JOURNAL OF APPROXIMATION THEORY 1, 476–483 (1968)

Approximation by Müntz Polynomials on Sequences

D. J. NEWMAN AND E. PASSOW

Yeshiva University, New York, New York 10033

AND

L. RAYMON

Temple University, Philadelphia, Pennsylvania 19122

Let S be a compact, nonempty, set of real numbers, let $P = \{\varphi_0, \varphi_1, \ldots\}$ be a sequence of distinct continuous real functions defined on S, and let P_n be the set of all linear combinations of $\varphi_0, \varphi_1, \ldots, \varphi_{n-1}$. The "degree of approximation" is defined to be

$$\rho_n(S,P) = \max_{f \in K} \min_{p \in P_n} \max_{x \in S} |f(x) - p(x)|,$$

where K is the set of "contractions," i.e., the set of all real functions f such that $|f(x) - f(y)| \leqslant |x - y|$ for all x, y in S.

In this terminology, Müntz's theorem [3] says that for $S = [0,1]$ and $P = \{1, x^{\lambda_1}, x^{\lambda_2}, \ldots\}$ (λ_i positive and distinct, $\lambda_i \not\to 0$), one has $\rho_n(S,P) \to 0$ as $n \to \infty$, if and only if $\sum (1/\lambda_k) = \infty$.

It is known [2] that $\frac{1}{2}\epsilon_n(S)$ is a lower bound for $\rho_n(S,P)$ over all possible choices of P, where $\epsilon_n(S)$, the "massivity" of S, is defined by

$$\epsilon_n(S) = \max_{y_1, y_2, \ldots, y_{n+1} \in S} \min_{i \neq j} |y_i - y_j|.$$

We shall say that P is "efficient" on S, if there exists a $C > 0$ such that $\rho_n < C\epsilon_n$. Also, P will be said to be "weakly efficient" on S, if there exists a $C > 0$, such that for each n, there exist $\varphi_{i_1}, \varphi_{i_2}, \ldots, \varphi_{i_n} \in P$, satisfying

$$\max_{f \in K} \min_{p \in Q_n} \max_{x \in S} |f(x) - p(x)| < C\epsilon_n(S),$$

where Q_n is the set of all linear combinations of $\varphi_{i_1}, \ldots, \varphi_{i_n}$. It follows immediately that P is weakly efficient on S whenever P is efficient on S. Jackson's theorem [1] says that $\{1, x, x^2, \ldots\}$ is efficient on $[0,1]$. It is known that this sequence is efficient on any real set of positive measure. One of the authors [4] has established that this sequence is efficient on the set consisting of $1, x_1, x_2, \ldots$ ($x_n > 0, \to 0$), if $\{x_n\}$ is logarithmically convex or if $x_{n+1}/x_n < C_1 < 1$. On the other hand, if the x_n do not have sufficiently "regular" growth, $\{1, x, x^2, \ldots\}$ may not be efficient on $\{1, x_1, x_2, \ldots\}$ (Ref. [4]).

In this article, too, we consider approximation on sequences $\{0, x_n\}$, where

$x_n \to 0$—but here we investigate approximation by linear combinations of given powers of x: $1, x^{\lambda_1}, \ldots, x^{\lambda_n}, \ldots$. First, we shall demonstrate how here Müntz's condition on the λ_i can be relaxed. We shall then investigate how fast $\rho_n \to 0$. Our goal is to find $\{x_k\}$ and/or $\{\lambda_k\}$ that will render P efficient on S In short, we seek Müntz–Jackson theorems on sequences. Our Müntz theorem on sequences is

THEOREM 1. *Let $S = \{0, x_n\}$ be a sequence of points in $[0,1]$ such that $x_n > x_{n+1}$, $\lim x_n = 0$, and let $P = \{1, x^{\lambda_i}\}$. Then $\rho_n(S,P) \to 0$ as $n \to \infty$, if $\{\lambda_n\}$ is any sequence of distinct positive numbers which does not tend to zero.*

We shall, in fact, prove that the sup-norm closure of the span of P on S, is $C(S)$—the set of all functions continuous on S.

We may assume, without loss of generality, that $\lambda_n \to \infty$; for, if not, the result is an immediate corollary of Müntz's theorem. Let $f(x)$ be an arbitrary continuous function in $C(S)$ with $f(0) = 0$. Given $\epsilon > 0$, choose N so large that

$$|f(x_n)| < \epsilon/2 \qquad \text{for all } n > N. \tag{1}$$

We observe that $\lim_{i\to\infty} (x/x_N)^{\lambda_i} = 0$, $x \in [0, x_{N+1}]$, and that the convergence is uniform on this interval. Thus, there exists an $i_1(\epsilon)$ such that

$$|f(x_N)| \left(\frac{x}{x_N}\right)^{\lambda_{i_1}} < \frac{\epsilon}{2(N+1)} \qquad \text{for all } x \in [0, x_{N+1}]. \tag{2}$$

Let $p_1(x) = f(x_N)(x/x_N)^{\lambda_{i_1}}$.

Since $(x/x_{N-1})^{\lambda_i} \to 0$ uniformly (as $i \to \infty$) on $[0, x_N]$, there exists an $i_2(\epsilon)$ such that

$$|f(x_{N-1}) - p_1(x_{N-1})| \left(\frac{x}{x_{N-1}}\right)^{\lambda_{i_2}} < \frac{\epsilon}{2(N+1)}$$

for all x in $[0, x_N]$. Let $p_2(x) = (f(x_{N-1}) - p_1(x_{N-1}))(x/x_{N-1})^{\lambda_{i_2}}$. In general, for $1 \leqslant k \leqslant N$, the uniform convergence of $(x/x_{N-k})^{\lambda_i}$ to 0 on $[0, x_{N-k+1}]$ assures the existence of an $i_k(\epsilon)$ such that

$$\left|f(x_{N-k}) - \sum_{j=0}^{k-1} p_j(x_{N-k})\right| \left(\frac{x}{x_{N-k}}\right)^{\lambda_{i_k}} < \frac{\epsilon}{2(N+1)} \qquad \text{for all } x \in [0, x_{N-k+1}]. \tag{3}$$

We then let

$$p_k(x) = \left[f(x_{N-k}) - \sum_{j=0}^{k-1} p_j(x_{N-k})\right] \left(\frac{x}{x_{N-k}}\right)^{\lambda_{i_k}}.$$

Let $p(x) = \sum_1^N p_k(x)$. Then $p(x)$ is a linear combination of the x^{λ_i}. From (2), (3) and the definition of p_k, it follows that:

(i) $p(0) = 0$;

(ii) $|p(x_n)| < \epsilon/2$ for all $n > N$, which, together with (1) gives $|f(x_n) - p(x_n)| < \epsilon$ for all $n > N$;

(iii) $|p(x_{N-k}) - f(x_{N-k})| < \epsilon(N-k)[2(N+1)]^{-1} < \epsilon$, $k = 0, 1, \ldots N-1$.

Hence $|f(x) - p(x)| < \epsilon$ for all x in S. The argument is completed by noting that if $f(0) \neq 0$, we have $|f(x) - (p(x) + f(0))| < \epsilon$. Q.E.D.

Note that the above argument extends, with some slight modification, to compact subsets of $[0, 1]$ with the property that every point is "isolated on the left." Without this restriction the argument fails, and, in fact, the corresponding theorem is false.

As the dual of Theorem 1, we have the following

COROLLARY.

$$f(x) = \sum_{n=1}^{\infty} c_n e^{-\mu_n x},$$

where $\mu_n \to \infty$, $\sum |c_n| < \infty$, *can have only a finite number of zeros in* (δ, ∞), $\delta > 0$.

This follows, since the completeness of P implies that there is no measure orthogonal to P, which in turn means that $\sum_{n=1}^{\infty} c_n x_n^{\lambda_i}$ cannot equal 0 for all i.

We now turn our attention to quantitative theorems. A sequence $S = \{0, x_n\}$, $x_n \in [0, 1]$, is called "thin," if there exists a C such that for all n, $x_n < C\epsilon_n(S)$.

THEOREM 2. *If* $S = \{0, x_n\}$ $(x_1 > x_2 > \ldots, x_n \to 0)$ *is a thin sequence, and* $P = \{1, x^{\lambda_i}\}$ $(\lambda_i > 0$ *and distinct*$)$, *then* P *is weakly efficient on* S *whenever* $\{\lambda_i\}$ *is unbounded.*

The proof is similar to the proof of Theorem 1. Let $f(x)$ belong to K, and suppose, without loss of generality, that $f(0) = 0$. For each N, we divide S into two parts, S_1, consisting of the $N-1$ largest elements of S, and S_2, the complement in S of S_1. We now proceed to define $p_k(x)$, $k = 1, 2, \ldots, N-1$, and $p(x)$ as in the proof of Theorem 1, replacing ϵ by the massivity, $\epsilon_N(S)$. Observe that p is a linear combination of $N-1$ monomials. In this manner, we obtain $|p(x) - f(x)| < \epsilon_N$ for all x in S_1. Also, $|p(x)| < \epsilon_N$ for all x in S_2. But since the sequence is thin and $f \in K$, we have $x \in S_2 \Rightarrow |f(x)| < C\epsilon_N$. Hence

$$|f(x) - p(x)| < (C+1)\,\epsilon_N \qquad \text{for all } x \in S,$$

completing the argument. (In case $f(0) \neq 0$, the constant subtracted is the Nth and last allowable monomial.)

Our next two theorems show that if $\lambda_n \to \infty$ in a sufficiently fast and regular manner, $P = \{1, x^{\lambda_i}\}$ is efficient on certain sets.

THEOREM 3. *Let* $S = \{0, x_n\}$ $(x_n > 0)$ *be a sequence for which there exist* C_1, C_2 *such that for all* n,

$$0 < \frac{1}{C_1} \leqslant \frac{x_{n+1}}{x_n} \leqslant C_2 < 1.$$

Then $P = \{1, x^{\lambda_i}\}$ $(\lambda_i > 0$ *and distinct*) *is efficient on* S, *whenever* $\lambda_{n+1}/n\lambda_n \to \infty$.

Proof. Note that the condition $x_{n+1}/x_n \leqslant C_2 < 1$ implies that $x_k < C\epsilon_k$, where $1/C = \min(1/C_2 - 1, 1)$, and hence that the sequence is thin. Thus, by Theorem 2, P is weakly efficient on S. Theorem 3 will be proved by choosing $\epsilon = \epsilon_N$ (N arbitrary) and demonstrating that the monomials selected in the proof of Theorem 2 can, under the more restrictive hypotheses of Theorem 3, be the first available monomials. To do this, we have to go back to the definition of the polynomials $p_k(x)$ (in the proof of Theorem 1) and show that $|p(x)| = |\sum_1^{N-1} p_k(x)| \leqslant C\epsilon_N$ for all x in $[0, x_{N+1}]$, where $\lambda_{i_k} = \lambda_k$. Theorem 3 follows then in the same way as does Theorem 2. Now,

$$p_{k+1}(x) = \left(\frac{x}{x_{N-k}}\right)^{\lambda_{k+1}} q_k(x),$$

where $q_k(x)$ consists of 2^k terms. Each term has the form $\pm f(x_{i_1})$ times the product of no more than k factors $(x_{i_2}/x_{i_3})^{\lambda_{i_4}}$, where

$$N - k \leqslant i_1 \leqslant N,\ N - k \leqslant i_2 < i_3,$$

and where i_4 are distinct integers, all $\leqslant k$. Since $f(x)$ is a contraction,

$$|f(x_{i_1})| \leqslant x_{i_1} \leqslant x_{N-k} \leqslant C_1{}^k x_N \leqslant C_1{}^k C\epsilon_N.$$

Also, on $[0, x_{N-k}]$, $|x/x_{N-k+1}|^{\lambda_{k+1}} < C_2^{\lambda_{k+1}}$. We then have the estimate

$$|p_k(x)| \leqslant C\epsilon_N 2^k C_1^{k \sum_0^{k-1} \lambda_i} C_2^{\lambda_k}.$$

A calculation now verifies that the conditions of our Theorem ensure that $|p(x)| < M\epsilon_N$ for all $x \in [0, x_{N+1}]$. Q.E.D.

THEOREM 4. $P = \{x^{C^n}\}$ $(C > 1)$ *is efficient on any set* $S = \{0, x_n\}$ *such that* $x_n > 0$, $x_{k+1}/x_k \leqslant M < 1$.

Proof. Let $f(x)$ be a contraction with $f(0) = 0$. Consider

$$p(x) = \sum_{j=0}^{n-1} f(x_j) \prod_{\substack{i=0 \\ i \neq j}}^{n-1} \frac{x^C - x_i{}^C}{x_j{}^C - x_i{}^C},$$

which interpolates to $f(x)$ at $0 = x_0, x_1, \ldots, x_{n-1}$. As in the previous theorem, $\{x_k\}$ is thin, i.e., there exists an $\alpha > 0$ such that $x_k \leqslant \alpha\epsilon_k$. Since $f(x)$ is a contraction, $|f(x)| \leqslant \alpha\epsilon_n$ for $x \in [0, x_n]$. We thus wish to show that there exists a β such that $|p(x)| \leqslant \beta\epsilon_n$ for $x \in [0, x_n]$.

Since, as a calculation verifies,

$$\text{for} \quad x \in [0, x_n], \left|\frac{x^C - x_i^C}{x_j^C - x_i^C}\right| < \left|\frac{x - x_i}{x_j - x_i}\right|,$$

we have

$$\begin{aligned}
|p(x)| &< \sum_{j=1}^{n-1} x_j \prod_{\substack{i=0\\ i\neq j}}^{n-1} \left|\frac{x - x_i}{x_j - x_i}\right| \\
&\leqslant x_n \sum_{j=1}^{n-1} \prod_{i=1}^{j-1} \frac{x_i}{x_i - x_j} \prod_{i=j+1}^{n-1} \frac{x_i}{x_j - x_i} \\
&\leqslant \alpha\epsilon_n \sum_{j=1}^{n-1} \prod_{i=1}^{j-1} \frac{1}{1 - (x_j/x_i)} \prod_{i=j+1}^{n-1} \frac{1}{1 - (x_i/x_j)} \prod_{i=j+1}^{n-1} (x_i/x_j) \\
&\leqslant \alpha\epsilon_n \sum_{j=1}^{n-1} \prod_{i=1}^{j-1} \frac{1}{1 - M^{j-i}} \prod_{i=j+1}^{n-1} \frac{1}{1 - M^{i-j}} \prod_{i=j+1}^{n-1} M^{i-j} \\
&\leqslant \alpha\epsilon_n \sum_{j=1}^{n-1} \prod_{i=1}^{\infty} \frac{1}{(1 - M^i)^2} \prod_{i=1}^{n-j-1} M^i \\
&\leqslant \alpha\epsilon_n \prod_{i=1}^{\infty} \frac{1}{(1 - M^i)^2} \sum_{j=1}^{n-1} M^{n-j-1} \\
&\leqslant \alpha\epsilon_n \left(\prod_{i=1}^{\infty} \frac{1}{(1 - M^i)^2} \sum_{i=1}^{\infty} M^i\right) \\
&\leqslant \beta\epsilon_n.
\end{aligned}$$

Thus, for $x \in [0, x_n]$,

$$|f(x) - p(x)| \leqslant |f(x)| + |p(x)| \leqslant (\alpha + \beta)\,\epsilon_n,$$

while for $x = x_k$, $k = 1, 2, \ldots, n-1$, we have $f(x) - p(x) = 0$. Q.E.D.

We now turn specifically to the set of points $0, C, C^2, C^3, \ldots, C^k, \ldots$, where C is a fixed number in $(0, 1)$. We are able to determine *exactly* when the sequence $\{1, x^{\lambda_1}, x^{\lambda_2}, \ldots, x^{\lambda_n}, \ldots\}$ is efficient on this set. The condition is given by the following

THEOREM 5. *The sequence* $\{1, x^{\lambda_n}\}$ $(\lambda_n > 0$ *and distinct*) *is efficient on the set* $\{0; C^k\}$ *if and only if* $\sum C^{\lambda_n} < \infty$.

Several curious corollaries follow for $\{0, C, C^2, \ldots\}$.

COROLLARY 1. *Efficiency is unaffected by a permutation of the sequence.*

COROLLARY 2. *Every sequence with unbounded* λ_n *has an efficient subsequence.*

COROLLARY 3. *Any subsequence of an efficient sequence is itself efficient.*

COROLLARY 4. *There exists a P such that P is weakly efficient but not efficient.*

PROOF OF THEOREM 5

We turn first to the necessity of our condition. If our sequence is efficient, there exists for each n a "polynomial" $P(x) = a_0 + a_1 x^{\lambda_1} + \ldots a_n x^{\lambda_n}$ such that, for all k,

$$|P(C^k)(-1)^k + C^k| \leqslant AC^n, \qquad A \text{ independent of } k \text{ and } n. \tag{4}$$

This results from the fact that the function f defined by

$$f(C^k) = (-1)^k C^k \cdot [(1 - C)/(1 + C)]$$

is a contraction.

Now write $\lambda_0 = 0$ for convenience, and then set

$$\begin{aligned} P(t) &= t(C^{\lambda_0} + t)(C^{\lambda_1} + t) \ldots (C^{\lambda_n} + t) \\ &= A_1 t + A_2 t^2 + A_3 t^3 \ldots . \end{aligned} \tag{5}$$

Next, form

$$Q = A_1(P(C)\cdot(-1) + C) + A_2(P(C^2) + C^2) + \ldots + A_k(P(C^k)\cdot(-1)^k + C^k) + \ldots.$$

On the one hand, by (4), the quantity Q satisfies

$$|Q| \leqslant (A_1 + A_2 \ldots) AC^n = p(1) AC^n, \tag{6}$$

while, on the other hand,

$$Q = (A_1 C + A_2 C^2 + A_3 C^3 + \ldots) - \sum_{i=0}^{n} a_i (A_1 C^{\lambda_i} - A_2 C^{2\lambda_i} + A_3 C^{3\lambda_i} - \ldots)$$

$$= p(C) - \sum_{i=0}^{n} a_i p(-C^{\lambda_i}) = p(C), \qquad \text{since each } p(-C^{\lambda_i}) = 0, \text{ by (5)}.$$

Combining this with (6), gives $p(C) \leqslant p(1)AC^n$, or, in other words,

$$\frac{1+C^{\lambda_1-1}}{1+C^{\lambda_1}} \cdots \frac{1+C^{\lambda_n-1}}{1+C^{\lambda_n}} \leqslant \frac{2A}{C(1+C)}.$$

Clearly, this implies the convergence of the infinite product

$$\prod_{n=1}^{\infty} \frac{1+C^{\lambda_n-1}}{1+C^{\lambda_n}}$$

which, in turn, yields the convergence of $\sum_{n=1}^{\infty} C^{\lambda_n}$.

Next, we prove the sufficiency by showing, in fact, that the "interpolating polynomials" perform the required approximation.

Since we are assuming $\sum C^{\lambda_n} < \infty$, we know that $\lambda_n > 1$ for n large enough. Also, since omitting a finite number of the λ_n does not effect efficiency, we may assume that $\lambda_n > 1$ for all n.

Let n be an arbitrary positive integer. Define $P(x;\lambda)$ as that linear combination of $x^{\lambda_1}, x^{\lambda_2}, \dots x^{\lambda_n}$ which satisfies $P(x;\lambda) = x^\lambda$ for $x = C, C^2, \dots, C^n$.

Existence and uniqueness both follow from the fact that the underlying matrix is a Vandermonde.

In particular, then, if

$$\lambda = \lambda_i, \quad i = 1, 2, \dots n, \qquad \text{then } P(x;\lambda) \text{ is } x^{\lambda_i}. \tag{7}$$

Now let $p_k(x)$, $k = 1, 2, \dots, n$, denote the basic "interpolating polynomials," defined by $p_k(C^m) = \delta_{km}$ for $m = 1, 2, \dots, n$, and by the requirement that each p is a linear combination of the x^{λ_i}, $i = 1, 2, \dots n$.

Obviously, we can write $P(x;\lambda)$ in terms of $p_k(x)$ as

$$P(x;\lambda) = \sum_{k=1}^{n} p_k(x) C^{\lambda k}. \tag{8}$$

Now we adopt a new point of view. We hold x fixed at C^m and consider λ as the variable. We find, then, that we can identify P.

Namely, let $Q(z)$ denote the $(m-n)$th partial sum of the power series for

$$\frac{1}{(1 - C^{\lambda_1} z)(1 - C^{\lambda_2} z) \dots (1 - C^{\lambda_n} z)}$$

(interpreted as 0 if $m \leqslant n$). We claim that

$$P(C^m, \lambda) = C^{m\lambda}(1 - Q(C^{-\lambda})(1 - C^{\lambda_1 - \lambda}) \dots (1 - C^{\lambda_n - \lambda})). \tag{9}$$

For, observe that our choice of Q forces the right-hand side to be a linear combination of $C^\lambda, C^{2\lambda}, \dots, C^{n\lambda}$, and also that this right-hand side reduces to $C^{m\lambda}$ for $\lambda = \lambda_i$, $i = 1, 2, \dots n$. By (7) and (8), these same properties hold for the left-hand side, and we conclude easily that the two sides must be identical. (Two nth-degree polynomials which agree at $n + 1$ points must be identical.)

If now, in (9), we replace the partial sum, Q, by the full sum

$$\frac{1}{(1-C^{\lambda_1-\lambda})\dots(1-C^{\lambda_n-\lambda})},$$

and if we replace the expression

$$-(1-C^{\lambda_1-\lambda})\dots(1-C^{\lambda_n-\lambda}) \quad \text{by} \quad (1+C^{\lambda_1-\lambda})\dots(1+C^{\lambda_n-\lambda}),$$

then we increase the moduli of all coefficients of the powers of C^{λ}.

Combining this observation with (8) tells us that for λ smaller than all the λ_i,

$$\sum_{k=1}^{n} |p_k(C^m)|\, C^{\lambda k} \leqslant C^{m\lambda}\left(1+\frac{(1+C^{\lambda_1-\lambda})\dots(1+C^{\lambda_n-\lambda})}{(1-C^{\lambda_1-\lambda})\dots(1-C^{\lambda_n-\lambda})}\right).$$

We now set $\lambda = 1$, recalling that the λ_i are >1 and that $\sum C^{\lambda_j} < \infty$, and conclude that

$$\sum_{k=1}^{n} |p_k(C^m)|\, C^{k} \leqslant AC^m, \tag{10}$$

where

$$A = 1+\prod_{j=1}^{\infty}\frac{1+C^{\lambda_j-1}}{1-C^{\lambda_j-1}}.$$

Finally, let $f(x)$ be a contraction with $f(0)=0$. Form its "interpolating polynomial," $P_f(x)=\sum_{k=1}^{n} p_k(x) f(C^k)$, and consider $|P_f(x)-f(x)|$.

For $x=C^m$, $m\leqslant n$, this is equal to 0. For $x=C^m$, $m>n$, however, we have

$$|P_f(x)-f(x)| = |\sum_{k=1}^{n} p_k(C^m) f(C^k) - f(C^m)| \leqslant \sum_{k=1}^{n} |p_k(C^m)|\, C^k + C^m.$$

(since $|f(C^k)| = |f(C^k)-f(0)| \leqslant C^k$).

By (10), this, in turn, is $\leqslant AC^m + C^m \leqslant (A+1)C^n$.

But this is exactly the required result, since, obviously, the sequence is thin. Q.E.D.

References

1. D. Jackson, "The Theory of Approximation" [American Mathematical Society Colloq. Publ., Vol. XI, New York, 1930.]
2. G. G. Lorentz, Lower bounds for the degree of approximation. *Trans. Am. Math. Soc.* **97** (1960), 25–34.
3. C. Müntz, Über den Approximationssatz von Weierstrass. *Schwarz-Festschrift* (1914), 302–312.
4. D. J. Newman, Efficiency of polynomials on sequences. *J. Approx. Theory* **1** (1968), 66–76.