Negative Theorems in Approximation Theory

Allan Pinkus

1. INTRODUCTION. Negative theorems have a rich tradition in mathematics. In fact mathematics seems to be unique among the sciences in that negative results are very much a part of the mathematical edifice. Most mathematical theories try to explain what is possible and also what is not. To understand the structure of a mathematical theory is also to understand its limitations.

There are many different types of negative theorems. Some simply say that something is impossible. For example, one of the classical problems of Greek mathematics asks whether it is possible, using only a ruler and a compass, to square the circle, i.e., to construct a square with the same area as a given circle. This was answered in the negative in 1882 when Lindemann proved that π is transcendental. (Lambert had already proved that π is irrational, but this is insufficient to prove the impossibility of squaring the circle.) We recall that Abel established the insolvability of the general quintic equation, i.e., one cannot find a formula for the roots of fifth degree polynomials that involves only the coefficients of the polynomials and radicals. And, of course, there is Wiles's proof of Fermat's Last Theorem. Each of these problems and results is a fundamental part of our mathematical heritage. Each has had a profound influence. Nevertheless, the results themselves simply say that something is impossible.

Other negative results delineate what cannot be done, as compared with what is possible. The best of these results also highlight the salient features that illustrate why they are not possible. For example, Gauss proved that a necessary and sufficient condition for the construction of a regular n-gon, using only ruler and compass, is that

$$n = 2^k p_1 \cdots p_r,$$

where each p_j is a Fermat prime, a prime of the form $2^{2^\ell} + 1$. And Galois proved that a polynomial over a field K of characteristic zero is solvable by radicals if and only if the Galois group of the polynomial over K is solvable. Both these deep results tell us when certain things are possible, but also when they are impossible.

In this note we consider some negative results in approximation theory, what they tell us, and why they are important. Certain of these simply tell us that something is impossible. However most of the results delineate what is impossible in order to determine more fully what is possible. The results we present all center about one theme, namely, that approximation processes are necessarily of limited capability. As approximation theory is concerned with the ability to approximate functions and processes by simpler and more easily calculated objects, it is important to know these limitations. This can save us from spending time and effort on futile attempts to improve what cannot be bettered. These approximation processes are also limited for specific reasons, and we should be aware of and try to understand these reasons, especially if we hope to amend and circumvent them in some way.

There are many negative results in approximation theory. We discuss three main categories of results, with a few elementary examples from each. The first category consists of inverse theorems, the second set of results is connected with limitations of linear processes, and the third group is concerned with n-width type results. There have been a multitude of papers written in each of these areas. We can certainly not do justice to any of these topics within these few pages. Rather it is our hope to whet the reader's appetite to learn more about the subject.

2. INVERSE THEOREMS. Let $C[a,b]$ be the space of real-valued continuous functions on the real interval $[a,b]$, endowed with the uniform norm $\|\cdot\|$ ($\|f\| = \max\{|f(x)| : x \in [a,b]\}$), and let $\widetilde{C}$ be the space of real-valued 2π-periodic continuous functions on $\mathbb{R}$, also endowed with the uniform norm. We denote by Π_n the space of algebraic polynomials of degree at most n, i.e.,

$$\Pi_n = \operatorname{span}\{1, x, \ldots, x^n\},$$

and by T_n the space of trigonometric polynomials of degree at most n, i.e.,

$$T_n = \operatorname{span}\{1, \sin x, \cos x, \ldots, \sin nx, \cos nx\}.$$

We also set

$$E_n(f) = \min_{p \in \Pi_n} \|f - p\|$$

for f in $C[a,b]$, and

$$\widetilde{E}_n(g) = \min_{t \in T_n} \|g - t\|$$

for g in $\widetilde{C}$. These are measures of the distance of f and g from—or the errors in the approximation of f and g by—functions from Π_n and T_n, respectively.

In 1885 Weierstrass [**25**] proved that the set of algebraic polynomials is dense in $C[a,b]$ and that trigonometric polynomials are dense in $\widetilde{C}$. This may be formulated as

$$\lim_{n\to\infty} E_n(f) = 0$$

for every f in $C[a,b]$, and

$$\lim_{n\to\infty} \widetilde{E}_n(g) = 0$$

for each g in $\widetilde{C}$. The subsequent major themes in this theory were successfully tackled only some twenty-five years later. These concerned exact estimates for the rates at which $E_n(f)$ and $\widetilde{E}_n(g)$ tend to zero in terms of some intrinsic properties of f and g, and the development of useful and practical methods of approximation. These two problems, in diverse settings, were and are at the heart of approximation theory.

In direct theorems we estimate from above the error in approximating a given function by functions from certain approximation classes—the classical classes that play this role are algebraic and trigonometric polynomials, splines, etc.—based on certain qualities or characteristics of the function being approximated. *Inverse theorems* (sometimes called converse theorems) infer a quality or a characteristic of the function from this rate of approximation. Thus inverse theorems also provide us with lower bounds on the approximation error. In this sense they are negative theorems. Direct and inverse theorems are generally paired. The best of these theorems, such as those that follow, are paired because they determine exact rates of approximation.

Example 1. To illustrate what we mean, consider the example of $\widetilde{C}$ and approximations from T_n. We let $\operatorname{Lip}\alpha$ ($0 < \alpha \leq 1$) denote the Lipschitz class of all functions g

in $\widetilde{C}$ that satisfy

$$\left|g(x) - g(y)\right| \le M|x-y|^{\alpha}$$

for some constant M and all real x and y.

The following direct theorem is due to Jackson [**10**] (it is contained in Jackson's prize-winning doctoral thesis written in 1911 under the supervision of Edmund Landau): *if g belongs to $\widetilde{C}$ and its rth derivative $g^{(r)}$ is in* Lip α *for some α satisfying* $0 < \alpha \le 1$, *then there exists a constant M such that*

$$\widetilde{E}_n(g) \le Mn^{-r-\alpha}$$

for every n. So we see that, at least for continuous periodic functions, the smoother g is the better it is possible to approximate g by trigonometric polynomials. Of course, Jackson's theorem furnishes only an upper bound on the rate of approximation. It does not say that we cannot in some cases do much better. The answer to the question of whether or not it is possible to do better was essentially answered almost before it was asked in the wonderful prize-winning 1912 memoir of Bernstein [**2**], submitted even prior to the publication of Jackson's thesis. The final form as presented here was completed by de la Vallée Poussin [**23**] (whose contribution is generally overlooked).

The Bernstein inverse theorem states: *if there exist r in* $\mathbb{N}$, *α in* $(0, 1)$, *and a constant M independent of n such that*

$$\widetilde{E}_n(g) \le Mn^{-r-\alpha}$$

for all n, then g is r-times differentiable and $g^{(r)}$ belongs to Lip α. In other words, if g is not r-times differentiable with $g^{(r)}$ in Lip α, where $0 < \alpha < 1$, then necessarily

$$\limsup_{n\to\infty} n^{r+\alpha}\widetilde{E}_n(g) = \infty.$$

This is the negative implication of Bernstein's theorem. (The case $\alpha = 1$ is more complicated and was ultimately settled by Zygmund [**26**]. It involves second, rather than first, differences.)

At first glance these two theorems, and more especially the latter, seem rather surprising. Why should the smoothness of a g in $\widetilde{C}$ depend upon the rate of decrease of $\widetilde{E}_n(g)$? Hindsight is a marvelous teacher. We now realize that the critical role in the Bernstein theorem is played by the Bernstein inequality:

$$\|t'\| \le n\|t\|$$

for every t in T_n. This measures, in some sense, the "stiffness" of the polynomials. (The inequality was first proved in the aforementioned Bernstein memoir [**2**] in a slightly weaker form.)

Here is a short outline of a proof for the case $r = 0$ and $0 < \alpha < 1$; i.e., we assume that $\widetilde{E}_n(g) \le Mn^{-\alpha}$ for all n. Let t_n denote the best approximant to g from T_n, and set $v_0 = t_1$ and $v_n = t_{2^n} - t_{2^{n-1}}$ for $n = 1, 2, \dots$. Note that

$$\|v_n\| \le \widetilde{E}_{2^n}(g) + \widetilde{E}_{2^{n-1}}(g) \le M_1 2^{-n\alpha}.$$

Because the t_n approximate g uniformly we can express g as the sum of a telescoping series:

$$g = \sum_{n=0}^{\infty} v_n.$$

From the mean value theorem and Bernstein's inequality it follows for given x and y that

$$\left|v_n(x) - v_n(y)\right| \le \|v_n'\| \, |x-y| \le 2^n \|v_n\| \, |x-y|.$$

Thus for every m

$$\begin{aligned}
\left|g(x) - g(y)\right| &= \left|\sum_{n=0}^{\infty} \left(v_n(x) - v_n(y)\right)\right| \le \sum_{n=0}^{m-1} \left|v_n(x) - v_n(y)\right| + 2\sum_{n=m}^{\infty} \|v_n\| \\
&\le \sum_{n=0}^{m-1} 2^n \|v_n\| \, |x-y| + 2\sum_{n=m}^{\infty} \|v_n\| \\
&\le M_1 |x-y| \sum_{n=0}^{m-1} 2^{n(1-\alpha)} + 2M_1 \sum_{n=m}^{\infty} 2^{-n\alpha} \\
&\le M_2 \left[|x-y| 2^{m(1-\alpha)} + 2^{-m\alpha}\right].
\end{aligned}$$

In the last step we use the fact that α lies in (0, 1). We now choose m so that $2^{-m} \sim |x-y|$, from which it follows that

$$\left|g(x) - g(y)\right| \le M_3 |x-y|^{\alpha}.$$

This puts g in $\operatorname{Lip}\alpha$.

When approximating by algebraic polynomials on a finite interval, the results are similar to those of Example 1, except that there are essential problems near the endpoints. This is due to the fact that for x near an endpoint there exist p in Π_n for which $|p'(x)|$ is as large as $n^2\|p\|$. Away from the endpoints a bound on the order of $n\|p\|$ still holds.

Example 2. Here is another (direct and inverse) theorem that contains both positive and negative information. A function f defined on $[-1, 1]$ is said to be *analytic* on this interval if it can be extended to a complex analytic function on some open set containing $[-1, 1]$. In this context the important open sets are the interiors of the ellipses with foci ± 1 and with semi-axes having sums ρ larger than one. Let E_ρ denote any such ellipse, say

$$E_\rho = \left\{ \left(\frac{1}{2}\left(\rho + \frac{1}{\rho}\right)\cos\theta, \frac{1}{2}\left(\rho - \frac{1}{\rho}\right)\sin\theta \right) : 0 \le \theta \le 2\pi \right\},$$

and let D_ρ signify the interior of E_ρ.

The following result is also in Bernstein [**2**]: *f has an analytic extension to D_r but to no D_ρ with $\rho > r$ if and only if*

$$\limsup_{n\to\infty} \left(E_n(f)\right)^{1/n} = \frac{1}{r}.$$

(There is a close connection between this result and the formula for the radius of convergence of a power series in terms of the coefficients of the power series.) The negative content of this theorem is that, if f is not analytic on $[-1, 1]$, then we cannot possibly obtain a geometric rate of convergence to zero for the error in any approximation to f by algebraic polynomials.

Example 3. *Splines* are piecewise polynomials with a high order of continuity. When $-\infty = t_0 < t_1 < \cdots < t_k < t_{k+1} = \infty$ we set

$$\mathcal{S}_{n,k}(t_1, \ldots, t_k) = \left\{ s \in C^{(n-1)}(\mathbb{R}) : s \in \Pi_n|_{(t_{i-1}, t_i)},\ i = 1, \ldots, k+1 \right\}.$$

Thus a function s belongs to $\mathcal{S}_{n,k}(t_1, \ldots, t_k)$ if it has a certain level of global smoothness and is a polynomial of degree at most n on each of the intervals (t_{i-1}, t_i). We say that $\mathcal{S}_{n,k}(t_1, \ldots, t_k)$ is the space of *splines of degree n with the simple knots* $\{t_1, \ldots, t_k\}$. When using splines one fixes the degree and permits the number (and placement) of the knots to vary. From the perspective of numerical computations, approximation by splines enjoys many advantages over approximation by algebraic and trigonometric polynomials. However, direct and inverse theorems for approximation by splines with equally spaced knots are, up to a point, similar to those for approximation by polynomials.

Let $\mathcal{S}^*_{n,k}$ denote the space of splines of degree $n \geq 1$ with k equally spaced knots in the interval $[a, b]$, and let

$$e_k(f) := \min_{s \in \mathcal{S}^*_{n,k}} \|f - s\|.$$

Then from Richards [**19**] we learn:

$$e_k(f) \leq M k^{-r-\alpha}$$

for fixed α *in* $(0, 1)$ *and* r *in* $\mathbb{N}$ *with* $r + \alpha < n + 1$ *if and only if* f *belongs to* $C[a, b]$ *and* $f^{(r)}$ *to* $\operatorname{Lip}\alpha$. *But if*

$$e_k(f) \leq M k^{-r}$$

for $r > n + 1$, *then necessarily* f *belongs to* Π_n *(a subclass of* $\mathcal{S}^*_{n,k}$*), in which event* $e_k(f) = 0$ *for every* k. In other words, when approximating by splines of a fixed degree we are also limited in the rate of convergence of the error to zero by the degree of the splines themselves.

We have described three fundamental direct and inverse theorems. There are many, many generalizations of these results. The excellent textbooks of Timan [**22**] and of DeVore and Lorentz [**6**] contain an abundance of information on direct and inverse theorems for approximation by algebraic and trigonometric polynomials. For approximation by splines with fixed or variable knots, see Schumaker [**20**] or DeVore and Lorentz [**6**]. The study of direct and inverse theorems remains an active area of research. This research now centers on approximation by members of various nonclassical and nonlinear classes such as wavelets, shift-invariant subspaces, radial basis functions, ridge functions, neural nets, multivariate splines, and the like.

3. LINEAR METHODS. Finding the best or even a good approximation is generally a difficult and costly procedure. In the setting of this paper best approximation operators are nonlinear. Linear methods of approximation are both preferable and simpler, and much effort has gone into their study. As it happens, however, linear processes often have intrinsic limitations. For example, it is quite natural to try to use interpolation to obtain good approximations. Let us assume that we are given a fixed triangular array of points with $n+1$ points in $[a,b]$ in the nth row of the array, and let $p_n(f)$ denote the (unique) polynomial in Π_n that interpolates f in $C[a,b]$ at the points of this nth row. It was Faber who showed in a 1914 paper [**7**] that for every such array there always exists an f in $C[a,b]$ for which

$$\lim_{n\to\infty} \left\| f - p_n(f) \right\| \neq 0.$$

This was subsequently generalized by Lozinski and Kharshiladze (see Lozinski [**16**]), who proved the following: *for any given sequence* (L_n) *of linear operators such that* L_n *maps* $C[a,b]$ *to* Π_n *and* $L_n p = p$ *for all* p *in* Π_n *(*L_n *is a projection of* $C[a,b]$ *onto* Π_n*), there exists an* f *in* $C[a,b]$ *such that*

$$\lim_{n\to\infty} \| f - L_n f \| \neq 0.$$

From this we see that there cannot exist a sequence of linear operators $L_n : C[a,b] \to \Pi_n$ satisfying

$$\left\| f - L_n(f) \right\| \leq M E_n(f) \tag{1}$$

for some constant M and for all n and f. (If this inequality were to hold, then each L_n would be a projection onto Π_n and the norms $\| f - L_n f \|$ would tend to zero as $n \to \infty$ for every f in $C[a,b]$.) The analogous result holds for $\widetilde{C}$ and trigonometric polynomials. One of the advantages when approximating by splines is that there do exist sequences of linear operators that satisfy (1), i.e., that provide asymptotically optimal rates of approximation.

Another problem we must be mindful of is that linear approximation processes may exhibit inherent limitations independent of any quality or characteristic of the function being approximated. We consider some elementary examples.

Example 4. One of the simplest and most natural examples of a linear approximation process on $\widetilde{C}$ is given by the sequence of partial sums s_n of a Fourier series. Strikingly, as du Bois-Raymond proved in 1875 [**4**], these sequences of partial sums do not necessarily converge for all functions in $\widetilde{C}$. In 1900 Fejér [**9**] (at the time a twenty-year-old student) gave an elementary and brilliant proof of the Weierstrass theorem. In a wonderful piece of lateral thinking, he proved that the Cesàro sums of the Fourier series of any g in $\widetilde{C}$ converge uniformly to g. In other words, assume that we are given the Fourier series of g,

$$g(x) \sim \sum_{k=-\infty}^{\infty} c_k e^{ikx},$$

where

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} g(x) e^{-ikx}\, dx$$

for every k in $\mathbb{Z}$. Consider the nth partial sum s_n of the Fourier series,

$$s_n(x) := \sum_{k=-n}^{n} c_k e^{ikx},$$

and set

$$\sigma_n(g; x) = \frac{s_0(x) + \cdots + s_n(x)}{n+1}.$$

We call σ_n the *nth Fejér operator*. Note that $\sigma_n(g; \cdot)$ belongs to T_n for each n. What Fejér proved was that, for each g in $\widetilde{C}$, $\sigma_n(g; \cdot)$ tends uniformly to g as $n \to \infty$. However the following was noted by Zygmund [**27**]: *if*

$$\|g - \sigma_n(g; \cdot)\| = o\left(\frac{1}{n}\right),$$

then g is a constant function. (The notation $a_n = o(\varepsilon_n)$ means that $a_n/\varepsilon_n \to 0$ as $n \to \infty$.)

Thus except in the very trivial case, and independent of any smoothness property of g, the Fejér operators $\sigma_n(g; \cdot)$ approximate g very slowly. The proof of this fact is really quite simple, so we present it here. As is easily checked,

$$\sigma_n(g; x) = \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) c_k e^{ikx},$$

which implies that

$$\frac{1}{2\pi} \int_0^{2\pi} \left[g(x) - \sigma_n(g; x)\right] e^{-ikx}\, dx = c_k \frac{|k|}{n+1}$$

for $-n \le k \le n$. Therefore

$$\left| c_k \frac{|k|}{n+1} \right| = \left| \frac{1}{2\pi} \int_0^{2\pi} \left[g(x) - \sigma_n(g; x)\right] e^{-ikx}\, dx \right| \le \|g - \sigma_n(g; \cdot)\|$$

and

$$|kc_k| \le \lim_{n\to\infty} n \left\| g - \sigma_n(g; \cdot) \right\|.$$

Accordingly, if $\|g - \sigma_n(g; \cdot)\| = o(1/n)$, then $c_k = 0$ for all $k \ne 0$, whence g is constant.

Saturation theory studies results of the foregoing type, i.e., it considers when a certain approximation rate for a linear process can be satisfied only by some trivial class of functions. In saturation theory one is also interested in determining the best rate of approximation that can be obtained by a nontrivial class of functions and in characterizing this class. For instance, in the case of the Fejér operators σ_n it can be shown that

$$\|g - \sigma_n(g; \cdot)\| = O\left(\frac{1}{n}\right)$$

if and only if g belongs to $\widetilde{C}$, g' is absolutely continuous, and g' is in Lip 1. (To write $a_n = O(\varepsilon_n)$ means that a_n/ε_n remains bounded for all n.) Saturation theory was first defined formally by Favard in 1949 [**8**], although some of the results of this theory predate Favard's paper.

Example 5. The important *Bernstein polynomials* (introduced by Bernstein in [**3**]) are given for f in $C[0, 1]$ by

$$B_n(f; x) := \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}.$$

The following saturation result due to Lorentz [**14**] (see also Bajšanski and Bojanić [**1**]) has its origins in a result of Voronovskaya [**24**]: *the only functions f in $C[0, 1]$ satisfying*

$$\left|f(x) - B_n(f; x)\right| \leq \varepsilon_n \frac{x(1-x)}{n}$$

for some sequence $\varepsilon_n \to 0$ are the linear functions.

Both $f \mapsto B_n(f)$ and $f \mapsto \sigma_n(f)$ are examples of positive ($f \geq 0$ implies $B_n(f) \geq 0$ and $\sigma_n(f) \geq 0$) linear operators. We note the following result: *if (L_n) is a sequence of positive linear operators from $C[a, b]$ to $C[a, b]$, then*

$$\lim_{n\to\infty} \left\|f - L_n(f)\right\| = 0 \tag{2}$$

for every f in $C[a, b]$ if and only if (2) holds for the three functions 1, x, and x^2. This is the content of the Bohman–Korovkin theorem of the 1950s (see Korovkin [**13**]). A similar result holds in $\widetilde{C}$, where the "test functions" are 1, $\sin x$, and $\cos x$.

Positivity is certainly a desirable quality for a linear approximation operator. However, all sequences (L_n) of positive linear operators mapping $C[a, b]$ to Π_n (or $\widetilde{C}$ to T_n) seem to exhibit a saturation property, or at least to have a fixed lower bound on the rate of approximation for anything outside a trivial class of functions. But we are unaware of any general theorem covering all cases. Is it true, for example, that for an arbitrary sequence of positive linear operators $L_n : C[a, b] \to \Pi_n$ the only functions for which

$$\left\|f - L_n(f)\right\| = o\left(\frac{1}{n^2}\right)$$

belong to Π_2 or are in some other fixed finite-dimensional set? It is known that for at least one of the functions 1, x, or x^2 we always have

$$\left\|f - L_n(f)\right\| \neq o\left(\frac{1}{n^2}\right).$$

4. *n*-WIDTHS. In section 2 we noted that for g in $\widetilde{C}$ the rate of decrease of $\widetilde{E}_n(g)$ is directly related to the smoothness of g. But we are not interested in smoothness. We want efficient and effective approximation procedures. As this is the case, then perhaps trigonometric polynomials are not good approximation classes. Maybe there are better ones. How do we find better approximation classes, and how can we compare the efficacy of different approximation classes? In 1936 Kolmogorov [**11**] proposed the following noteworthy idea. Given a set of functions, rather than considering approximation from some fixed linear subspace (such as T_n in our example) why don't we try to find a "best" approximating subspace of the same dimension?

As an example, recall that if g and $g^{(r)}$ belong to $\widetilde{C}$, then

$$\widetilde{E}_n(g) \leq Mn^{-r}\|g^{(r)}\| \tag{3}$$

for some constant M independent of g and n (Jackson's theorem). Define

$$\widetilde{B}_r = \left\{ g : g, g^{(r)} \in \widetilde{C},\ \|g^{(r)}\| \le 1 \right\}.$$

For each g in $\widetilde{B}_r$ we see in light of (3) that

$$\widetilde{E}_n(g) \le Mn^{-r},$$

which we can rewrite as

$$E(\widetilde{B}_r; T_n) = \sup_{g \in \widetilde{B}_r} \min_{t \in T_n} \|g - t\| \le Mn^{-r}. \tag{4}$$

The quantity $E(\widetilde{B}_r; T_n)$ is a measure of the distance from T_n to $\widetilde{B}_r$ in that it is the supremum of the errors obtained in approximating members g of $\widetilde{B}_r$ by elements t from T_n. (This is called "worst case" error.)

Kolmogorov's idea was to consider how well one might possibly be able to approximate the set $\widetilde{B}_r$ in the sense suggested by (4), not necessarily by T_n, but by any linear subspace of $\widetilde{C}$ whose dimension is the same as that of T_n, namely, $2n + 1$. In other words, he defined

$$d_{2n+1}(\widetilde{B}_r) = \inf_{X_{2n+1}} \sup_{g \in \widetilde{B}_r} \inf_{h \in X_{2n+1}} \|g - h\|,$$

where the left-most infimum varies over all subspaces X_{2n+1} of $\widetilde{C}$ of dimension at most $2n + 1$. This is now called the *Kolmogorov* $(2n + 1)$*-width of* $\widetilde{B}_r$. It is a measure of how well any $(2n + 1)$-dimensional subspace can possibly approximate $\widetilde{B}_r$.

In general, given a set A in a normed linear space X, the *Kolmogorov n-width $d_n(A)$ of A in X* is defined by:

$$d_n(A) = \inf_{X_n} \sup_{f \in A} \inf_{h \in X_n} \|f - h\|_X,$$

where X_n varies over all n-dimensional subspaces of X. The quantity $d_n(A)$ delineates what is feasible and what is not in this approximation context. On the one hand it is a measure of the extent to which A may be approximated by n-dimensional subspaces of X. On the other hand the n-widths $d_n(A)$ are lower bounds on how well it is possible to approximate A. That is, for each $\varepsilon > 0$ and each n-dimensional subspace X_n of X there necessarily exists an f in A (dependent upon X_n) satisfying

$$\inf_{h \in X_n} \|f - h\|_X \ge d_n(A) - \varepsilon.$$

In this sense n-width estimates are negative results. We can never do better than $d_n(A)$ when approximating A by n-dimensional subspaces.

It happens, but it is rare, that one can identify a subspace X_n for which

$$E(A; X_n) = \sup_{f \in A} \inf_{h \in X_n} \|f - h\|_X = d_n(A).$$

In general one is more interested in identifying *asymptotically optimal subspaces*, meaning sequences (X_n) of n-dimensional subspaces X_n such that

$$E(A; X_n) \le M d_n(A)$$

for some constant M and for all n. These are good approximating subspaces for A in that they yield a rate of approximation that, up to the constant M, cannot be bettered. Unless the constant matters, it is generally not worth spending the time and effort to search for better subspaces. This is why it is important to pin down the asymptotics of n-widths.

Example 6. The example of $\widetilde{B}_r$ in $\widetilde{C}$ represents one of those rare cases where we can identify "best" subspaces. Tikhomirov [**21**] proved that

$$E(\widetilde{B}_r; T_n) = d_{2n+1}(\widetilde{B}_r) \approx n^{-r}.$$

However we should also note that the more stable space of periodic splines of degree $r-1$ with $2n+2$ equally spaced simple knots (a subspace of dimension $2n+2$) is optimal for $d_{2n+2}(\widetilde{B}_r)$, and additionally

$$d_{2n+1}(\widetilde{B}_r) = d_{2n+2}(\widetilde{B}_r).$$

There are also linear operators that map $\widetilde{C}$ to T_n and to the indicated space of splines that attain this same error bound.

Example 7. Let

$$B_r := \left\{ f : f, f^{(r)} \in C[a,b],\ \|f^{(r)}\| \le 1 \right\}.$$

Then $d_n(B_r)$ is also asymptotically of the order n^{-r}, and the spaces Π_{n-1} of algebraic polynomials of degree at most $n-1$ are asymptotically optimal subspaces. However, the Π_{n-1} are not optimal subspaces, nor are they very useful in practice. Optimal subspaces for $d_n(B_r)$ are splines of degree $r-1$ with certain specific $n-r$ simple knots. We can characterize these optimal knots, but their exact values are not easily calculated. Nonetheless, the exact same order of approximation can be achieved by using the more practical space of splines of degree $r-1$ with $n-r$ equally spaced simple knots.

Kolmogorov n-widths have been studied in a variety of different settings and for many diverse sets of functions. The interested reader is referred to Pinkus [**18**], Korneĭchuk [**12**], Lorentz, von Golitschek, and Makovoz [**15**], and DeVore [**5**]. Perhaps just as important as the Kolmogorov n-widths themselves is the idea behind the concept that has spawned a myriad of variations. For instance, in our examples we considered "worst case" error. Perhaps "worst case" error is far from typical and we should look at average errors or some variant thereof (see, for example, Maiorov and Wasilkowski [**17**]). Other variations include the *Gel′fand n-width*, which considers the best n bits of linear information with which to approximate, and the *linear n-width*, which asks for the best rank n linear approximation operators. In addition, numerous nonlinear n-width concepts have been proposed for providing effective lower bounds on different classes of nonlinear approximation schemes.

5. EPILOGUE. In these pages we have described results that demarcate the effectiveness of approximation processes. Inverse theorems, which depend on the approximating subspaces, define certain qualities or characteristics of the functions being approximated in terms of the rate at which the error in the best approximation tends to zero. Accordingly, for functions without these qualities, inverse theorems provide a lower bound on the rate of approximation. Linear approximation processes often exhibit inherent limitations on the rates of approximation independent of the functions being approximated. In the study of n-widths one obtains structural lower bounds on how well it is possible to approximate classes of functions by n-dimensional subspaces, independent of the choice of the approximating subspaces. Researchers today are very much interested in multivariate and nonlinear approximation. Many different approximation procedures have been suggested. It is only by the study of the approximation properties of these multivariate and nonlinear approximations that we will more fully understand their advantages and disadvantages. Negative results such as inverse theorems, the limitations of linear approximation processes, and n-width type results have an important role to play in this endeavor.

ACKNOWLEDGMENTS. This article is based on a lecture for graduate students presented at Texas A&M University in Fall 2002 as part of the Frontiers in Mathematics Lecture Series. I would like to express my appreciation to the faculty of the mathematics department of Texas A&M University for their warm hospitality. I also want to thank Yoav Benyamini and Carl de Boor for their constructive criticisms.

REFERENCES

1. B. Bajšanski and R. Bojanić, A note on approximation by Bernstein polynomials, *Bull. Amer. Math. Soc.* **70** (1964) 675–677.
2. S. N. Bernstein, Sur l'ordre de la meilleure approximation des fonctions continues par les polynomes de degré donné, *Mém. Cl. Sci. Acad. Roy. Belg.* **4** (1912) 1–103.
3. ———, Démonstration du théorème de Weierstrass fondée sur la calcul des probabilités, *Comm. Kharkov Math. Soc.* **13** (1912/13) 1–2.
4. P. du Bois-Raymond, Versuch einer Classification der willkürlichen Functionen reeller Argumente nach ihren Änderungen in den kleinsten Intervallen, *J. Reine Angew. Math.* **79** (1875) 21–37.
5. R. A. DeVore, Nonlinear approximation, *Acta Numer.* **7** (1998) 51–150.
6. R. A. DeVore and G. G. Lorentz, *Constructive Approximation*, Springer-Verlag, Berlin, 1993.
7. G. Faber, Über die interpolatorische Darstellung stetiger Funktionen, *Jahresber. Deutsch. Math.-Verein.* **23** (1914) 190–210.
8. J. Favard, Sur l'approximation dans les espaces vectoriels, *Ann. Mat. Pura Appl.* **29** (1949) 97–110.
9. L. Fejér, Sur les fonctions bornées et intègrables, *C. R. Acad. Sci. Paris* **131** (1900) 984–987.
10. D. Jackson, Über die Genauigkeit der Annäherung stetiger Funktionen durch ganze rationale Funktionen gegebenen Grades und trigonometrische Summen gegebener Ordnung, Preisschrift und Dissertation, Universität Göttingen, 1911.
11. A. Kolmogoroff, Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse, *Ann. of Math. (2)* **37** (1936) 107–110.
12. N. Korneĭchuk, *Exact Constants in Approximation Theory*, Cambridge University Press, Cambridge, 1991.
13. P. P. Korovkin, *Linear Operators and Approximation Theory*, Hindustan Publ. Corp., Delhi, 1960; the Russian original appeared in 1959.
14. G. G. Lorentz, Inequalities and the saturation classes of Bernstein polynomials, in *On Approximation Theory* (Proceedings of Conference in Oberwolfach, 1963), P. L. Butzer and J. Korevaar, eds., Birkhäuser, Basel, 1964, pp. 200–207.
15. G. G. Lorentz, M. von Golitschek, and Y. Makovoz, *Constructive Approximation: Advanced Problems*, Springer-Verlag, Berlin, 1996.
16. S. Lozinski, On a class of linear operators, *Dokl. Akad. Nauk SSSR* **61** (1948) 193–196 (Russian).
17. V. E. Maiorov and G. W. Wasilkowski, Probabilistic and average linear widths in L_∞-norm with respect to r-fold Wiener measure, *J. Approx. Theory* **84** (1996) 31–40.
18. A. Pinkus, *n-Widths in Approximation Theory*, Springer-Verlag, Berlin, 1985.
19. F. Richards, On the saturation class for spline functions, *Proc. Amer. Math. Soc.* **33** (1972) 471–476.
20. L. L. Schumaker, *Spline Functions: Basic Theory*, Wiley, New York, 1981.
21. V. M. Tikhomirov, Diameters of sets in functional spaces and the theory of best approximations, *Uspekhi Mat. Nauk* **15** (1960) 81–120 (Russian); translated in *Russian Math. Surveys* **15** (1960) 75–111.
22. A. F. Timan, *Theory of Approximation of Functions of a Real Variable*, Pergamon, Oxford, 1963.
23. Ch. J. de la Vallée Poussin, L'approximation des fonctions d'une variable réelle, *Enseign. Math.* **20** (1918) 5–29.
24. E. V. Voronovskaya, Determination de la forme asymptotique d'approximation des fonctions par les polynomes de M. Bernstein, *Dokl. Akad. Nauk SSSR* **4** (1932) 79–85 (Russian).

25. K. Weierstrass, Über die analytsche Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen, *Sitzungsberichte der Akademie zu Berlin* (1885) 633–639, 789–805; an expanded version of this paper with ten additional pages also appeared in *Mathematische Werke*, vol. 3, Mayer & Müller, Berlin, 1903, pp. 1–37.
26. A. Zygmund, Smooth functions, *Duke Math. J.* **12** (1945) 47–76.
27. ———, The approximation of functions by typical means of their Fourier series, *Duke Math. J.* **12** (1945) 695–704.

ALLAN PINKUS was born in Montreal, Canada in 1946. He received his B. Sc. from McGill University and his Ph.D. from the Weizmann Institute. Since 1977 he has been at the Technion. His research interests center about approximation theory. Further biographical information can be found at http://math.technion.ac.il/˜pinkus.
Technion, Haifa 32000, Israel
pinkus@tx.technion.ac.il