_i are classes Λ with characteristics $\chi = s/(p+\alpha) \geqq m$, then

$$H_\epsilon(T) \leqq \text{const. } \epsilon^{-m}. \tag{8}$$

The formula defining a type T may have a more complicated structure than (5), but it reduces to this case by finite iteration, and by induction we obtain: the entropy of each type T satisfies (8), where $m = \max \chi < \chi_0$.

The proof of the Theorem 8 is completed by a category argument. Let $\chi_0 = s_0/(p_0 + \alpha_0)$; we consider the set X of all functions f on the s_0-dimensional unit square with the smoothness $p_0 + \alpha_0$(2.3). A norm can be defined for $f \epsilon X$ in a natural way: $\|f\|_X$ is the maximum of the absolute values of all partial derivatives of f of orders $\leqq p$ and of the smallest constants with which the derivatives of order p satisfy the Lipschitz α_0-condition. In a standard way one proves that X is a Banach space; its completeness in particular is essential. A ball $U_r(0)$ in X of center 0 and radius r is a class $\Lambda(p_0 + \alpha_0, s_0, r, \cdots, r, r)$ and has (2.3) the entropy $\approx \epsilon^{-\chi_0}$. Any other ball in X is a translation of a ball $U_r(0)$ and has the same entropy. From this and (8) it is clear: a type T cannot contain any ball of X.

On the other hand, the functions $f, g, \cdots$ in a formula (1) determining a class T, as well as their derivatives are equicontinuous. This allows us to apply Arzela's theorem and prove that T is closed under uniform convergence, and hence under convergence in the norm of the space X. Thus, sets $T \cap X$ are nowhere dense in X. Since there are countably many types T, the union $\bigcup_T (T \cap X)$ is a set of the first category in X. By Baire's theorem, its complement in X is a nonempty set. It consists of functions not representable by superpositions of functions with characteristics $\chi < \chi_0$.

We have proved the theorem by establishing not only that there are functions in X not representable by superpositions of our type, but even that "almost all" functions (in the category sense) in X are not so representable. A fine point of the proof will be noted: it uses two norms; the uniform norm to compute the entropies, and the norm $\|f\|_X$, in which the space X is complete.

References

1. N. I. Akhiezer, Theory of Approximation, F. Ungar, New York, 1956.

2. V. I. Arnol'd, On functions of three variables, Dokl. Akad. Nauk SSSR, **114** (1957) 679–681.

3. Yu. A. Brudnyĭ and A. F. Timan, Constructive characteristics of compact sets in Banach spaces and ϵ-entropy, Dokl. Akad. Nauk, **126** (1959) 927–930.

4. V. D. Erohin, On conformal maps of rings and on the fundamental basis of the space of functions analytic in an elementary neighborhood of an arbitrary continuum, Dokl. Akad. Nauk, **120** (1958) 689–692.

5. ———, On the asymptotic of the ϵ-entropy of analytic functions, Dokl. Akad. Nauk, **120** (1958) 949–952.

6. I. C. Gohberg and M. G. Kreĭn, Fundamental aspects of defect numbers, root numbers and indexes of linear operators, Uspehi Mat. Nauk, **12** (1957) no. 2 (74) 43–118.

7. P. R. Halmos, Entropy in ergodic theory, Mimeographed notes, University of Chicago, 1959.

8. F. Hausdorff, Set theory, Chelsea, New York, 1957.

9. D. Jackson, The theory of approximation, Amer. Math. Soc. Colloquium Publ., **11**, New York, 1930.

10. A. N. Kolmogorov, Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse, Ann. of Math. (2) **37** (1936) 107–111.

11. ———, Asymptotic characteristics of some completely bounded metric spaces, Dokl. Akad. Nauk, **108** (1956) 585–589.

12. ———, On the representation of continuous functions of many variables by superpositions of continuous functions of one variable and addition, Dokl. Akad. Nauk, **114** (1957) 679–681.

13. A. N. Kolmogorov and V. M. Tihomirov, ϵ-entropy and ϵ-capacity of sets in function spaces, Uspehi Mat. Nauk, **14** (1959) no. 2 (86) 3–86.

14. G. G. Lorentz, Lower bounds for the degree of approximation, Trans. Amer. Math. Soc., **97** (1960) 25–34.

15. B. S. Mityagin, The approximative dimension and bases in Kernel spaces, Uspehi Mat. Nauk, **16** (1961), No. 4 (100) 63–132.

16. V. M. Tihomirov, The widths of sets in functional spaces and the theory of best approximations, Uspehi Mat. Nauk, **15** (1960) no. 3 (93) 81–120.

17. A. G. Vituškin, On the 13th problem of Hilbert, Dokl. Akad. Nauk, **95** (1954) 701–704.

18. ———, Estimation of the complexity of the tabulation problems, Moscow, 1959. There is an English translation: Theory of the transmission and processing of information (!), Pergamon Press, New York, 1961.