JOURNAL OF APPROXIMATION THEORY 7, 239–255 (1973)

Uniform Approximation of Functions Through Partitioning*

ROBERT MICHAEL HARRIS

Applied Physics Section, Aerospace Corp., Box 95085, Los Angeles, California 90045

Communicated by E. W. Cheney

Received November 2, 1970

It is the purpose of this paper to present a method for the computation of best uniform approximation, through the replacement of the uniform norm by another norm designated by $\|\cdot\|_n$ and also by two pseudonorms designated by $ps\|\cdot\|_n$ and $(A)\|\cdot\|_n$. Existence, uniqueness, characterization and computation of best approximations for $\|\cdot\|_n$, $ps\|\cdot\|_n$, and $(A)\|\cdot\|_n$ are examined.

1. INTRODUCTION

Let X be a compact topological space and let $C(X)$ denote the space of continuous real-valued functions on X. Let $\|\cdot\|$ be the uniform norm defined on $C(X)$ by $\|f\| = \sup\{|f(x)|: x \in X\}$. Finally, let Φ be the m-dimensional linear subspace of $C(X)$ generated by some m fixed linearly independent functions ϕ_i, $i = 1,\ldots, m$.

The linear approximation problem can be stated as follows. Given $f \in C(X)$, find the elements $\phi^* \in \Phi$ such that

$$\|f - \phi^*\| = \inf\{\|f - \phi\|: \phi \in \Phi\}.$$

Such a ϕ^* is called a best approximation to f on X.

The problem is then usually broken into four parts:

(i) Do such elements ϕ^* exist?

(ii) Characterize the ϕ^*.

(iii) Establish uniqueness or nonuniqueness of best approximations.

(iv) Compute a best or a good approximation. (Even though best approximations may not be computable, "good" approximations may be.)

* This paper is part of the author's Ph.D. Thesis at the University of Utah.

DEFINITION 1. Φ is said to be a Haar space (Tchebyshev space) if no nonzero $\phi \in \Phi$ has more than $m - 1$ zeros.

For univariate approximation, if Φ is a Haar space, the Remez Algorithm works nicely in obtaining the best approximation. For multivariate approximation, best approximations are not necessarily unique, and the Remez Algorithm does not work.

This paper presents an alternative to existing methods of multivariate approximation. In Section 2, I approximate the L_∞ or uniform norm, which is designated by $\|\cdot\|$ throughout the paper, by a norm I designate by $\|\cdot\|_n$. In Sections 3 and 4 I approximate the L_∞ norm by two pseudonorms designated by $ps\|\cdot\|_n$ and $(A)\|\cdot\|_n$. Computational results are given in Section 5.

2. n-NORMS

The following norm was motivated by the search for a form of approximation which

(i) is nearly a best uniform approximation,

(ii) maintains the finite nature of the method of discretization while making use of information about the given function f on all of the space in question.

Let X be a compact measurable metric space with positive measure μ. Let U_n be a partition of X into n sets $\{E_i^n\}_{i=1}^n$ such that $\mu E_i^n \neq 0$, $i = 1,\ldots, n$. Assume for simplicity that $\mu(E_i^n \cap E_j^n) = 0$ for $i \neq j$. For $f \in C(X)$, define

$$\|f\|_n = \max_{1 \leqslant i \leqslant n} \left[\left(\frac{1}{\mu E_i^n} \int_{E_i^n} f^2 \, d\mu\right)^{1/2}\right].$$

It is easy to verify that $\|\cdot\|_n$ is a norm. Frequently the norm $\|\cdot\|_n$ will be called the "n-norm."

Let $\{\phi_1,\ldots, \phi_m\}$ be a set of linearly independent functions in $C(X)$. Define $\mathscr{P}_m = \{P(A, x) = \sum_{i=1}^m a_i\phi_i(x)$ where $A = (a_1,\ldots, a_m) \in R^m\}$. $P(A, x)$ will be referred to as a generalized polynomial.

THEOREM 1. *Given $f \in C(X)$, there exists a best approximation to f from the class of generalized polynomials in the norm* $\|\cdot\|_n$.

This is a corollary to the theorem which states: A finite-dimensional subspace of a normed linear space contains at least one point of minimum distance from a fixed point. (See Cheney [1, p. 20]).

THEOREM 2. *The best $\|\cdot\|_n$ approximation to f from $\mathscr{P}_m$ is unique.*

Proof. Let

$$r_k^n(A) = \left[\frac{1}{\mu E_k^n} \int_{E_k^n} (f(x) - P(A, x))^2 \, d\mu\right]^{1/2}, \qquad k = 1, \ldots, n.$$

Note. The superscript n will be deleted on $r_k^n(A)$ when no ambiguity arises from doing so.

If $f \in \mathscr{P}_m$, then the result is obvious since the ϕ_i's were assumed to be linearly independent. For $f \notin \mathscr{P}_m$ each $r_k(A)$ is strictly convex.

Let $\delta(A) = \max_{1 \leqslant k \leqslant n} r_k(A)$. Then $\delta(A)$ is strictly convex hence has a unique minimum.

CHARACTERIZATION THEOREM. *Let $r_k(A)$ and $\delta(A)$ be as above. Then A is a minimum for δ if and only if*

$$\Theta \in H\left[\sum_{i=1}^{m} a_i A_i^k + B^k : r_k(A) = \delta(A)\right].$$

Θ denotes the origin of R^m. $H[\ ]$ denotes the "convex hull of" and

$$A_i^k = [A_{i1}^k, A_{i2}^k, \ldots, A_{im}^k]$$

$$B^k = [B_1^k, B_2^k, \ldots, B_m^k]$$

where

$$A_{ij}^k = \frac{2}{\mu E_k^n} \int_{E_k^n} \phi_i \phi_j \, d\mu,$$

$$B_i^k = \frac{-2}{\mu E_i^n} \int_{E_k^n} f \cdot \phi_i \, d\mu.$$

Proof (necessity). Let

$$C^k = \frac{1}{\mu E_k^n} \int_{E_k^n} f^2 \, d\mu.$$

Then

$$r_k(A) = \left[\sum_{i,j=1}^{m} \tfrac{1}{2} A_{ij}^k a_i a_j + \sum_{i=1}^{m} B_i^k a_i + C^k\right]^{1/2}.$$

Suppose $\Theta \notin H[\sum_{i=1}^m a_i A_i^k + B^k : r_k(A) = \delta(A)]$. Then by the theorem on linear inequalities (see Cheney [1, p. 19]), there is an h such that

$$\left\langle \sum_{i=1}^{m} a_i A_i^k + B^k, h \right\rangle > 0 \qquad \text{for} \quad k \in M = \{i : r_i(A) = \delta(A)\}.$$

Let $\alpha = \min_{k \in M} \langle \sum a_i A_i^k + B^k, h \rangle$. We have $\alpha > 0$. For $k \in M$ and $\lambda > 0$, look at

$$\begin{aligned}
[r_k(A - \lambda h)]^2 &= \sum_{i,j=1}^{m} \tfrac{1}{2} A_{ij}^k (a_i - \lambda h_i)(a_j - \lambda h_j) + \sum_{i=1}^{m} B_i^k (a_i - \lambda h_i) + C^k \\
&= [r_k(A)]^2 - \sum_{i,j=1}^{m} \tfrac{1}{2} A_{ij}^k (\lambda a_i h_j + \lambda a_j h_i - \lambda^2 h_i h_j) - \sum_{i=1}^{m} B_i^k (\lambda h_i) \\
&= [r_k(A)]^2 - \lambda \left\langle \sum_{i=1}^{m} a_i A_i^k + B^k, h \right\rangle + \frac{\lambda^2}{2} \sum_{i,j=1}^{m} A_{ij}^k h_i h_j \\
&\qquad\qquad \text{(noting that } A_{ij}^k = A_{ji}^k) \\
&\leqslant \delta^2(A) - \lambda\alpha + \frac{\lambda^2}{2} \sum_{i,j=1}^{m} A_{ij}^k h_i h_j \\
&< \delta^2(A)
\end{aligned}$$

for λ sufficiently small. Therefore $r_k(A - \lambda h) < \delta(A)$ for $k \in M$. For $k \notin M$, $r_k(A) < \delta(A)$ hence remains so in some neighborhood of A. Therefore, there is a λ such that $r_k(A - \lambda h) < \delta(A)$ for each k.

(*Sufficiency*). Suppose A is not a minimum. Then there is an h such that $\delta(A - h) < \delta(A)$. As noted in the proof of uniqueness, δ is convex. Thus

$$\begin{aligned}
\delta(A - \lambda h) &= \delta((1 - \lambda)A + \lambda(A - h)) \leqslant (1 - \lambda)\,\delta(A) + \lambda\delta(A - h) \\
&= \delta(A) - \lambda(\delta(A) - \delta(A - h)) < \delta(A) \qquad \text{for} \quad 0 < \lambda < 1.
\end{aligned}$$

As a result $r_k(A - \lambda h) \leqslant \delta(A - \lambda h) < \delta(A) = r_k(A)$ for $k \in M$. Written out, this is

$$\sum_{i,j=1}^{m} \tfrac{1}{2} A_{ij}^k (a_i - \lambda h_i) \cdot (a_j - \lambda h_j)$$

$$+ \sum_{i=1}^{m} B_i^k (a_i - \lambda h_i) + C^k < \sum_{i,j=1}^{m} \tfrac{1}{2} A_{ij}^k a_i a_j + \sum_{i=1}^{m} B_i^k a_i + C^k.$$

This gives

$$\sum_{i,j=1}^{m} \tfrac{1}{2} A_{ij}^k (-a_i \lambda h_j - a_j \lambda h_j + \lambda^2 h_i h_j) + \sum_{i=1}^{m} B_i^k (-\lambda h_i) < 0.$$

After a change of sign and writing this in inner product notation, we obtain

$$\lambda \left\langle \sum_{i=1}^{m} a_i A_i^k + B^k, h \right\rangle + \frac{\lambda^2}{2} \sum_{i,j=1}^{m} A_{ij}^k h_i h_j > 0,$$

giving

$$\left\langle \sum_{i=1}^{m} a_i A_i^k + B^k, h \right\rangle \geqslant 0, \qquad k \in M.$$

It is easy to see that if this quantity is zero, then a small change in h will make it positive if it is assumed $\sum_{i=1}^{m} a_i A_i^k + B^k$ is not zero for $k \in M$. In this case, by the "theorem on linear inequalities,"

$$\Theta \notin H \left[\sum_{i=1}^{m} a_i A_i^k + B^k : k \in M \right].$$

(The following conclusion of the proof was suggested by the referee.)

If one of the vectors inside the brackets of $H[\cdots]$ were zero (implying $\Theta \in H[\cdots]$). Then we could have

$$\begin{aligned} 0 &= \sum_{i=1}^{m} a_i A_{ij}^k + B_j^k \\ &= \frac{2}{\mu E_k^n} \left[\sum_{i=1}^{m} \int_{E_k^n} a_i \phi_i \phi_j \, d\mu - \int_{E_k^n} f \phi_j \, d\mu \right] \\ &= \frac{2}{\mu E_k^n} \int_{E_k^n} \left\{ \sum_{i=1}^{m} a_i \phi_i - f \right\} \phi_j \, d\mu, \qquad j = 1,\ldots, m. \end{aligned}$$

This implies

$$r_k^n(A) \leqslant r_k^n(B) \qquad \text{for all} \quad B \in R^m,$$

since the above equations are the well-known normal equations for approximation in the least-squares sense.

Hence $\| P(A, \cdot) - f \|_n = r_k^n(A) \leqslant r_k^n(B) \leqslant \| P(B, \cdot) - f \|_n \; \forall B \in R^m$, i.e., A is a minimum for δ.

The following theorem gives the basic properties of $\| \cdot \|_n$. The first is monotonicity. The second is its convergence to the uniform norm. These two properties hold then with the following hypotheses.

THEOREM 3. *Let the diameter of* $E_i^n = \sup_{x, y \in E_i^n} d(x, y)$ *where* d *is the metric for* X *and let*

$$\delta(U_n) = \max_{1 \leqslant i \leqslant n} [\textit{diameter of } E_i^n].$$

Assume U_m *is a refinement of* U_n, $m > n$. *Then*

(i) $\|f\|_n \leqslant \|f\|_m \leqslant \|f\|$,

(ii) $\|f\| \leqslant \|f\|_n + \omega(\delta_n)$

where ω is the modulus of continuity of f on X and δ_n is an abbreviation for $\delta(U_n)$.

Proof. (i) It suffices to consider $m = n + 1$. Then without loss of generality $E_i^{n+1} = E_i^n$, $1 \leqslant i \leqslant n - 1$ and $E_n^{n+1} \cup E_{n+1}^{n+1} = E_n^n$. Thus $r_i^n = r_i^{n+1}$, $1 \leqslant i \leqslant n - 1$. Suppose

$$\frac{1}{\mu E_j^{n+1}} \int_{E_j^{n+1}} f^2 \, d\mu < \frac{1}{\mu E_n^n} \int_{E_n^n} f^2 \, d\mu \qquad \text{for} \quad j = n, n + 1.$$

Then

$$\begin{aligned}
\int_{E_n^n} f^2 \, d\mu &= \int_{E_n^{n+1}} f^2 \, d\mu + \int_{E_{n+1}^{n+1}} f^2 \\
&< \left[\frac{\mu E_n^{n+1} + \mu E_{n+1}^{n+1}}{\mu E_n^n}\right] \int_{E_n^n} f^2 \, d\mu \\
&= \int_{E_n^n} f^2 \, d\mu.
\end{aligned}$$

This is a contradiction.

Therefore $\|f\|_n \leqslant \|f\|_{n+1}$. And $\|f\|_n \leqslant \|f\|_m$ follows inductively by defining an appropriate sequence of refinements. Next,

$$\begin{aligned}
\|f\|_m &= \max_{1 \leqslant i \leqslant m} \left[\frac{1}{\mu E_i^m} \int_{E_i^m} f^2 \, d\mu\right]^{1/2} \\
&\leqslant \max_{1 \leqslant i \leqslant m} \left[\frac{1}{\mu E_i^m} \int_{E_i^m} \|f\|^2 \, d\mu\right]^{1/2} = \|f\|.
\end{aligned}$$

(ii) Pick $x_0 \in X$ so that $|f(x_0)| = \|f\|$. Now consider the partition U_n. x_0 lies in E_i^n for some i. For $x \in E_i^n$, $|f(x) - f(x_0)| \leqslant \omega(\delta_n)$. Hence $|f(x_0)| \leqslant |f(x)| + \omega(\delta_n)$ and we obtain

$$\begin{aligned}
\|f\| &= \left[\frac{1}{\mu E_i^n} \int_{E_i^n} |f(x_0)|^2 \, d\mu\right]^{1/2} \\
&\leqslant \left[\frac{1}{\mu E_i^n} \int_{E_i^n} [|f(x)| + \omega(\delta_n)]^2 \, d\mu\right]^{1/2} \\
&\leqslant \left[\frac{1}{\mu E_i^n} \int_{E_i^n} |f(x)|^2 \, d\mu\right]^{1/2} + \left[\frac{1}{\mu E_i^n} \int_{E_i^n} \omega^2(\delta_n) \, d\mu\right]^{1/2} \\
&\leqslant \|f\|_n + \omega(\delta_n).
\end{aligned}$$

COROLLARY. *If the sequence of refinements* $\{U_k\}_{k=1}^{\infty}$ *is such that* $\lim_{k\to\infty} \delta_k = 0$, *then*

$$\lim_{k\to\infty} \|f\|_k = \|f\| \qquad \textit{since} \quad \omega(\delta_k) \to 0 \quad \textit{as} \quad k \to \infty.$$

From now on, we assume the sequence of norms $\{\|\cdot\|_n\}$ corresponds to a sequence of refinements $\{U_n\}$ where $\delta_n \to 0$ as $n \to \infty$.

The next two theorems provide the development needed to prove Theorems 6, 7, 8, and 9 which are the important results of this paper. Theorems 4, 5, and 6 are analogous to results given by Cheney [1, pp. 85–87] in his discussion of the method of discretization. Theorem 4 gives a simple estimate on the growth of $\| P(A, \cdot)\|_n$ for fixed A, as n increases.

THEOREM 4. *To each* $\alpha > 1$, *there corresponds a* $\delta > 0$ *such that* $\| P(A, \cdot)\| < \alpha \| P(A, \cdot)\|_n < \alpha \| P(A, \cdot)\|$ *for all generalized polynomials* $P(A, x)$ *and for any partition* U_n *of* X *with* $\delta_n \leqslant \delta$.

Proof. The second inequality holds by Theorem 3(i). Next let $\sigma = \min \| P(A, \cdot)\|$ on the compact set defined by $\sum_{i=1}^{m} | a_i | = 1$. Since the ϕ_i's are linearly independent, $\sigma > 0$. Let

$$\Omega(\delta) = \max_{1\leqslant i\leqslant m} [\max_{d(x,y)\leqslant\delta} | \phi_i(x) - \phi_i(y)|].$$

$\Omega(\delta) \to 0$ as $\delta \to 0$, since ϕ_i is uniformly continuous on X, for each $i = 1, 2,..., m$. Pick $\delta > 0$ so that $\Omega(\delta) < \sigma$. And choose U_n so that $\delta_n \leqslant \delta$. Let $P(A, x)$ by any generalized polynomial. Let $x_0 \in X$ be such that $| P(A, x_0)| = \| P(A, \cdot)\|$. Then $x_0 \in E_i^n$ for some i. By the mean value theorem for integrals, pick $x \in E_i^n$ so that

$$| P(A, x)| = \left[\frac{1}{\mu E_i^n} \int_{E_i^n} P^2(A, u)\, d\mu\right]^{1/2}.$$

Then

$$\begin{aligned}\sigma \sum_{i=1}^{m} | a_i | &\leqslant \| P(A, \cdot)\| = | P(A, x_0)| \\ &\leqslant | P(A, x_0) - P(A, x)| + | P(A, x)| \\ &\leqslant \sum | a_i | \, | \phi_i(x_0) - \phi_i(x)| + \| P(A, \cdot)\|_n \\ &\leqslant \Omega(\delta_n) \sum | a_i | + \| P(A, \cdot)\|_n .\end{aligned}$$

Hence

$$\sum_{i=1}^{m} | a_i | \leqslant \left[\frac{\| P(A, \cdot)\|_n}{\sigma - \Omega(\delta)}\right]$$

and

$$\| P(A, \cdot)\| \leqslant \frac{\Omega(\delta) \| P(A, \cdot)\|_n}{\sigma - \Omega(\delta)} + \| P(A, \cdot)\|_n$$

$$= \| P(A, \cdot)\|_n \left[1 + \frac{\Omega(\delta)}{\sigma - \Omega(\delta)}\right].$$

Since $[1 + \Omega(\delta)/(\sigma - \Omega(\delta))] \to 1^+$ as $\delta \to 0$ and is independent of A, the result is established.

The following theorem shows how the n-norm of the error curve $f - P$ converges to the uniform norm of the error curve $f - P$.

THEOREM 5. *$\| f - P(A, \cdot)\|_n$ converges to $\| f - P(A, \cdot)\|$ as $\delta_n \to 0$ according to the inequality*

$$\| f - P(A, \cdot)\| < \| f - P(A, \cdot)\|_n + \omega(\delta_n) + \beta \| P(A, \cdot)\| \, \Omega(\delta_n),$$

where β is independent of f and P, ω is the modulus of continuity of f and Ω is as in Theorem 4.

Proof. Let σ be as in the previous theorem, i.e., $\sigma \sum | a_i | \leqslant \| P(A, \cdot)\|$. Let $x_0 \in X$ be such that $| f(x_0) - P(A, x_0)| = \| f - P(A, \cdot)\|$. For some k, $x_0 \in E_k{}^n$. By the mean value theorem for integrals, let $x \in E_k{}^n$ be such that

$$| f(x) - P(A, x)| = \left[\frac{1}{\mu E_k{}^n} \int_{E_k{}^n} (f(u) - P(A, u))^2 \, d\mu\right]^{1/2}.$$

Then

$$\begin{aligned} \| f - P(A, \cdot)\| &= | f(x_0) - P(A, x_0)| \\ &\leqslant | f(x_0) - f(x)| + | f(x) - P(A, x)| + | P(A, x) - P(A, x_0)| \\ &\leqslant \omega(\delta_n) + \| f - P(A, \cdot)\|_n + \textstyle\sum | a_i | \, | \phi_i(x) - \phi_i(x_0)|. \end{aligned}$$

But

$$\begin{aligned} \textstyle\sum | a_i | \, | \phi_i(x) - \phi_i(x_0)| &\leqslant \textstyle\sum | a_i | \max_i | \phi_i(x) - \phi_i(x_0)| \\ &\leqslant \sigma^{-1} \| P(A, \cdot)\| \, \Omega(\delta_n). \end{aligned}$$

Let $\beta = \sigma^{-1}$.

The next theorem shows that the uniform norm of the error curve $f - P(A_n, \cdot)$ converges to the uniform norm of the error curve $f - P(A, \cdot)$ where $P(A_n, \cdot)$ is the best n-norm approximation of f on X and $P(A, \cdot)$ is a best uniform approximation to f on X.

THEOREM 6. *If $P(A_n, x)$ is the best approximation to $f(x)$ in the norm $\|\cdot\|_n$ and $P(A, x)$ is a best approximation in $\|\cdot\|$, then $\|f - P(A_n, \cdot)\|$ converges to $\|f - P(A, \cdot)\|$ according to the estimate*

$$\|f - P(A_n, \cdot)\| - \|f - P(A, \cdot)\| \leqslant \omega(\delta_n) + 2\alpha\beta \|f\| \Omega(\delta_n)$$

where α, β, and Ω are as in previous theorems.

Proof.

$$\|f - P(A_n, \cdot)\|_n \leqslant^1 \|f - P(A, \cdot)\|_n \leqslant^2 \|f - P(A, \cdot)\| \leqslant^3 \|f - P(A_n, \cdot)\|$$

where 1 follows by definition of $P(A_n, \cdot)$, 2 follows by Theorem 3(i), and 3 by the definition of $P(A, \cdot)$. Therefore

$$\|f - P(A_n, \cdot)\| - \|f - P(A, \cdot)\| \leqslant \|f - P(A_n, \cdot)\| - \|f - P(A_n, \cdot)\|_n \leqslant \omega(\delta_n) + \beta \| P(A_n, \cdot)\| \Omega(\delta_n)$$

by Theorem 5.

$$\| P(A_n, \cdot)\|_n \leqslant \| P(A_n, \cdot) - f\|_n + \|f\|_n \leqslant \| 0 - f\|_n + \|f\|_n \leqslant 2 \|f\|.$$

Using Theorem 4, we obtain $\| P(A_n, \cdot)\| \leqslant \alpha \| P(A_n, \cdot)\|_n \leqslant 2\alpha \|f\|$. Therefore $\|f - P(A_n, \cdot)\| - \|f - P(A, \cdot)\| \leqslant \omega(\delta_n) + 2\alpha\beta \|f\| \Omega(\delta_n)$.

I obtain next an estimate on the total approximation process. The process is that of approximating $\|\cdot\|$ by $\|\cdot\|_n$ and approximating $f(x)$ by $P(A_n, x)$ in the norm $\|\cdot\|_n$.

THEOREM 7.

$$\|f - P(A, \cdot)\| - \|f - P(A_n, \cdot)\|_n \leqslant \omega(\delta_n) + \frac{2 \|f\| \Omega(\delta_n)}{\sigma - \Omega(\delta_n)}$$

where σ is as in the previous theorems. Moreover, $\|f - P(A_n, \cdot)\|_n$ converges monotonically to $\|f - P(A, \cdot)\|$.

Proof. Let W_n be the modulus of continuity of $f(x) - P(A_n, x)$. Since $\|f - P(A, \cdot)\| \leqslant \|f - P(A_n, \cdot)\|$ and by Theorem 3(ii) $\|f - P(A_n, \cdot)\| \leqslant \|f - P(A_n, \cdot)\|_n + W_n(\delta_n)$, we have

$$\|f - P(A, \cdot)\| \leqslant \|f - P(A_n, \cdot)\|_n \leqslant W_n(\delta_n).$$

Thus

$$\begin{aligned}
&\|f - P(A, \cdot)\| - \|f - P(A_n, \cdot)\|_n \\
&\quad \leqslant W_n(\delta_n) = \sup\{|f(x) - P(A_n, x) - [f(y) - P(A_n, y)]|: d(x, y) \leqslant \delta_n\} \\
&\quad \leqslant \omega(\delta_n) + \Omega(\delta_n) \sum | a_{n_i} |.
\end{aligned}$$

From Theorem 4, we again use the estimate

$$\sum_{i=1}^{m} | a_{n_i} | \leqslant \frac{\| P(A_n, \cdot)\|_n}{\sigma - \Omega(\delta_n)}$$

and since as shown in Theorem 6, $\| P(A_n, \cdot)\|_n \leqslant 2 \|f\|$, we obtain

$$\|f - P(A, \cdot)\| - \|f - P(A_n, \cdot)\|_n \leqslant \omega(\delta_n) + \frac{2 \|f\| \, \Omega(\delta_n)}{\sigma - \Omega(\delta_n)}.$$

The convergence is monotone since $\|f - P(A_n, \cdot)\|_n \leqslant^1 \|f - P(A_m, \cdot)\|_n \leqslant^2 \|f - P(A_m, \cdot)\|_m$ for $m > n$ where 1 follows by definition of $P(A_n, x)$ and 2 follows by Theorem 3(i).

Remark. The value of this theorem lies in the fact that in computing $P(A_n, x)$, $\|f - P(A_n, \cdot)\|_n$ is a good estimate for $\|f - P(A, \cdot)\|$. I make use of this fact in Section 5. (Actually no computation was performed for the n-norm, however, it will be shown that the above observation carries over to the pseudo norm $(A)\| \cdot \|_n$, yet to be defined, which was used in computation.)

The following theorem is a simple modification of a theorem proved by Weinstein [3].

THEOREM 8. *Let* $\rho = \inf\{\|f - P(A, \cdot)\|: A \in R^m\}$ *and* $\mathscr{A} = \{A \in R^m: \|f - P(A, x)\| = \rho\}$. *Given any* $\epsilon > 0$, *there exists a* $\delta = \delta(\epsilon) > 0$ *such that for any* U_n *with* $\delta_n \leqslant \delta$, *then* $\sigma(A_n, \mathscr{A}) = \inf\{\sigma(A_n, A): A \in \mathscr{A}\} < \epsilon$ *where* σ *denotes the usual Euclidean metric on* R^m.

Proof. By Theorem 6, $\|f - P(A_n, \cdot)\| \leqslant \rho + K(\delta_n)$ where $K(\delta_n) \to 0$ as $\delta_n \to 0$. Suppose there is a subsequence $\{A_{n_i}\}$ such that $\sigma(A_{n_i}, \mathscr{A}) \geqslant \epsilon$ for each i.

Since $\| P(A_{n_i}, \cdot)\| \leqslant \|f\| + \|f - P(A_{n_i}, \cdot)\| \leqslant \|f\| + \rho + K(\delta_{n_i})$, we know $\{P(A_{n_i}, x)\}_{i=1}^{\infty}$ is a bounded subset of a finite-dimensional space and therefore has a limit point $P(A^*, x)$. Then $\|f - P(A^*, \cdot)\| \leqslant \rho$ which implies $A^* \in \mathscr{A}$. However, $\sigma(A_{n_i}, \mathscr{A}) \geqslant \epsilon$ for $i = 1, 2,...$ implying $0 = \sigma(A^*, \mathscr{A}) \geqslant \epsilon$.

The following is a corollary whose proof is contained in the proof of Theorem 8.

COROLLARY 1. *There exists a subsequence of* $\{P(A_n, x)\}_{n=1}^{\infty}$ *which converges uniformly to* $P(A^*, x)$ *where* $P(A^*, x)$ *is some best uniform approximation to* $f(x)$.

The next corollary indicates that Corollary 1 can be strengthened to read "sequence" instead of "subsequence" if f has the unique best approximation $P(A^*, x)$.

COROLLARY 2. *If* f *has the unique best uniform approximation* $P(A^*, x)$ *then* $\lim_{n\to\infty} \| P(A_n, \cdot) - P(A^*, \cdot)\| = 0$. *For this corollary, an estimate on the rate of convergence can be obtained if the* ϕ_i*'s are assumed to satisfy theHaar condition. Because then*

$$\|f - P(B, \cdot)\| - \|f - P(A^*, \cdot)\| \geqslant \gamma \| P(B, \cdot) - P(A^*, \cdot)\|$$

by the "strong unicity theorem." Using Theorem 6 *again, we obtain* $\| P(A_n, \cdot) - P(A^*, \cdot)\| \leqslant \gamma^{-1}[\omega(\delta_n) + 2\alpha\beta \|f\| \Omega(\delta_n)]$ *where* γ *is dependent on* f*, whereas,* α *and* β *are not.*

THEOREM 9. *Let* $\Phi = \{\phi_1, \phi_2, ...\}$ *be a fundamental set in* $C(X)$. *Then there is a sequence of partitions* $\{U_{m_n}\}_{n=1}^{\infty}$ *such that the generalized polynomial of "degree n" of best approximation to f in the* m_n*-norm converges uniformly to f as* $n \to \infty$.

Proof. For each set $\{\phi_1, ..., \phi_n\}$, a constant $\beta_n = \sigma_n^{-1}$ may be computed as in Theorem 5, where $\sigma_n = \max\{\| \sum_{i=1}^{n} a_i\phi_i \| : \sum | a_i | = 1\}$. Pick $\alpha > 1$. Then for each n there is a partition U_{m_n} with mesh δ_{m_n} so that δ_{m_n} corresponds to α as obtained in Theorem 4. Assuming $\delta_{m_n} \to 0$ sufficiently fast gives that $\alpha\beta_n\Omega_n(\delta_{m_n}) \to 0$.

From the fundamentality of Φ, it follows that there are vectors $\{C_i\}_{i=1}^{\infty}$ and polynomials $P_n(C_n, x)$ of "degree n" such that $\lim_{n\to\infty} \|f - P_n(C_n, \cdot)\| = 0$. If we designate the polynomial of degree n of best approximation in the m_n-norm by $P_n(A_{m_n}, x)$ and a polynomial of degree n of best uniform approximation by $P_n(A, x)$, then by Theorem 6

$$\begin{aligned}\|f - P_n(A_{m_n}, \cdot)\| &\leqslant \|f - P_n(A, \cdot)\| + \omega(\delta_n) + 2\alpha\beta_n \|f\| \Omega(\delta_n)\\ &\leqslant \|f - P_n(C_n, \cdot)\| + \omega(\delta_n) + 2\alpha\beta_n \|f\| \Omega(\delta_n).\end{aligned}$$

And since the R.H.S. tends to zero as $n \to \infty$, the result is established.

3. Alternatives to the n-Norm

In practice it is easier to use the following pseudonorm to facilitate computation. Define

$$ps\,\|f\|_n = \max_{1\leqslant i\leqslant n}\left[\frac{1}{\mu E_i^n}\left|\int_{E_i^n} f\,d\mu\right|\right].$$

It should be noted that I may be giving up speed of convergence in choosing to work with $ps\,\|\cdot\|_n$ since it is easily verified that

$$ps\,\|f\|_n \geqslant \|f\|_n\,.$$

However, this choice reduced the problem to looking at error surfaces which are piecewise hyperplanes rather than at surfaces which are composed of quadratic hypersurfaces. This is because the residuals $r_k(A)$ for $ps\,\|\cdot\|_n$ represent hyperplanes whereas the residuals $r_k(A)$ for $\|\cdot\|_n$ are quadratic forms. Therefore choosing $ps\,\|\cdot\|_n$ for computation reduces to a problem which has already been solved, i.e., finding the minima for polytopes which are surfaces which are piecewise hyperplanes.

I now present the development for $ps\,\|\cdot\|_n$, which proceeds like that for $\|\cdot\|_n$.

Theorem 10. *Given $f \in C(X)$, there exists a best approximation to f from $\mathscr{P}_m$ in the pseudonorm $ps\,\|\cdot\|_n$.*

Proof. It is easily verified that "pseudonorm" can replace "norm" in the following statement: A finite dimensional linear subspace of a normed (pseudonormed) linear space contains at least one point of minimum distance from a fixed point.

As is usually the case for pseudonorms, best approximations prove to be nonunique. The following simple example illustrates this point.

Let $X = [-1, 1]$, $f = x - 1/2$, $x \geqslant 0$; $-x - 1/2$, $x < 0$. Let $\phi_1 = 1$, $\phi_2 = x$, $\phi_3 = x^2$ and $U_2 = E_1^2 \cup E_2^2$ where $E_1^2 = [-1, 0]$ and $E_2^2 = [0, 1]$. Let μ be one-dimensional Lebesgue measure. Then $\alpha[x^2 - 1/3]$ is a best approximation to f for each $\alpha \in [0, 1]$.

Let

$$b_i = \frac{1}{\mu E_i^n}\int_{E_i^n} f\,d\mu \qquad \text{and} \qquad c_{ik} = \frac{1}{\mu E_i^n}\int_{E_i^n} \phi_k\,d\mu.$$

Then

$$ps\,\|f - P(A, \cdot)\|_n = \max_{1\leqslant i\leqslant n}\left|\sum_{k=1}^{m} c_{ik}a_k - b_i\right|.$$

Let $r_i(A) = \sum_{k=1}^{m} c_{ik}a_k - b_i$. Given a vector A, let $\sigma_i(A) = \operatorname{sgn} r_i(A)$. Set $C_i = [c_{i1}, c_{i2}, \ldots, c_{im}]$ and designate $ps\,\|f - P(A, \cdot)\|$ by $\delta(A)$. Then we obtain the well-known

CHARACTERIZATION THEOREM. *A vector A is a minimum for δ if and only if the origin of m-space lies in the convex hull of the set of vectors* $\{\sigma_i C_i : |r_i(A)| = \delta(A)\}$.

THEOREM 11. (i) $ps\,\|f\|_n \leqslant ps\,\|f\|_m \leqslant \|f\|$, $n < m$.

(ii) $\|f\| < ps\,\|f\|_n + \omega(\delta_n)$ *for* δ_n *sufficiently small.*

Proof. (i) The proof is entirely analogous to that of Theorem 3(i).

(ii) $\|f\| = 0$ is trivial. Assume $\|f\| > 0$. Let $x_0 \in X$ be such that $|f(x_0)| = \|f\|$. For some i, $x_0 \in E_i^n$. For all $x \in E_i^n$ and δ_n sufficiently small, $f(x)$ does not change sign. Then

$$\begin{aligned}\|f\| - \omega(\delta_n) &= \frac{1}{\mu E_i^n} \int_{E_i^n} |f(x_0)| - \omega(\delta_n)\, d\mu \\ &\leqslant \frac{1}{\mu E_i^n} \int_{E_i^n} |f(x)|\, d\mu \\ &= \frac{1}{\mu E_i^n} \left| \int_{E_i^n} f(x)\, d\mu \right| \leqslant ps\,\|f\|_n .\end{aligned}$$

Remark. Theorems 4–9 hold with "$\|\cdot\|_n$" replaced by "$ps\,\|\cdot\|_n$" and "the" best approximation replaced by "a" best approximation.

4. CHANGES IN $ps\,\|\cdot\|_n$ NECESSITATED BY COMPUTATION

It would be desirable for the preceding estimates to go through essentially unchanged when in the computational process, the integral is computed by an integration formula. It is clear that given a particular integration formula and assuming the necessary conditions on f and the ϕ_i's in order to obtain an error bound, then as the mesh δ_n of U_n goes to zero, the error in the integration also goes to zero. Hence only the error bound for the formula need be combined with the estimate. However, if the formula $\int t = \sum w_i f_i$ has nonnegative weights and has precision zero, then the weights and the points where the weights are to be taken are associated with a positive measure μ. More precisely, let

$$(A)\ \|f\|_n = \max_{1 \leqslant k \leqslant n} \left| \sum_i w_{ik}^n f_{ik} \right|$$

with $\sum_i w_{ik}^n = 1$ for each k. Then $(A)\|\cdot\|$ is a pseudonorm on $C(X)$. The point of note is that the w_{ik}^n correspond to a positive measure μ_n, and hence the measure is now varying with the partition U_n. We obtain existence and nonuniqueness the same way as for $ps\|\cdot\|_n$. The characterization theorem is also similar. But as soon as U_m is introduced as a refinement of U_n, U_m must be acceptable in terms of the geometry for which the integration scheme is designed. Hereafter it will be assumed that this condition is satisfied.

The following example shows that the monotonicity part of Theorem 3(i), i.e., $(A)\|f\|_n \leqslant (A)\|f\|_m$, $n < m$ does not necessarily hold. Let $U_1 = \{[0, \pi]\}$, $U_2 = \{[0, \pi/2], [\pi/2, \pi]\}$ and $f(x) = \sin x$. Let the rule be Simpson's with $h = \pi/2$ for U_1 and $h = \pi/4$ for U_2. Then $(A)\|\sin x\|_1 = 4/6$ and $(A)\|\sin x\|_2 = 2\sqrt{2}/6 < 4/6$. Therefore $(A)\|\cdot\|_n$ cannot replace $\|\cdot\|_n$ in this part of the theorem. However, it does follow that $(A)\|f\|_n \leqslant \|f\|$ since

$$r_k = \sum_i |w_{ik}^n f_{ik}| \leqslant \sum w_{ik}^n |f_{ik}| \leqslant \max_i |f_{ik}| \sum_i w_{ik} = \max_i |f_{ik}| \leqslant \|f\|.$$

Theorem 3(ii) which states $\|f\| \leqslant \|f\|_n + \omega(\delta_n)$ also holds for $(A)\|f\|_n$. Theorems 4–9 are true with $\|\cdot\|_n$ replaced by $(A)\|\cdot\|_n$ with the exception that the monotone convergence in Theorem 7 must be deleted.

5. Computation and Conclusions

The following examples were calculated on the Univac 1108 computer at the University of Utah computer center. The computation was carried out using the pseudonorm $(A)\|\cdot\|_n$. The computation of best approximations was accomplished by taking the systems of linear inequalities to be solved and placing them into linear programming form. The residuals $r_i(A)$ are of the form $\sum_{j=1}^m \alpha_{ij}a_j - \beta_i$ where $A = [a_1, \ldots, a_m]$ and the α_{ij}'s and β_i's are known. We seek the smallest number ϵ such that $\max_{1\leqslant i\leqslant n} |r_i(A)| \leqslant \epsilon$. This "usually" overdetermined system can be placed in linear programming form by adding the appropriate "slack" variables. The integration formulas used were Simpson's rule for X one-dimensional and Simpson's product rule for X two-dimensional. In each of the following tables, the fifth column headed by the letters cyc. indicates how many matrix pivoting operations were needed to compute the best approximation. All values are given to four decimal places.

Example 1. $X = [0, 1], f(x) = x^2, \phi_i(x) = 1, \phi_2(x) = x.$

$$U_n = \left\{\left[\frac{i-1}{n}, \frac{i}{n}\right]\right\}_{i=1}^n$$

TABLE 1

Norm	Best approximation	$\|\cdot\|$ Error	$(A)\|\cdot\|_n$ Error	Cyc.
$\|\cdot\|$	$x - 0.1250$	0.1250	—	—
$(A)\|\cdot\|_{10}$	$x - 0.1467$	0.1467	0.0900	28
$(A)\|\cdot\|_{20}$	$x - 0.1367$	0.1367	0.1125	53
$(A)\|\cdot\|_{50}$	$x - 0.1299$	0.1299	0.1200	128

Note that the best $(A)\|\cdot\|_n$ approximations appear to be converging slowly to the best uniform approximation. In the latter part of this section, an algorithm is suggested which may speed up the convergence of this procedure.

EXAMPLE 2. $X = [-1, 1], f(x) = x^5, \phi_1(x) = 1, \phi_2(x) = x, \phi_3(x) = x^2, \phi_4(x) = x^3, \phi_5(x) = x^4$.

$$U_n = \left\{\left[-1 + \frac{2(i-1)}{n}, -1 + \frac{2i}{n}\right]\right\}_{i=1}^{n}.$$

TABLE 2

Norm	Best Approximation	$\|\cdot\|$ Error	$(A)\|\cdot\|_n$ Error	Cyc.
$\|\cdot\|$	$1.2500x^3 - 0.3125x$	0.0625	—	—
$(A)\|\cdot\|_{20}$	$1.1400x^3 - 0.2597x$	0.1197	0.0475	56
$(A)\|\cdot\|_{35}$	$1.1822x^3 - 0.2790x$	0.0968	0.0538	91
$(A)\|\cdot\|_{50}$	$1.2015x^3 - 0.2883x$	0.0868	0.0563	122

For $n = 50$, the associated linear programming problem consists of 102 equations in 214 unknowns. For $n > 50$, the problems becomes too unwieldy for the capacity of the Univac 1108.

EXAMPLE 3. $X = [-1, 1] \times [-1, 1]$, $f(x, y) = x^3y^3$, $\phi_1 = 1$, $\phi_2 = x$, $\phi_3 = y$, $\phi_4 = x^2$, $\phi_5 = xy$, $\phi_6 = y^2$, $\phi_7 = x^3$, $\phi_8 = x^2y$, $\phi_9 = xy^2$, $\phi_{10} = y^3$, $\phi_{11} = x^4$, $\phi_{12} = x^3y$, $\phi_{13} = x^2y^2$, $\phi_{14} = xy^3$, $\phi_{15} = y^4$. U_{n^2} was obtained by taking cross products of intervals of the form $[-1 + 2(i-1)/n, -1 + 2i/n]$.

The convergence of P_n to P in Example 3 seems to be slower than that obtained in Example 2. One possible explanation is that the mesh for Example 3 is much larger than that for Example 2.

TABLE 3

Norm	Best approximation	$\|\cdot\|$ Error	$(A)\|\cdot\|_n$ Error	Cyc.
$\|\cdot\|$	$0.5000x^3 + 0.7500xy^2 - 0.3750x$	0.1250	—	—
$(A)\|\cdot\|_{25}$	$0.3333x^3 + 0.5200xy^2 - 0.1733x$	0.3200	0.0409	74
$(A)\|\cdot\|_{36}$	$0.3704x^3 + 0.5556xy^2 - 0.2063x$	0.2803	0.0463	116
$(A)\|\cdot\|_{49}$	$0.3742x^3 + 0.5918xy^2 - 0.2214x$	0.2654	0.0504	141

Each of Examples 1, 2, and 3, illustrate Corollary 2 to Theorem 8 of Section 2 which states that if the best approximation $P(A, x)$ to $f(x)$ in $\|\cdot\|$ is unique, then the sequence $\{P(A_n, x)\}$ of best approximations to f in $(A)\|\cdot\|_n$ converges uniformly to $P(A, x)$.

Each example in this section illustrates Theorems 6 and 7 of Section 2 which show that the uniform norm of the error functions $f - P(A_n, \cdot)$ converges to the uniform norm of $f - P(A, \cdot)$ from above (Column 3 in the tables) and that $(A)\|\cdot\|_n$ of $f - P(A_n, \cdot)$ converges to the uniform norm of $f - P(A, \cdot)$ from below (column 4 in the tables).

EXAMPLE 4. $X = [0, 1] \times [0, 1], f(x, y) = xy, \phi_1 = 1, \phi_2 = x, \phi_3 = y, \phi_4 = x^2, \phi_5 = y^2$. U_{n^2} was obtained by taking cross products of intervals of the form $[(i-1)/n, i/n]$.

The set of best uniform approximations to f are of the form $af_1 + bf_2 a > 0$, $b > 0$, $a + b = 1$ and

$$f_1 = (1/2)(x^2 + y^2) - 1/4, \qquad f_2 = x + y - (1/2)(x^2 + y^2) - 1/4.$$

In particular, $(1/2)[f_1 + f_2] = x/2 + y/2 - 1/4$ is a best approximation and it is listed in the table.

TABLE 4

Norm	Best approximation	$\|\cdot\|$ Error	$(A)\|\cdot\|_n$ Error	Cyc.
$\|\cdot\|$	$0.5000x + 0.5000y - 0.2500$	0.2500	—	—
$(A)\|\cdot\|_{16}$	$0.5000x + 0.5000y - 0.2500$	0.2500	0.1406	38
$(A)\|\cdot\|_{25}$	$0.5000x + 0.5000y - 0.2500$	0.2500	0.1640	56
$(A)\|\cdot\|_{49}$	$0.5000x + 0.5000y - 0.1854$	0.3146	0.1837	104

The result in row 4, column 3 of Table 4 is understandable since Theorem 6 does not guarantee that $\|f - P(A_n, \cdot)\|$ converges monotonically to $\|f - P(A, \cdot)\|$.

EXAMPLE 5. $X = [0, 1] \times [0, 1], f(x, y) = e^{xy}, \phi_1 = 1, \phi_2 = x, \phi_3 = y$. U_{n^2} was obtained as in Example 4.

TABLE 5

Norm	Best approximation	$\|\cdot\|$ Error	$(A)\|\cdot\|_n$ Error	Cyc.
$(A)\|\cdot\|_{16}$	$0.5905 + 0.7621x + 0.7621y$	0.6041	0.2347	40
$(A)\|\cdot\|_{25}$	$0.5864 + 0.7775x + 0.7775y$	0.6399	0.2675	58
$(A)\|\cdot\|_{49}$	$0.5816 + 0.7973x + 0.7973y$	0.5421	0.3097	108

The fact that the convergence in the examples is slow, motivates a search for a more intelligent method of partitioning which will speed up the rate of convergence. Since it was shown that $(A)\|f - P(A_n, \cdot)\|_n \leqslant \|f - P(A, \cdot)\|$, we would like to choose the partitions so that the error in $(A)\|\cdot\|_n$ rises to meet the error in $\|\cdot\|$ as rapidly as possible. This problem is investigated in (4).

As an alternative to choosing a large value of n to obtain the desired approximation, I propose the following algorithm which begins with n small.

ALGORITHM. *Start with a partition U_{n_0} where n_0 is small and compute the best $(A)\|\cdot\|_{n_0}$ approximation. Find the members of the partition where the moduli of the residuals $r_i(A)$ take on their maximum value. Subdivide these members according to some prescribed scheme and repeat the process with the new partition U_{n_1} obtaining U_{n_2}, etc.*

REFERENCES

1. E. W. CHENEY, "Introduction to Approximation Theory," McGraw-Hill, New York, 1966.
2. R. M. HARRIS, Uniform Approximation of Functions: Approximation by Partitioning, Ph.D. Thesis, University of Utah, 1970.
3. S. E. WEINSTEIN, Approximation of Functions of Several Variables: Product Chebyshev Approximation. I, *J. Approximation Theory* **2** (1969), 433–447.
4. S. E. WEINSTEIN, Uniform Approximation of Functions through Optional Partitioning, In preparation.