

# A view of Approximation Theory

by T. J. Rivlin

**This is a selective survey of Approximation Theory which touches on the concepts of best approximation, good approximation, approximation of classes of functions, approximation of functionals, and optimal estimation.**

## Introduction

I present a view of the subject called Approximation Theory which makes no pretense at being complete or objective. It focuses much of its attention on aspects of the subject which have particularly interested me and, especially, in which I and my friends have worked. In taking this approach I am following in the footsteps of—among many others—one of the founding fathers of the subject, A. A. Markov (1856–1922), who, apparently in reply to the question “What is mathematics?”, said: “Mathematics is that which Gauss, Chebyshev, Lyapunov, Steklov and I study” (as quoted in [1]).

## Early work

The broad river of contemporary Approximation Theory derives from the confluence of two prominent 19th-century tributaries.

### • Best approximation

The idea of best approximation according to P. L. Chebyshev (1821–1894) is best exemplified in the

©Copyright 1987 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

following prototypical setting. Let  $I = [-1, 1]$  and  $p_n$  denote the polynomials of degree  $\leq n$ . For  $f \in C(I)$  we put

$$\|f\| = \max_{x \in I} |f(x)|.$$

$p^* \in p_n$  is called a *best approximation* to  $f$  on  $I$  out of  $p_n$  if  $\|f - p^*\| \leq \|f - p\|$ , for all  $p \in p_n$ . Such a  $p^*$  always exists, and is, indeed, unique. We also put  $E_n(f) = \|f - p^*\|$ . We can now state the famous *Alternation Theorem* (Chebyshev–Borel, cf. [2]).

$p^* \in p_n$  is a best approximation to  $f \in C(I)$  if, and only if, there exist  $n + 2$  distinct points of  $I$ ,  $-1 \leq x_1 < \dots < x_{n+2} \leq 1$ , for which

$$f(x_i) - p^*(x_i) = \epsilon(-1)^i \|f - p^*\|, \\ i = 1, \dots, n + 2; \epsilon = \pm 1.$$

The main themes of the Chebyshevian approach are existence, uniqueness, and characterization of best approximations. The setting we have just presented has been generalized in subsequent work in every conceivable way: with respect to the function being approximated, the approximating class, and the measure of error. The problem of best approximation is a typical extremal problem, and as is generally the case with such problems, exact solutions are hard to come by. However, in one interesting case Chebyshev [3] obtained a best approximation in closed form, namely, approximation of  $x^n$  out of  $p_{n-1}$  on  $I$ . He proceeded as follows. Suppose  $p^* \in p_{n-1}$  is a best approximation to  $x^n$  on  $I$ . Then  $e(x) = x^n - p^*(x) \in p_n$ . By the Alternation Theorem there exist  $n + 1$  points satisfying  $-1 \leq x_1 < \dots < x_{n+1} \leq 1$  such that

$$|e(x_i)| = \|e\| =: M, \quad i = 1, \dots, n + 1.$$

But then  $x_1 = -1$  and  $x_{n+1} = 1$ , for otherwise  $e'(x) = 0$  at  $n$  interior points of  $I$ . Now  $M^2 - e^2(x) \geq 0$ ,  $x \in I$ , and so

$M^2 - e^2 \in P_{2n}$  has  $x_2, \dots, x_n$  as zeros of order 2 and  $\pm 1$  as simple zeros, thus accounting for all  $2n$  of its zeros. But  $(1 - x^2)[e'(x)]^2$  has exactly the same zeros and hence

$$(1 - x^2)[e'(x)]^2 = n^2(M^2 - e^2(x)). \quad (1)$$

Thus, if  $x = \cos \theta$ ,

$$e(x) = M \cos n\theta = MT_n(x) = 2^{1-n}T_n(x)$$

$$=: \hat{T}_n(x),$$

the monic Chebyshev polynomial of degree  $n$ . Finally then,

$$p^*(x) = x^n - \hat{T}_n(x)$$

and

$$E_{n-1}(x^n) = 2^{1-n}.$$

The Chebyshev polynomial which makes its appearance as a solution of (1) has played an important role in Approximation Theory (cf. [2]).

#### • Good approximation

The second approach originates with Weierstrass and his famous *Approximation Theorem*: Given  $f \in C(I)$  and  $\varepsilon > 0$ , then there exist  $n$  and  $p \in P_n$  such that  $\|f - p\| < \varepsilon$ . In other words, continuous functions can be (well) approximated by polynomials. Weierstrass' proof (1885) involved convolution of  $f$  with an appropriate kernel. In 1898 Lebesgue [4] demonstrated that it sufficed to show that the function  $|x|$  could be approximated by polynomials in order to prove the same for all continuous functions. This approach led to interest in the quantitative aspects of approximating  $|x|$ . For instance, S. N. Bernstein (1880–1968) showed, in his prize-winning memoir [5], that

$$\frac{1}{2\pi} \frac{1}{2n-1} \leq E_n(|x|) \leq \frac{2}{\pi \left( \left[ \frac{n}{2} \right] + 1 \right)}, \quad (2)$$

and later showed [6] that

$$\lim_{n \rightarrow \infty} 2n E_{2n}(|x|) = 0.282 \dots =: \mu.$$

In connection with this last result he remarked: "It would be very interesting to determine if  $\mu$  is a new transcendental or if it can be expressed in terms of known transcendentals. I note, as a curious coincidence, that  $1/(2\sqrt{\pi}) = 0.282 \dots$ " [However, Varga and Carpenter [7] showed in 1983, using careful numerical computation, that  $\mu \neq 1/(2\sqrt{\pi})$ .]

Speaking of approximation of  $|x|$ , let us jump ahead to D. J. Newman's [8] striking result of 1964 about approximation of  $|x|$  on  $I$  by *rational functions* of degree  $n$  (i.e., rational functions whose numerator and denominator are both in  $P_n$ ). Newman showed that, if we denote the (uniform) error of the best rational approximation of degree  $n$  to  $|x|$  on  $I$  by  $E_{n,n}(|x|)$ , we have

$$\frac{1}{2} e^{-9\sqrt{n}} \leq E_{n,n}(|x|) \leq 3e^{-\sqrt{n}}, \quad (3)$$

a result which compares most favorably to (2).

We conclude our brief excursion into the realm of approximation of  $|x|$  with a recent result of Blatt, Iserles, and Saff. They show [9] that the rational function of degree  $n$  of *best* uniform approximation to  $|x|$  on  $I$  has all its zeros and poles on the imaginary axis. Moreover, if  $r_n^*(x)$  is this best approximation, then

$$\lim r_n^*(z) = \begin{cases} z, & \operatorname{Re} z > 0, \\ -z, & \operatorname{Re} z < 0. \end{cases}$$

Note that the Blatt, Iserles, and Saff result is in the Chebyshevian stream, while Newman's result is in the Weierstrassian line.

Weierstrass' method of constructing good polynomial approximations to a function by convolving it with an appropriate kernel was pursued by Landau, and, most successfully, by the latter's student, Dunham Jackson. In his prize-winning Göttingen doctoral dissertation (1911) Jackson gave the definitive quantitative result on the error of best approximation of continuous functions by polynomials of given degree (cf. [10]). He proved that if  $f \in C(I)$ , then there exists a constant  $C (\leq 6)$  such that

$$E_n(f) \leq C \omega_f\left(\frac{1}{n}\right). \quad (4)$$

Here  $\omega_f(\delta)$  is the *modulus of continuity* of  $f$  on  $I$ , i.e., for  $\delta > 0$ ,

$$\omega_f(\delta) = \sup\{|f(x) - f(y)| : x, y \in I, |x - y| \leq \delta\},$$

a notion introduced by Jackson in his dissertation. Observe that if  $f(x) = |x|$ , so that  $\omega_f(\delta) = \delta$ , (2) shows that Jackson's Theorem is sharp in relating  $E_n(f)$  to  $n$ .

This seems to be the appropriate point to interject Bourbaki's view of the place of Approximation Theory in the world of mathematics. What follows is the only direct reference by that author to this area that I can find. "On the other hand, Weierstrass himself discovered the possibility of uniform approximation of a continuous real-valued function in one or more variables on a bounded set by polynomials. This result immediately aroused lively interest and led to many 'quantitative' studies (\*). Footnote. See e.g. C. De La Vallée Poussin, *Leçons sur l'approximation des fonctions d'une variable réelle*, Paris (Gauthier-Villars), 1919" (see [11]). We turn next to more recent developments and generalizations.

## Later developments

### • Approximation of classes of functions

We have seen that for functions belonging to a given class of functions, upper bounds for  $E_n(f)$  can be obtained. These bounds involve constants which depend on the particular  $f$

being approximated, and best constants for a given  $f$  can rarely be determined. It seems plausible, then, to try to find the best constant that will work for all functions in the class. We formulate this problem in great generality, but will investigate only a special case.

Let  $X$  be a Banach space,  $K$  a given subset of  $X$ , and  $V_n$  an  $n$ -dimensional subspace of  $X$ . Then we consider

$$E(K, V_n) = \sup_{f \in K} \inf_{v \in V_n} \|f - v\|, \quad (5)$$

that is, we seek the worst function in  $K$  with respect to approximation out of  $V_n$ . For example, suppose  $X = C^*$ , the  $2\pi$ -periodic continuous functions on  $\mathbb{R}$ , and  $K = \{f \in C^*: f^{(k-1)} \text{ absolutely continuous, } |f^{(k)}(x)| \leq 1, \text{ a.e., } k \geq 1\}$ . Then if  $\mathcal{T}_n$  is the set of trigonometric polynomials of degree  $n$ , we can state the theorem of Favard (cf. [12] for details),

$$E(K, \mathcal{T}_{n-1}) = \frac{M_k}{n^k}, \quad n = 1, 2, \dots,$$

where

$$M_k = \begin{cases} \frac{4}{\pi} \sum_{j=0}^{\infty} \frac{1}{(2j+1)^{k+1}}, & k \text{ odd,} \\ \frac{4}{\pi} \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)^{k+1}}, & k \text{ even.} \end{cases}$$

Moreover, there exists a worst function for each  $k$ ,  $n = 1, 2, \dots$ , and  $M_k \leq M_1 = \pi/2$  for  $k = 1, 2, \dots$ .

In the setting of (5) we may, following Kolmogorov, seek the best  $n$ -dimensional subspace of  $X$  for approximating elements of  $K$ , thus being led to

$$d_n(K) = \inf_{V_n} E(K, V_n),$$

the  $n$ -width of  $K$  in  $X$ . The subject of  $n$ -widths now has a large literature for which the reader should consult [13]. In the examples we discussed above, Tihomirov (cf. [12]) showed that

$$d_{2n-1}(K) = \frac{M_k}{n^k}$$

and  $\mathcal{T}_{n-1}$  is the best approximating subspace.

#### • Approximation of functionals

Chebyshev's setting for polynomial approximation of continuous functions may be mimicked to approximate a linear functional by linear combinations of given linear functionals. More precisely, suppose  $L, L_1, \dots, L_n$  are linear functionals on  $K \subset C(I)$ . Consider the problem of finding

$$\inf_a \sup_{f \in K} |Lf - \sum_{i=1}^n a_i L_i f| =: \mathcal{E}(L, L_1, \dots, L_n), \quad (6)$$

where  $a := (a_1, \dots, a_n)$ . As a typical example define  $L, L_i$ ,  $i = 1, \dots, n$ , by

$$Lf = \int_{-1}^1 f w, \quad L_i f = f(x_i), \quad i = 1, \dots, n, \quad (7)$$

where  $w$  is a (non-negative) "weight function" on  $I$  and  $x^{(n)} := (x_1, \dots, x_n)$  is a given set of distinct points of  $I$ . The problem is now the old and important one of *numerical integration* (or *quadrature*), and we shall restrict our attention to this example, retaining the notation given by (7) and writing the *quadrature formula* and its *error* as, respectively,

$$Q_n(a, x^{(n)})f = \sum_{i=1}^n a_i f(x_i)$$

and

$$\mathcal{E}(x^{(n)}, w) = \mathcal{E}(L, L_1, \dots, L_n).$$

A striking 19th-century result, associated with the names of Gauss and Jacobi, is that if  $p_0, p_1, \dots, p_n, \dots$  are polynomials (of degree equal to their index) orthogonal with respect to a given  $w$  and  $x^{(n)} = \xi^{(n)}$  is the set of zeros of  $p_n$ , then there are numbers  $a = \lambda(\xi^{(n)})$  such that  $Lf = Q_n(a, \xi^{(n)})f$  for all  $f \in P_{2n-1}$ ,  $n = 1, 2, \dots$ . This result may be expressed in our setting, fixing  $n$  and  $w$  and putting  $K = P_{2n-1}$ , as

$$0 = \mathcal{E}(\xi^{(n)}, w) = \inf_{x^{(n)}} \mathcal{E}(x^{(n)}, w),$$

the infimum in  $\mathcal{E}(\xi^{(n)}, w)$  being attained for  $a = \lambda$ .

Let us also remark that given  $w$ ,  $\epsilon > 0$  and putting  $K = C(I)$  there exist  $n$  and  $Q_n(a, x^{(n)})$  such that

$$\sup_{f \in K} |Lf - Q_n(a, x^{(n)})f| < \epsilon.$$

Indeed, we may choose  $x^{(n)} = \xi^{(n)}$  and  $a = \lambda$  as above (cf. [14]), thus obtaining an analog of the Weierstrass theorem.

An interesting quadrature illustration of (6) is due to A. Sard [15]. Given  $x^{(n)}$  he chooses

$$K = \{f \in C^{(k-1)}(I) : f^{(k-1)} \text{ absolutely cont., } \|f^{(k)}\|_2 \leq 1\}, \quad (8)$$

where  $0 < k \leq n$ . If we put  $Rf = Lf - Q_n(a, x^{(n)})f$ , Sard's problem is

$$\inf_a \sup_{\substack{f \in K \\ Rf|_{P_{k-1}} = 0}} |Rf|. \quad (9)$$

Note that he, reasonably, restricts the admissible  $a$  by requiring that the quadrature formula be exact for polynomials of degree at most  $k-1$ . I say "reasonably," because if  $p \in P_{k-1}$  and  $Rp \neq 0$ , then since  $cp \in K$  for any constant  $c$  and  $R$  is linear,

$$\sup_{f \in K} |Rf| = \infty.$$

Sard's solution, in modern terminology, uses the unique spline of order  $2k$  [i.e., piecewise polynomial of degree  $2k-1$  in each  $(x_i, x_{i+1})$ ,  $i = 1, \dots, n-1$  and of

degree  $k - 1$  outside  $(x_1, x_n)$  which is in  $C^{(2k-2)}(I)$ ,  $s_f$ , satisfying

$$s_f(x_i) = f(x_i), \quad i = 1, \dots, n.$$

Then

$$Ls_f =: \sum_{i=1}^n a_i^* f(x_i)$$

defines the optimal choice of  $a$ . In other words, the optimal quadrature formula  $Q_n(a^*, x^{(n)})$  is obtained by integrating the interpolating natural spline at the nodes. In particular when  $k = n$  we recover the classical Newton-Cotes quadrature formula, since the interpolating spline becomes a polynomial of degree at most  $n - 1$ .

Since we have mentioned splines let us say a few more words about this class of functions which have become so popular as approximators in recent years (cf. [16]). Splines are simply piecewise functions, the pieces generally being polynomials of some fixed degree. A familiar example of a spline is a continuous piecewise linear function. Piecewise functions were widely used before the word spline was brought into mathematics by Schoenberg. A spline is, originally, a thin strip which is grooved and pinned at points by a set of heavy weights (called "ducks") with arms fitting in the groove. It is used to draw a smooth curve through the points. Schoenberg observed that natural "splines" of order 4 with knots (= points at which pieces are joined) at the abscissas of the points, which interpolated the given values at the knots, were a reasonable mathematical model of the material spline. Mathematical splines came into heavy use after 1960 in conjunction with the explosive growth of computers. By now the theory and practice of approximation by univariate splines has reached a high level of development (cf. [17]) and the notion of multivariate splines is at the frontier of research in Approximation Theory (cf. [18]).

### Optimal recovery

A comparison of (5) and (6) is quite instructive. In (5) there is the tacit assumption that each  $f$  is known and its best approximation is sought. But in (6) all we know is that  $f$  is an element of  $K$  and the limited information about it provided by  $L_1, \dots, L_n$ . It is this latter point of view, estimation from incomplete information, that we wish to pursue. Such problems had already received attention (cf. [19] and [20]) by the mid-70s, at which time Micchelli and Rivlin [21] gave a precise formulation of the notion of optimal estimation which they called "optimal recovery." By this they meant estimating (= recovering) some required feature of a function from limited, and possibly error-contaminated, information about it as effectively as possible. From another point of view, what they presented was a quantification of the concept of "complexity" as applied to many areas of numerical approximation. This approach has

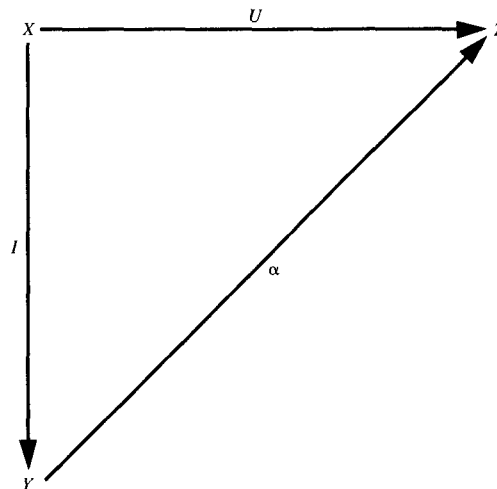


Figure 1

Schematic of optimal recovery.

a large Soviet literature. It is also the subject of the books [22, 23] and the lecture notes [24] and [25].

Our formulation of the optimal recovery problem is the following. Let  $X$  be a linear space,  $Y$  and  $Z$  normed linear spaces.  $K$  is a subset of  $X$ ,  $U$  a linear operator from  $X$  into  $Z$ , and  $I$  a linear operator (the information operator) from  $X$  into  $Y$ . We wish to recover  $Ux$  from information close to  $Ix$ , that is, for each  $x \in K$  we know a  $y \in Y$  satisfying  $\|Ix - y\| \leq \epsilon$ , for some fixed  $\epsilon \geq 0$ . We call any function  $\alpha: IK + \epsilon S \rightarrow Z$  ( $S = \{y \in Y: \|y\| \leq 1\}$ ) an *algorithm*. Each algorithm,  $\alpha$ , produces an error

$$E_\alpha(K, \epsilon) = \sup_{\substack{x \in K \\ \|Ix - y\| \leq \epsilon}} \|Ux - \alpha y\|,$$

and

$$E(K, \epsilon) = \inf_\alpha E_\alpha(K, \epsilon) \quad (10)$$

is the *intrinsic error* in the recovery problem. If  $E_{\alpha^*}(K, \epsilon) = E(K, \epsilon)$ , we say that  $\alpha^*$  is an *optimal algorithm* and effects the *optimal recovery* of  $Ux$ . Figure 1 gives a brief outline of the setting.

Each problem requires specification of  $X, Y, Z, K, I, U$ , and  $\epsilon$ . For instance, if we choose  $X$  to be the space of functions having  $k - 1$  ( $k \geq 1$ ) continuous derivatives on  $[-1, 1]$  and a  $k$ th derivative in  $L^2[-1, 1]$ ;  $K$  as defined in (8),  $Y = \mathbb{R}^n$ ,  $Z = \mathbb{R}$ ,  $I$  defined by  $If = (f(x_1), \dots, f(x_n))$  for some given  $x^{(n)} (n \geq k)$ ,  $U = \int_{-1}^1 f w$ , and  $\epsilon = 0$ , then we have the problem of Sard again. But note that the admissible

algorithms need not be linear, *a priori*, and the second restriction in (9) is not present. It can be shown [21] that in this more general setting Sard's solution remains optimal. The determination of the intrinsic error is a difficult problem (see [25] for more information).

One general approach to an optimal recovery problem is to obtain first the intrinsic error, or at least a lower bound for the intrinsic error. If  $K$  is a balanced (i.e., contains  $-x$  if it contains  $x$ ) convex set, then it is easy to see (cf. [21]) that

$$E(K, \varepsilon) \geq \sup_{\substack{x \in K \\ \|x\| \leq \varepsilon}} \|Ux\| =: e(K, \varepsilon). \quad (11)$$

Moreover, in the important case that  $U$  is a linear functional, not only does (11) become an equality [thus providing us with a "dual" extremal problem to (10)] but we are assured, under minor additional assumptions about  $K$ , that there exists a linear optimal algorithm (details are given in [21]). Observe that the extremal problem on the right-hand side of the inequality (11) is of a more familiar kind than (10). The second step in this strategy is to find an algorithm,  $\alpha$ , that satisfies  $E_\alpha(K, \varepsilon) \leq e(K, \varepsilon)$ , thus obtaining an optimal algorithm and the intrinsic error. In general, of course, "finding" the  $\alpha$  is the nub of the problem and is an art. There are, however, settings, for example when  $X$  is a Hilbert space,  $K$  its unit ball, the kernel of  $I$  is closed, and  $\varepsilon = 0$ , in which the optimal algorithms can be obtained in a straightforward fashion. Details are given in [21] and [25].

We turn now to some problems in which  $X = H^\infty$ , the functions analytic and bounded in  $D$ :  $|z| < 1$ .  $H^\infty$  is a normed linear space, the norm being given by

$$\|f\| = \sup_{z \in D} |f(z)|.$$

Suppose  $z_1, \dots, z_n \in D$  are given. Let

$$K = \{f \in H^\infty : \|f\| \leq 1\}$$

and  $If = (f(z_1), \dots, f(z_n))$  (with the convention that if some of the points coincide the corresponding function values are replaced by consecutive derivatives in the obvious way, so that  $Y = \mathcal{C}^n$ ). Suppose  $\zeta \in D$  is given,  $Uf = f(\zeta)$ ,  $Z = \mathcal{C}$ , and  $\varepsilon = 0$ . We are thus considering the problem of optimal interpolation in  $H^\infty$ .

Put

$$B_n(z) = \prod_{i=1}^n \frac{z - z_i}{1 - \bar{z}_i z}.$$

Since  $B_n(z_i) = 0$ ,  $i = 1, \dots, n$ , (11) yields

$$E(K, 0) \geq \sup_{\substack{f \in K \\ f(z_i)=0, i=1, \dots, n}} |f(\zeta)| \geq |B_n(\zeta)|. \quad (12)$$

Indeed, in view of the remark following (11), equality holds in (12), but this fact is not needed in what follows.

Next define  $a_j(\zeta)$ ,  $j = 1, \dots, n$  by

$$\begin{aligned} f(\zeta) - \sum_{j=1}^n a_j(\zeta) f(z_j) \\ = \frac{1}{2\pi i} \int_{\partial D} \frac{B_n(\zeta)}{B_n(z)} \frac{1 - |\zeta|^2}{1 - \bar{\zeta}z} \frac{1}{z - \zeta} f(z) dz, \end{aligned} \quad (13)$$

using the calculus of residues. The absolute value of the right-hand side of (13) is bounded by

$$\begin{aligned} \frac{1}{2\pi} \int_0^{2\pi} |B_n(\zeta)| \frac{1 - |\zeta|^2}{|1 - \bar{\zeta}e^{i\phi}|} \frac{1}{|e^{i\phi} - \zeta|} d\phi \\ = |B_n(\zeta)| \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - |\zeta|^2}{|e^{i\phi} - \zeta|^2} d\phi \\ = |B_n(\zeta)|. \end{aligned}$$

The last integral on the right-hand side, being the integral of the Poisson kernel, is 1. Thus, for any  $f \in K$  we have

$$|f(\zeta) - \sum_{j=1}^n a_j(\zeta) f(z_j)| \leq |B_n(\zeta)|.$$

If we define the algorithm  $\alpha^*$  by

$$\alpha^*: (f(z_1), \dots, f(z_n)) \rightarrow \sum_{j=1}^n a_j(\zeta) f(z_j),$$

then  $E_{\alpha^*}(K, 0) \leq |B_n(\zeta)|$ , and in view of (13),  $\alpha^*$  is an optimal algorithm for interpolation. Note that it is a linear optimal algorithm, whose existence the general theory had affirmed. We add the remark that the case  $\varepsilon > 0$  has been studied by Osipenko [26] and Dita [27]. The choice of (13) may seem to be akin to pulling a rabbit out of the air. However, the left-hand side is a linear functional on  $H^\infty$  and the theory of extremal problems for such linear functionals represented by rational kernels is well-known (cf. [28]) and leads to (13).

As a final example we turn to the problem of optimal numerical integration in  $H^\infty$ . All specifications in the optimal recovery problem are the same as in the interpolation problem except that

$$Uf = \int_{-1}^1 f(t) w(t) dt,$$

where  $w(t)$  is a given weight-function. We have equality in (11) and obtain for the intrinsic error

$$\begin{aligned} E(K, 0, w) &= \sup_{\substack{f \in K \\ f(z_i)=0, i=1, \dots, n}} \left| \int_{-1}^1 f w \right| \\ &= \sup_{f \in K} \left| \int_{-1}^1 f B_n w \right|. \end{aligned} \quad (14)$$

In our notation for the intrinsic error we have now added its dependence on  $w$ . In the context of the numerical integration problem it is natural to seek the choice of the nodes  $z_1, \dots, z_n$  which minimizes the intrinsic error, i.e., what might be called optimal sampling. It is known that the minimal intrinsic error satisfies

$$b_1 e^{-b_2 \sqrt{n}} \leq \inf_{(z_1, \dots, z_n)} E(K, 0, 1) \leq c_1 e^{-c_2 \sqrt{n}},$$

$b_1, b_2, c_1, c_2 > 0$ . For details and generalizations see [29–31].

Assume that the weight function satisfies

$$m(\mu) \leq w(t)(1-t^2)^{\frac{1-\mu}{2}} \leq M(\mu), \quad -1 < t < 1, \mu > -1$$

for  $m, M > 0$ . Then we can present some interesting error estimates for the intrinsic error,  $E(K, 0, w, z^{(n)})$ , where the dependence of  $E$  on the nodes  $z^{(n)} := (z_1, \dots, z_n)$  is now made explicit in the notation. In particular, we put  $z^{(n)} = E^{(n)}$  when  $z_j = -1 + (2j)/(n+1)$ ,  $j = 1, \dots, n$  (equally spaced points), and  $z^{(n)} = z^{(n)}(\lambda)$  when the  $z_j$  are the zeros of  $P_n^{(\lambda)}$ , the ultraspherical orthogonal polynomials with respect to the weight function  $(1-x^2)^{\lambda-1/2}$ ,  $\lambda \geq 0$ , e.g., Legendre zeros ( $\lambda = 1/2$ ), Chebyshev zeros ( $\lambda = 0$ ), 1).

In view of (14) we have

$$\int_{-1}^1 B_n^2(t) w(t) dt \leq E(K, 0, w, z^{(n)}) \leq \int_{-1}^1 |B_n(t)| w(t) dt$$

for any choice of real nodes. It was shown in [32] that there are positive constants  $A(\mu)$ ,  $B(\mu)$ ,  $C(\mu, \lambda)$ , and  $D(\mu, \lambda)$  such that

$$A(\mu) \leq E(K, 0, w, E^{(n)}) (n \log n)^{\frac{1+\mu}{2}} \leq B(\mu) \quad (15)$$

and

$$C(\mu, \lambda) \leq E(K, 0, w, z^{(n)}(\lambda)) n^{1+\mu} \leq D(\mu, \lambda). \quad (16)$$

The proofs rely on appropriate pointwise bounds on  $B_n(t)$  from above and below. We remark that while (15) and (16) give the correct order of dependence of the intrinsic error on  $n$ , our invocation of the duality (14) has wiped out the role of the quadrature formula and we have no information about an optimal algorithm.

## References

1. A. A. Youshkevitch, *Dictionary of Scientific Biography*, Vol. IX, C. C. Gillispie, Ed., Scribner's, New York, 1974, p. 125.
2. T. J. Rivlin, *The Chebyshev Polynomials*, John Wiley, New York, 1974.
3. P. L. Chebyshev, *Oeuvres*, Vol. I, Chelsea, New York, 1961, p. 284.
4. H. Lebesgue, "Sur l'Approximation des Fonctions," *Bull. Soc. Math.* **22**, 278–287 (1898).
5. S. N. Bernstein, "Sur l'Ordre de la Meilleure Approximation des Fonctions Continues par des Polynomes," *Acad. Royale Belg., Classe des Sciences*, Ser. II, 4, 1912.
6. S. N. Bernstein, "Sur la Meilleure Approximation de  $|x|$  par des Polynomes de Degrés Donnés," *Acta Math.* **37**, 1–57 (1913).
7. R. S. Varga and A. J. Carpenter, "On the Bernstein Conjecture in Approximation Theory," *Constr. Approx.* **1**, 333–348 (1985).
8. D. J. Newman, "Rational Approximation to  $|x|$ ," *Michigan Math. J.* **11**, 11–14 (1964).
9. H.-P. Blatt, A. Iserles, and E. B. Saff, "Remarks on the Behavior of Zeros of Best Approximating Polynomials and Rational Functions," *Proceedings of the IMA Conference on Algorithms for the Approximation of Functions and Data*, July 1985; to appear.

10. D. Jackson, *The Theory of Approximation*, Vol. 11, Amer. Math. Soc., Colloquium Publications, New York, 1930.
11. N. Bourbaki, *Elements of Mathematics, General Topology, Part 2*, Hermann, Paris, 1966, Ch. X, p. 348.
12. G. G. Lorentz, *Approximation of Functions*, Holt, Rinehart and Winston, New York, 1966.
13. A. Pinkus, *n-Widths in Approximation Theory*, Springer-Verlag, Berlin, 1985.
14. G. Szegő, *Orthogonal Polynomials*, Amer. Math. Soc. Colloq. Publ. 23, Providence, RI, 1975.
15. A. Sard, "Best Approximate Integration Formulas; Best Approximation Formulas," *Amer. J. Math.* **71**, 80–91 (1949).
16. L. L. Schumaker, *Spline Functions: Basic Theory*, Wiley, New York, 1981.
17. C. de Boor, *A Practical Guide to Splines*, Applied Math. Sciences Vol. 27, Springer-Verlag, New York, 1978.
18. W. Dahmen and C. A. Micchelli, "Recent Progress in Multivariate Splines," *Approximation Theory IV*, C. K. Chui et al., Eds., Academic Press, Inc., New York, 1983.
19. M. Golomb and H. F. Weinberger, "Optimal Approximation and Error Bounds," *On Numerical Approximation*, R. Langer, Ed., University of Wisconsin Press, Madison, WI, 1959, pp. 117–190.
20. M. Golomb, *Lectures on Theory of Approximation*, Applied Math. Division, Argonne National Laboratory, Argonne, IL, 1962.
21. C. A. Micchelli and T. J. Rivlin, "A Survey of Optimal Recovery," *Optimal Estimation in Approximation Theory*, C. A. Micchelli and T. J. Rivlin, Eds., Plenum Publishing Co., New York, 1977, pp. 1–54.
22. J. F. Traub and H. Wozniakowski, *A General Theory of Optimal Algorithms*, Academic Press, Inc., New York, 1980.
23. J. F. Traub, G. W. Wasilkowski, and H. Wozniakowski, *Information, Uncertainty, Complexity*, Addison-Wesley Publishing Co., Reading, MA, 1983.
24. T. J. Rivlin, "The Optimal Recovery of Functions," *Contemp. Math.* **9**, 121–151 (1982).
25. C. A. Micchelli and T. J. Rivlin, "Lectures on Optimal Recovery," *Lecture Notes in Mathematics No. 1129, Proceedings: Numerical Analysis 1984*, Springer-Verlag, Berlin, 1985, pp. 21–93.
26. K. Yu. Osipenko, "Best Methods of Approximating Analytic Functions Given with an Error," *Math. U.S.S.R. Sbornik* **46**, 353–374 (1983).
27. P. Dita, "Optimal Analytic Extrapolation Revisited," *J. Phys. A* **17**, 957–968 (1984).
28. P. Duren, *Theory of  $H^p$  Spaces*, Academic Press, Inc., New York, 1970.
29. B. D. Bojanov, "Best Quadrature Formula for a Certain Class of Analytic Functions," *Zastos. Mat.* **XIV**, 441–447 (1974).
30. H. L. Loeb, "A Note on Optimal Integration in  $H_\infty$ ," *C.R. Akad. Bulgare Sci.* **27**, 615–618 (1974).
31. D. J. Newman, "Quadrature in  $H^p$ , Lectures III, IV," *Approximation with Rational Functions*, CBMS Regional Conference Series in Mathematics No. 41, Amer. Math. Soc., Providence, RI, 1979.
32. N. Dyn, C. A. Micchelli, and T. J. Rivlin, "Blaschke Products and Optimal Recovery in  $H^*$ ," *CALCOLO*; to appear.

Received August 8, 1986; accepted for publication October 28, 1986

**Theodore J. Rivlin** IBM Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598. Dr. Rivlin is a Research staff member and senior consultant in the Mathematical Sciences Department of the Research Division at Yorktown Heights.

He has been with IBM since 1959. He received the Ph.D. degree in mathematics from Harvard University, Cambridge, Massachusetts, in 1953; was instructor in mathematics at The Johns Hopkins University, Baltimore, Maryland, from 1952 to 1955; and was research assistant in the Institute for Mathematics Sciences, New York University, New York, in 1955-56. From 1956 to 1959 he was senior mathematical analyst in a research group at Fairchild Engine and Airplane Company, Deer Park, New York. He was Visiting Professor of Computer Science at Stanford University, California, for the academic year 1969-70, and was Adjunct Professor of Mathematics at the Graduate Center, City University of New York, from 1966-1976. In the academic year 1976-77 he was Science Research Council Senior Research Fellow at the Department of Mathematics, Imperial College of Science and Technology, University of London. Dr. Rivlin's major interest is Approximation Theory, and he is the author of the books *An Introduction to the Approximation of Functions* (1969) and *The Chebyshev Polynomials* (1974), co-editor of *Optimal Estimation in Approximation Theory* (1977), and an Associate Editor of *The Journal of Approximation Theory*.