

LOWER BOUNDS FOR APPROXIMATION BY NONLINEAR MANIFOLDS

BY
HUGH E. WARREN

1. Introduction. If L is a normed linear space and K and F are subsets of L , it is a general problem of approximation theory to compute the number

$$(1) \quad D(K, F) = \sup_{x \in K} \inf_{y \in F} \|x - y\|,$$

which is the precision to which F approximates K . An instance of classical interest is that in which L is $C([0, 1])$, K is Λ^ω —i.e. those functions bounded by one, and whose modulus of continuity is dominated by the modulus of continuity function ω —and F is the family P_n of polynomials of degree $n - 1$. For this case it is known that there are positive constants A and B , independent of ω and n , such that

$$(2) \quad A\omega(1/n) \leq D(\Lambda^\omega, P_n) \leq B\omega(1/n),$$

that is, $D(\Lambda^\omega, P_n) = O(\omega(1/n))$. (See for example [4].) The expression of D in terms of upper and lower bounds as in (2) is typical of known results.

For arbitrary compact sets K in $C([0, 1])$, the theorem of Weierstrass shows that $D(K, P_n) \rightarrow 0$ as $n \rightarrow \infty$. With a view to abstracting this result, one can replace each P_n by an arbitrary linear subspace G_n of dimension n , and can study the behavior of $D(K, G_n)$ as $n \rightarrow \infty$. In this connection, A. N. Kolmogorov [1] defined the n -width of any subset K of a normed linear space L to be

$$(3) \quad D_n(K) = \inf_{G_n} D(K, G_n),$$

where the infimum is over all n -dimensional subspaces G_n of L . In the space $C([0, 1])$, with Λ^ω as before, it turns out that $D_n(\Lambda^\omega) = O(\omega(1/n))$ as $n \rightarrow \infty$. Thus in the limit as $n \rightarrow \infty$, the classical subspaces P_n of polynomials approximate Λ^ω as well as any sequence $\{G_n\}$ of linear subspaces.

A. G. Vitushkin has studied $D(K, F)$ for families of sets F more general than linear manifolds [6]. Specifically, he takes L to be $C(X)$, the real valued continuous functions, with supremum norm, on a compact metric space X . The set K is some compact subset of $C(X)$. Vitushkin chooses F to be of the form

$$(4) \quad F = \{P(x) : x = (x_1, \dots, x_n) \in \mathbb{R}^n\}$$

where $P(x)$ is an algebraic polynomial in the coordinates x_1, \dots, x_n of x with coefficients in $C(X)$. If in (4) the polynomial degree of P is d , F will be said to

depend polynomially of degree d on n parameters. The n parameters are the coordinates x_1, \dots, x_n .

Now for a general normed linear space L and a compact subset K in L , define

$$(5) \quad D_{n,d}(K) = \inf_F D(K, F),$$

where the infimum is over all subsets F in L which depend polynomially of degree d on n parameters. In the case $L = C([0, 1])$, Vitushkin's results state that

$$(6) \quad D_{n,d}(\Lambda^\omega) = O\left(\omega\left(\frac{1}{n \log d}\right)\right)$$

as n and d tend to infinity. (See §36 in [6].)

To show in (6) that for some constant A

$$(7) \quad D_{n,d}(\Lambda^\omega) \leq A\omega\left(\frac{1}{n \log d}\right)$$

is relatively easy (see §§36 and 29 in [6]). The proof of the corresponding lower bound for $D_{n,d}$ is intricate and is based on the theory of "variations of sets" as expounded by Vitushkin in [5] and [6]. The present paper gives lower bounds for $D_{n,d}$ in general real linear spaces L . Although Vitushkin's results will be included as special cases, the method of proof is different. In particular, no use is made of "variations of sets".

The main theorems on $D_{n,d}$ (see §4) will be stated in terms of linear functionals on L . It will be assumed from now on that L is a linear space over the real numbers. If $\lambda_1, \dots, \lambda_m$ are linear functionals on L , the mapping $T: L \rightarrow \mathbb{R}^m$ given by $T(f) = (\lambda_1(f), \dots, \lambda_m(f))$ is also linear. If as in (4) $F = \{P(x) : x \in \mathbb{R}^n\}$ is a subset of L which depends polynomially of degree d on n parameters, then $p_i(x) = \lambda_i(P(x))$ is an ordinary real polynomial of degree d or less in n variables. Moreover,

$$(8) \quad T(F) = \{p(x) = (p_1(x), \dots, p_m(x)) : x \in \mathbb{R}^n\}.$$

If the transformation T has norm ≤ 1 , then for any compact K in L , $D(K, F) \geq D(T(K), T(F))$. Using a technique of H. S. Shapiro [3], one can find a lower bound for $D(T(K), T(F))$ by determining an upper bound for the number of sign sequences $s = (s_1, \dots, s_m)$ which have the form $\text{sgn } p(x) = (\text{sgn } p_1(x), \dots, \text{sgn } p_m(x))$.

The number of sign sequences $\text{sgn } p(x)$ is not greater than the number of topological components of $\mathbb{R}^n - \bigcup_{i=1}^m \{x : p_i(x) = 0\}$. In §§2 and 3 following, a suitable bound for the number of these components is given in terms of m , n , and d .

2. Partitions of a topological manifold by submanifolds.

THEOREM 1. *Let M be a connected topological n -manifold, and let M_1, \dots, M_b be connected $(n-1)$ -manifolds embedded in M so that:*

- (1) *the M_i are topologically closed and locally flat in M ;*

(2) the intersection of any given j of the M_i , $1 \leq j \leq n$, is either empty or is an $(n-j)$ -manifold that has finitely many components and is locally flat in the intersection of any $j-1$ of the given M_i ;

(3) any intersection of more than n of the M_i is empty.

Let b_j be the number of topological components among all intersections of any j of the M_i with M .

Under these hypotheses the set $M - \bigcup_{i=1}^m M_i$ has at most $\sum_{j=0}^n b_j$ topological components.

I preface the proof of this with two lemmas.

LEMMA 1.1. *If M is a connected n -manifold and M^* is a connected locally flat submanifold of M of dimension $n-1$, then $M - M^*$ has at most two components.*

Proof. One can assume M^* is not empty, since otherwise the conclusion is trivially true. If C_i is a component of $M - M^*$, put $E_i = M^* \cap \text{Cl}(C_i)$, where "Cl" denotes closure. Since M^* is locally flat, each point x of M^* has a neighborhood N in M for which there is a homeomorphism

$$(9) \quad h: (M \cap N, M^* \cap N, x) \rightarrow (R^n, R^{n-1}, 0).$$

Give R^n rectilinear coordinates x_1, \dots, x_n so that $R^{n-1} = \{(x_1, \dots, x_n) : x_n = 0\}$. Put $R_+^n = \{(x_1, \dots, x_n) : x_n > 0\}$, and $R_-^n = \{(x_1, \dots, x_n) : x_n < 0\}$. Because h is a homeomorphism $(M - M^*) \cap N$ has exactly two components—namely $h^{-1}(R_+^n)$ and $h^{-1}(R_-^n)$. Since $M^* \cap N = h^{-1}(R^{n-1} \cap \text{Cl}(R_+^n)) = h^{-1}(R^{n-1} \cap \text{Cl}(R_-^n))$, both components of $(M - M^*) \cap N$ contain $M^* \cap N$ in their closure. Because $M^* \cap N$ is open in M^* and because each set $E_i = M^* \cap \text{Cl}(C_i)$ is a union of sets of the type $M^* \cap N$, each E_i is open in the relative topology of M^* . The sets E_i are automatically closed in M^* since $\text{Cl}(C_i)$ is closed in M . Being both open and closed in the connected set M^* , each E_i must be empty or all of M^* . If E_i is empty, C_i is both open and closed in the connected set M . Hence $E_i = \emptyset$ implies $C_i = M$. But this is a contradiction because C_i lies in $M - M^*$ which, since $M^* \neq \emptyset$, is a proper subset of M . Consequently every E_i coincides with M^* .

Thus each point x of M^* is a limit point of every component of $M - M^*$. But x is a limit point of only those components which contain one of the two sets $h^{-1}(R_+^n)$, $h^{-1}(R_-^n)$. Hence $M - M^*$ has at most two components.

DEFINITION. For $b \geq 0$ let A_n denote the collection of $(b+1)$ -tuples of manifolds $(M; M_1, \dots, M_b)$, where M is an n -manifold and the M_i are $(n-1)$ -manifolds, which together satisfy the hypotheses of Theorem 1.

LEMMA 1.2. *If $(M; M_1, \dots, M_b) \in A_n$, and if k is the number of components of $M - \bigcup_{i=1}^b M_i$, k' is the number of components of $M - \bigcup_{i=1}^{b-1} M_i$, and k'' is the number of components of $M_b - \bigcup_{i=1}^{b-1} M_i$, then $k \leq k' + k''$.*

Proof. Let $\{C_j\}$ be the components of $M - \bigcup_{i=1}^{b-1} M_i$, and let k_j be the number of components of $M_b \cap C_j$. Since the M_i are closed, each C_j is an n -manifold and each

component of $M_b \cap C_j$ is a connected $(n-1)$ -manifold which is locally flat in C_j because M_b is locally flat in M . The components of $M_b \cap C_j$ are by definition disjoint. Removing them one at a time and applying Lemma 1.1 at each step shows that M_b divides C_j into at most $k_j + 1$ components.

Since $M_b - \bigcup_{i=1}^{b-1} M_i = \bigcup_j (M_b \cap C_j)$ and since the C_j are disjoint, $k'' = \sum_j k_j$. Hence $k \leq \sum_j (1 + k_j) = k' + k''$.

Proof of Theorem 1. Finding an upper bound for the number of components of $M - \bigcup_{i=1}^b M_i$ will be referred to as the "problem of $(M; M_1, \dots, M_b)$."

Let $b_{i,j}$ be the number of components among all intersections of M_i with any j of M_{i+1}, \dots, M_b . Let the components among all intersections $M_i \cap M_{i+1}, \dots, M_i \cap M_b$ be denoted by $M(i, 1), \dots, M(i, b_{i,1})$.

Lemma 1.2 reduces the problem of $(M; M_1, \dots, M_b)$ to the problem of $(M; M_1, \dots, M_{b-1})$ and the problem of $(M_b; M(b, 1), \dots, M(b, b_{b,1}))$. Successive application of Lemma 1.2 to the problems of $(M; M_1, \dots, M_i)$, $i = b, \dots, 1$, reduces the problem of $(M; M_1, \dots, M_b)$ to the problem of $(M; \emptyset)$ and the problems of $(M_i; M(i, 1), \dots, M(i, b_{i,1}))$, $i = b, \dots, 2$, and $(M_1; \emptyset)$. In particular, if β_i is a bound for the number of components of $M_i - \bigcup_{j=1}^{b_{i,1}} M(i, j)$, then the number of components of $M - \bigcup_{i=1}^b M_i$ is at most $1 + \sum_{i=1}^b \beta_i$.

It is easy to check that $(M_i; M(i, 1), \dots, M(i, b_{i,1})) \in A_{n-1}$, $i = 1, \dots, b$. As an induction hypothesis suppose that the theorem is true for dimensions less than n . Then in particular the theorem is supposed true for dimension $n-1$; therefore the number of components of $M_i - \bigcup_{j=1}^{b_{i,1}} M(i, j)$ is at most $\sum_{j=0}^{b_{i,1}-1} b_{i,j}$, $i = 1, \dots, b$. Since $b_{i+1} = \sum_{j=1}^b b_{i,j}$, it follows that the number of components of $M - \bigcup_{i=1}^b M_i$ is at most

$$(10) \quad 1 + \sum_{i=1}^b \sum_{j=0}^{b_{i,1}-1} b_{i,j} = \sum_{j=0}^b b_j.$$

Since the theorem is self-evident in the case $n=1$, its truth for all dimensions follows by complete induction. In fact when $n=1$, M is topologically either a Euclidean line or a circle and M_1, \dots, M_b are b distinct points. These b points divide a Euclidean line into $b+1 = b_1 + b_0$ components; they divide a circle into $b < b+1 = b_1 + b_0$.

3. Partitions of R^n by real varieties. In this section p, q, p_i , and q_i will stand for polynomials in n variables with *real* coefficients. The set of points $x = (x_1, \dots, x_n)$ in R^n at which p vanishes will be denoted by $N(p)$. The space R^n will be thought of as embedded in complex Euclidean n -space C^n . The set of $z = (z_1, \dots, z_n)$ in C^n at which p vanishes will be denoted by $N_C(p)$. By adding an $(n+1)$ -st variable, one can make p into a homogeneous polynomial on projective n -space in the usual way. Let $P(p)$ and $P_C(p)$ denote the zeros of this homogeneous version of p in real and complex projective n -space respectively.

An intersection $\bigcap_{i=1}^m N(p_i)$ will be called *regular* if at each point z of $\bigcap_{i=1}^m N_C(p_i)$ the gradients of p_1, \dots, p_m at z are linearly independent over the complex numbers.

If the gradients of p_1, \dots, p_m are linearly dependent at z , the point z will be referred to as a *singular point* of the intersection. Note that a single $N(p)$ is regular when $N_c(p)$ has no singular points in the usual algebraic sense. Also if $m > n$, an intersection of sets $N(p)$ is regular if and only if the corresponding complex intersection is empty.

A number of sets $N(p_1), \dots, N(p_m)$ are said to form a *regular configuration* if all intersections $\bigcap_{k=1}^j N(p_{i(k)}), j=1, \dots, m$ are regular.

LEMMA 2.1. *If M_1, \dots, M_b are the topological components of the several $N(q_1), \dots, N(q_m)$ which form a regular configuration, then the $(b+1)$ -tuple $(R^n; M_1, \dots, M_b)$ satisfies the hypotheses of Theorem 1.*

Proof. The connectedness conditions are obviously satisfied. For $j=1, \dots, n$, select any j of the $N(q_i)$ in some arbitrary order and let E_k be the intersection of the first $k, k=1, \dots, j$. If x is some point of E_j , by the regularity conditions on the gradients at x of the q_i involved, one can apply the implicit function theorem to produce a neighborhood U of x and a homeomorphism

$$(11) \quad h: (U, U \cap E_1, \dots, U \cap E_j, x) \rightarrow (R^n, R^{n-1}, \dots, R^{n-j}, 0).$$

The convention in (11) is that R^k corresponds to the points (x_1, \dots, x_n) in R^n such that $x_{k+1} = \dots = x_n = 0$, and $0 = (0, \dots, 0)$. This gives the manifold properties and local flatness. The condition that the intersection of any m of the M_i be empty for $m > n$ is plain from the definition of regularity above and the remarks following. The finiteness condition on the number of components has been proved by H. Whitney for arbitrary real varieties, see Theorem 3 in [8].

LEMMA 2.2. *Let p be a real polynomial. If $V = \bigcap_{i=1}^m N(q_i)$ is regular or if $V = R^n$, then for all but finitely many real numbers α , $V(\alpha) = N(p-\alpha) \cap V$ is regular.*

Proof. Put $V_c = \bigcap_{i=1}^m N_c(q_i)$ and $V_p = \bigcap_{i=1}^m P_c(q_i)$. If $m > n$, V_c is empty and there is nothing more to prove. If $m = n$, Lemma 2.1 shows that V_c consists of finitely many points. If α avoids the finite set of values of p at these points, $N_c(p-\alpha) \cap V_c$ is empty and $V(\alpha)$ is trivially regular. In the case $m < n$, or in the case $V = R^n$, put $V_p(\alpha) = P_c(p-\alpha) \cap V_p$. Consider the pencil of varieties $|V_p(\alpha)|$, where now α ranges over the complex projective line. Call α *singular* if $V_p(\alpha)$ has a singular point in the usual algebraic sense which is neither in the singular subvariety of V_p nor in $\bigcap_{\alpha} V_p(\alpha)$. By the strong theorem of Bertini (see for example [7]) the singular α form a proper subvariety of the projective line and hence are finite in number.

Since V is regular, the singular subvariety of V_p does not intersect V_c . Also $\bigcap_{\alpha} V_p(\alpha)$ lies at infinity, outside C^n , and in particular does not intersect V_c . It follows that for all but the finitely many singular α , $V_c(\alpha)$ contains no singular points and hence that $V(\alpha)$ is regular.

LEMMA 2.3. Let q_1, \dots, q_m be given, and put $S = \bigcap_{i=1}^m N(q_i)$. Given any p , there is a $\beta > 0$, such that whenever $0 < \alpha < \beta$, every topological component of the set $E = \mathbb{R}^n - S \cup N(p)$ contains one of the sets $E_\alpha = \mathbb{R}^n - S \cup N(p - \alpha) \cup N(p + \alpha)$.

Proof. The components C_i of E are finite in number, again by H. Whitney [8]. For each i choose a point a_i in C_i . Put $\beta = \min_i |p(a_i)|$, and note that $\beta > 0$. A component C_i is contained in some component J of $\mathbb{R}^n - S$, and C_i is the maximal connected set of points x in J containing a_i and such that $\text{sgn } p(x) = \text{sgn } p(a_i)$. For $0 < \alpha < \beta$, there is a component K of $J \cap \{x : |p(x)| > \alpha\}$ that contains a_i . This K is a component of E_α and is contained in C_i .

LEMMA 2.4. Given p_1, \dots, p_m , one can choose real numbers $\alpha_1, \dots, \alpha_m$ such that the hypersurfaces $N(p_1 - \alpha_1), N(p_1 + \alpha_1)$ together form a regular configuration and such that every topological component of $E = \mathbb{R}^n - \bigcup_{i=1}^m N(p_i)$ contains one of

$$(12) \quad F = \mathbb{R}^n - \bigcup_{i=1}^m N(p_i - \alpha_i) \cup N(p_i + \alpha_i).$$

Proof. Choose β as in Lemma 2.3 with p_1 in the role of p and the other p_i in the role of the q_i . By Lemma 2.2 (with $V = \mathbb{R}^n$), for all but finitely many α , $N(p_1 - \alpha)$ is regular. Hence we can choose α_1 between zero and β so that $N(p_1 - \alpha_1)$ and $N(p_1 + \alpha_1)$ are both regular.

If $\alpha_1, \dots, \alpha_k$ have been determined, choose β as in Lemma 2.3 with p_{k+1} in the role of p and $p_i - \alpha_i, p_i + \alpha_i, i = 1, \dots, k$, and $p_j, j = k + 2, \dots, m$, in the role of the q_i . Choose $\alpha_{k+1} \in (0, \beta)$ so that the hypersurfaces $N(p_i - \alpha_i), N(p_i + \alpha_i), i = 1, \dots, k + 1$, form a regular configuration. Such a choice is possible, in view of Lemma 2.2, because the number of intersections to consider is finite, and because all intersections among $N(p_j - \alpha_j), N(p_j + \alpha_j), j = 1, \dots, k$, are already regular. The lemma now follows by finite induction.

LEMMA 2.5. If the degree of p is d , the number of topological components of $N(p)$ does not exceed $2d^n$.

Proof. By the result of Whitney already cited, $N(p)$ has finitely many components. Suppose first that $N(p)$ is regular.

Let B_r be the closed ball of radius r about the origin in \mathbb{R}^n . Let S_r be the boundary of B_r . Choose r so large that every component of $N(p)$ intersects the interior of B_r and, by Lemma 2.2, so that $S_r \cap N(p)$ is regular. Oleĭnik and Petrovskii have shown (see [2, Lemma 1]), that a small variation in the coefficients of p does not change the topology of $B_r \cap N(p)$. In the coefficient space of the general real polynomial q of degree d , the points giving a variety $P(q)$ with any singularity form a proper subvariety. One concludes that a small variation of the coefficients of p will insure that the number of components of $N(p)$ will not decrease and that also $P(p)$ will be nonsingular. Assume that such a variation has been made.

For $j = 0, \dots, n$, embed \mathbb{R}^j in projective n -space as the points (x_1, \dots, x_{n+1}) with $x_{n+1} = \dots = x_{j+2} = 0, x_{j+1} \neq 0$. Call a subset of \mathbb{R}^j bounded if it has no limit points

in R^k for $k < j$. Since projective n -space is the union of R^n, \dots, R^0 , any unbounded component of $N(p)$ has a bounded component of $P(p) \cap R^j$ in its closure for some $j < n$. By again varying slightly the coefficients of p , we may assume that the $P(p) \cap R^j, j=2, \dots, n-1$, as well as $P(p)$, are nonsingular.

Oleĭnik and Petrovskiĭ have shown in [2] that a nonsingular algebraic hypersurface of degree d in R^j has at most $(d-1)^j$ bounded components, $j \geq 2$. By the classical theorem of Bezout a curve of order d in projective 2-space intersects the line at infinity (with multiplicities counted) in at most d real points. It follows that a plane curve (singular or nonsingular) of order d has at most d unbounded components in R^2 .

Since the restriction of $P(p)$ to R^j is again a nonsingular variety of order $\leq d$ and at most two topological components of $N(p)$ count against one bounded component of $P(p) \cap R^j, j=n-1, \dots, 3$, or arbitrary component of $P(p) \cap R^2$, the number of unbounded components if $N(p)$ does not exceed

$$(13) \quad 2[(d-1)^{n-1} + \dots + (d-1)^2 + (d-1) + 1].$$

The sum of (13) and $(d-1)^n$ is a bound for the number of all components of $N(p)$. For $d \geq 2$ and $n \geq 3$ this sum does not exceed d^n . When $n=2$, the number of bounded components does not exceed $(d-1)^2$, and there are at most d unbounded components. The sum again is $\leq d^2$. The case $d=1$ is clear. One concludes that a regular $N(p)$ has at most d^n topological components.

In the case of an arbitrary $N(p)$, consider a closed ball B_r containing all the bounded components of $N(p)$ in its interior. By the uniform continuity of p on B_r , there is an $\varepsilon > 0$ such that the ε -neighborhoods in R^n of the several bounded components of $N(p)$ are disjoint, and such that for all small $\alpha > 0$, each of these ε -neighborhoods will contain a bounded component of $N(p-\alpha)$ or $N(p+\alpha)$. By Lemma 2.2, $N(p-\alpha)$ and $N(p+\alpha)$ will be regular for all small $\alpha > 0$. Thus the number of bounded components of $N(p)$ is not more than $2(d-1)^n$. Similarly, estimates for the number of unbounded components of $N(p)$ come out to be double that for the regular case, and the general bound of $2d^n$ follows.

THEOREM 2. *Let p_1, \dots, p_m be real polynomials in n variables, each of degree d or less. The number of topological components of the set $R^n - \bigcup_{i=1}^m N(p_i)$ does not exceed $\sum_{k=0}^n 2(2d)^n 2^k C_{m,k}$, where $C_{m,k}$ is the usual binomial coefficient, except that $C_{m,k} = 0$ for $m < k$.*

Proof. By Lemma 2.4 it suffices to bound the number of components of the set F in (12). By Lemma 2.1 we may apply Theorem 1 to this case. To compute the parameters b_i needed in Theorem 1, note that if q_1, \dots, q_k have degrees $\leq d$ and $d = q_1^2 + \dots + q_k^2$, then p has degree $\leq 2d$ and $N(p) = \bigcap_{i=1}^k N(q_i)$. From this and Lemma 2.5 we see that any intersection of k hypersurfaces in the regular configuration cutting out the set F in (12) has at most $2(2d)^n$ topological components. Since two sets $N(p_i - \alpha_i)$ and $N(p_i + \alpha_i)$ have empty intersection, the parameter b_k

is bounded above by $2(2d)^n 2^k C_{m,k}$, $k=0, \dots, n$. If $k > n$, $b_k=0$, since the configuration of hypersurfaces is regular. Theorem 2 now follows from Theorem 1 and these bounds for the b_k .

4. Lower bounds for $D_{n,d}(K)$.

THEOREM 3. *Let p_1, \dots, p_m be real polynomials in n variables, each of degree at most $d \geq 1$. If $m \geq n$, the number of sign sequences $\text{sgn } p(x) = (\text{sgn } p_1(x), \dots, \text{sgn } p_m(x))$ that consist of terms $+1, -1$ does not exceed $(4edm/n)^n$.*

Proof. The sequence valued function $\text{sgn } p(x)$ has values with terms $+1, -1$ only on the set $E = \mathbb{R}^n - \bigcup_{i=1}^m N(p_i)$. Since the p_i are continuous, $\text{sgn } p(x)$ is constant on topological components of E . Thus the number of distinct values of $\text{sgn } p(x)$ does not exceed the number of components of E , which by Theorem 2 does not exceed

$$(14) \quad 2(2d)^n \sum_{k=0}^n 2^k C_{m,k}.$$

For $n=1, 2$, one can easily verify that except for $m=n=1$, (14) is less than $(4edm/n)^n$. When $m=n=1$, $\text{sgn } p(x)$ has at most $2 < 4ed$ values anyway. If $n \geq 3$, Stirling's formula shows that $4/n! < (e/n)^n$. Thus for $m \geq n \geq 3$,

$$(15) \quad \begin{aligned} 2(2d)^n \sum_{k=0}^n 2^k C_{m,k} &< 2(2d)^n \sum_{k=0}^n 2^k m^k / k! \\ &< 2(2d)^n \sum_{k=0}^n 2^k m^n / n! < 4(4dm)^n / n! < (4edm/n)^n, \end{aligned}$$

and the theorem is proved.

COROLLARY 3.1. *If $d \geq 2$ and $m \geq 8n \log_2 d$, then the number of distinct sequences of terms $+1, -1$ taken on by $\text{sgn } p(x)$ is less than 2^m .*

Proof. By Theorem 3 it suffices to show that $m/n \geq 8 \log_2 d$ implies $4ed(m/n) < 2^{m/n}$, or equivalently that $t \geq 8 \log_2 d$ implies $\log_2 (4edt) < t$. In fact, if $t \geq 8 \log_2 d$, then $t \geq 8$, since $d \geq 2$, and

$$(16) \quad \begin{aligned} \log_2 (4edt) &= \log_2 4e + \log_2 d + \log_2 t \\ &< 4t/8 + t/8 + 3t/8 = t. \end{aligned}$$

COROLLARY 3.2. *If $d \geq 2$ and $m \geq 18n \log_2 d$, there is a sign sequence $s = (s_1, \dots, s_m)$, $s_i = \pm 1$, which differs in more than $m/10$ places from any sequence in the range of $\text{sgn } p(x)$.*

Proof. First, $d \geq 2$ and $m \geq 18n \log_2 d$ imply that $(4edm/n)^n < 2^{m/2}$, or equivalently $t \geq 18 \log_2 d \geq 18$ implies $\log_2 (4edt) < t/2$. In fact, under these conditions,

$$(17) \quad \begin{aligned} \log_2 (4edt) &= 2 + \log_2 e + \log_2 d + \log_2 t \\ &\leq t/9 + t(\log_2 e)/18 + t/18 + t(\log_2 18)/18 < t/2. \end{aligned}$$

Consequently the range of $\text{sgn } p(x)$ contains fewer than $2^{m/2}$ distinct sequences of terms $+1, -1$.

The number of sign sequences s which differ from a particular sequence $\text{sgn } p(x)$ in at most k places is $1 + C_{m,1} + \dots + C_{m,k}$, that is, the sum of those which differ in exactly no place, one place, and so on. Using Stirling's lower estimate for $k!$, we see that since $m \geq k$, for $k \geq 1$,

$$(18) \quad \begin{aligned} 1 + C_{m,1} + \dots + C_{m,k} &< 1 + km^k/k! \\ &< 1 + (em/k)^k (k/2\pi) < (3m/k)^k. \end{aligned}$$

Now if $t \geq 10$, $3t < 2^{t/2}$, by an argument like that of (16) and (17). Thus if $k \leq m/10$, the number of sequences s which differ from a particular $p(x)$ in at most k places is less than $2^{m/2}$.

Now the corollary follows. For if $m \geq 18n \log_2 d$, the number of sequences s which differ from any value of $\text{sgn } p(x)$ in at most $m/10$ places is less than $2^{m/2} \cdot 2^{m/2} = 2^m$. But there are in all 2^m possible sequences s .

Let L be a real normed linear space with norm $\| \cdot \|$. Define the norms $\| \cdot \|_M$ and $\| \cdot \|_s$ on R^m by

$$(19) \quad \|(x_1, \dots, x_m)\|_M = \max_i |x_i|, \quad \|(x_1, \dots, x_m)\|_s = |x_1| + \dots + |x_m|.$$

For compact subsets K of L , define $\chi(K, m)$ to be the supremum of all numbers δ for which there exist linear functionals $\lambda_1, \dots, \lambda_m$ on L such that

(a) the mapping $T: L \rightarrow R^m$ given by $Tf = (\lambda_1(f), \dots, \lambda_m(f))$ has norm ≤ 1 with respect to the M -norm on R^m ;

(b) for each sequence (s_1, \dots, s_m) , $s_i = \pm 1$, there is a g in K such that

$$s_i \lambda_i(g) \geq \delta, \quad i = 1, \dots, m.$$

THEOREM 4. *If K is a compact subset of the real normed linear space L , then for $d \geq 2$,*

$$(20) \quad D_{n,d}(K) \geq \chi(K, -[-8n \log_2 d]).$$

Proof. Let m be the integer $-[-8n \log_2 d]$. The case $\chi(K, m) = 0$ is trivial. Otherwise, let δ be any positive number $< \chi(K, m)$. Choose linear functionals $\lambda_1, \dots, \lambda_m$ satisfying (a) and (b) above, and form the mapping T as in (a).

Let $F = \{P(x) : x \in R^n\}$ be, as in (4), a subset of L depending polynomially of degree d on n parameters. Since P has degree d , each of the functions $p_i(x) = \lambda_i(P(x))$ is a real polynomial in n variables of degree d or less. Because $m \geq 8n \log_2 d$, by Corollary 3.1, there is a sign sequence $s = (s_1, \dots, s_m)$, $s_i = \pm 1$, such that $s \neq (\text{sgn } p_1(x), \dots, \text{sgn } p_m(x))$ for any x in R^n . Choose $g \in K$ as in (b) above such that $s_i \lambda_i(g) \geq \delta$, $i = 1, \dots, m$. Now for each $x \in R^n$, Tf and $TP(x)$ differ in at least one coordinate in sign and hence differ in that coordinate by at least δ . Therefore,

$$(21) \quad D(K, F) \geq D(\{f\}, F) \geq D(\{Tf\}, T(F)) = \inf_x \|Tf - TP(x)\|_M \geq \delta.$$

Since δ can approach $\chi(K, m)$ from below, the theorem follows.

Define $\psi(K, m)$ to be the same as $\chi(K, m)$ except replace " M -norm" in condition (a) above by " S -norm".

THEOREM 5. *If K is as in Theorem 4 and $m \geq 18n \log_2 d$, then for $d \geq 2$,*

$$(22) \quad D_{n,d}(K) \geq (m/10)\psi(K, m).$$

Proof. Proceed as in the proof of Theorem 4. However, choose the sequence s such that, as in Corollary 3.2, s differs from each $(\operatorname{sgn} p_1(x), \dots, \operatorname{sgn} p_m(x))$ in more than $m/10$ places. Instead of (21) one has

$$(23) \quad D(K, F) \geq \inf_x \|Tf - TP(x)\|_s > (m/10)\delta,$$

and the theorem follows.

5. Approximation in the uniform norm. Let X be a set and let $B(X)$ be the Banach space, with uniform norm, of the bounded real valued functions on X . Point evaluation is a linear functional on $B(X)$. In fact, if x_1, \dots, x_m are points in X , the mapping $T: X \rightarrow \mathbb{R}^m$ defined by $Tf = (f(x_1), \dots, f(x_m))$ has norm one with respect to the M -norm on \mathbb{R}^m . This observation leads as follows to a practical method of estimating $\chi(K, m)$ for a compact subset K of $B(X)$.

For any m points x_1, \dots, x_m in X put

$$(24) \quad \Omega(K; x_1, \dots, x_m) = \min_{s_i = \pm 1} \sup_{g \in K} \min_{1 \leq i \leq m} s_i g(x_i).$$

Ω measures the extent to which K contains functions of arbitrary oscillation about zero on $\{x_1, \dots, x_m\}$.

THEOREM 6. *If K is a compact subset of $B(X)$, then*

$$(25) \quad \chi(K, m) \geq \sup \Omega(K; x_1, \dots, x_m),$$

where the supremum is over all sets of points x_1, \dots, x_m in X .

Proof. Ignoring a trivial case, we may assume that the supremum in (25) is positive. Let δ be any number less than this supremum. There are then points x_1, \dots, x_m in X such that $\Omega(K; x_1, \dots, x_m) > \delta$. For $i = 1, \dots, m$, define the linear functional λ_i on $B(X)$ by $\lambda_i(f) = f(x_i)$. The mapping $Tf = (\lambda_1(f), \dots, \lambda_m(f))$ has norm one with respect to the M -norm on \mathbb{R}^m . By the definition of Ω and the choice of the x_i , for every sequence (s_1, \dots, s_m) , $s_i = \pm 1$, there is a $g \in K$ such that $s_i \lambda_i(g) \geq \delta$, $i = 1, \dots, m$. Hence $\chi(K, m) \geq \delta$. Let δ approach $\sup \Omega(K; x_1, \dots, x_m)$ to finish the proof.

COROLLARY 6.1. *Let Λ^ω be as in §1. Then in $B([0, 1])$*

$$(26) \quad D_{n,d}(\Lambda^\omega) \geq \frac{1}{16} \omega\left(\frac{1}{n \log_2 d}\right)$$

for $d \geq 2$.

Proof. Let $m = -[-8n \log_2 d]$ and put $x_i = (i-1)/(m-1)$, $i = 1, \dots, m$. Clearly $\Omega(\Lambda^\omega; x_1, \dots, x_m) = \omega(1/(2m-2))$. By the subadditivity of ω and the definition of m , $\omega(1/(2m-2)) \geq (1/16)\omega(1/(n \log_2 d))$. Now (26) can be read off from Theorems 4 and 6.

Note that (26) holds *a fortiori* if we compute $D_{n,d}(\Lambda^\omega)$ in the subspace $C([0, 1])$ of $B([0, 1])$.

6. Approximation in the L_1 -norm. This section illustrates the use of Theorem 5. Let I^k be the k -dimensional cube $[0, 1] \times \dots \times [0, 1]$ (k factors). Let $\| \cdot \|$ denote the Euclidean norm on I^k . Let $\Lambda_{\alpha,k}$ be the real valued functions f on I^k which satisfy $|f(x+t) - f(x)| \leq \|t\|^\alpha$ for all x and $x+t$ in I^k ($0 < \alpha \leq 1$) and which are bounded by one. Let $L_1(I^k)$ be the usual Banach space of real Lebesgue integrable functions on I^k .

THEOREM 7. In $L_1(I^k)$, for $d \geq 2$,

$$(27) \quad D_{n,d}(\Lambda_{\alpha,k}) \geq (n \log_2 d)^{-\alpha/k} \frac{4^{-\alpha} 18^{-\alpha/k} k!}{10(k+\alpha) \cdots (1+\alpha)}.$$

Proof. Let r be the least integer such that $r^k \geq 18n \log_2 d$. Put $m = r^k$. Divide I^k into m disjoint cubes K_i each of side $1/r$. Define linear functionals λ_i on $L_1(I^k)$ by $\lambda_i(f) = \int_{K_i} f d\mu$, where μ is Lebesgue product measure on I^k . Note that the mapping $Tf = (\lambda_1(f), \dots, \lambda_m(f))$ has norm one with respect to the S -norm on R^m .

Given any sign sequence (s_1, \dots, s_m) , $s_i = \pm 1$, define g on I^k by $g(x) = s_i(\text{dist}(x, \partial K_i))^\alpha$ for x in K_i , where ∂K_i is the boundary of K_i and "dist" is Euclidean distance. It is straightforward to check that $g \in \Lambda_{\alpha,k}$. For $i = 1, \dots, m$,

$$(28) \quad \begin{aligned} s_i \lambda(g_i) &= \int_{K_i} (\text{dist}(x, \partial K_i))^\alpha d\mu \\ &= 2^k k! \int_0^{1/2r} \int_0^{x_1} \cdots \int_0^{x_{k-1}} x_k^\alpha dx_k \cdots dx_1 \\ &= m^{-1} r^{-\alpha} 2^{-\alpha} k! / [(k+\alpha) \cdots (1+\alpha)]. \end{aligned}$$

Thus $\psi(\Lambda_{\alpha,k}, m)$ is bounded below by (28). Since $r \geq 2$, $1/r \geq 2^{-1}/(r-1)$, and $r^{-\alpha} \geq 2^{-\alpha}(18n \log_2 d)^{-\alpha/k}$. Now (27) follows from Theorem 5.

REFERENCES

1. A. N. Kolmogorov, *Ueber die beste Annäherung von Funktionen einer gegebenen Funktionenklasse*, Ann. of Math. (2) 37 (1936), 107-110.
2. O. A. Oleĭnik and I. B. Petrovskii, *On the topology of real algebraic surfaces*, Izv. Akad. Nauk Ser. Mat. 13 (1949), 389-402; English transl., Transl. Amer. Math. Soc. (1) 7 (1962), 399-417.
3. H. S. Shapiro, *Some negative theorems of approximation theory*, Michigan Math. J. 11 (1964), 211-217.
4. A. F. Timan, *Theory of approximation of functions of a real variable*, Fizmatgiz, Moscow, 1960. (Russian)

5. A. G. Vitushkin, *On multidimensional variations*, GITTL, Moscow, 1955. (Russian)
6. ———, *Estimation of the complexity of the tabulation problem*, Fizmatgiz, Moscow, 1959. (Russian)
7. B. L. van der Waerden, *Einführung in die algebraische Geometrie*, Dover, New York, 1945.
8. H. Whitney, *Elementary structure of real algebraic varieties*, Ann. of Math. **66** (1957), 545–556.

UNIVERSITY OF OREGON,
EUGENE, OREGON